



HAL
open science

Y recombination arrest and degeneration in the absence of sexual dimorphism

Thomas Lenormand, Denis Roze

► **To cite this version:**

Thomas Lenormand, Denis Roze. Y recombination arrest and degeneration in the absence of sexual dimorphism. *Science*, 2022, 375 (6581), pp.663-666. 10.1126/science.abj1813 . hal-03755019

HAL Id: hal-03755019

<https://hal.science/hal-03755019>

Submitted on 20 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Y recombination arrest and degeneration in the absence of sexual dimorphism

Authors: Thomas Lenormand^{1*}, Denis Roze^{2,3}

Affiliations:

¹ CEFE, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France.

² CNRS, IRL 3614, Roscoff, France.

³ Sorbonne Université, Station Biologique de Roscoff, France.

*Correspondence to: thomas.lenormand@cefe.cnrs.fr

Abstract:

Current theory proposes degenerated sex chromosomes, such as the mammalian Y, evolve via three steps: recombination arrest, linking male-beneficial alleles to the Y chromosome; Y degeneration due to the inefficacy of selection in the absence of recombination; and dosage compensation correcting the resulting low expression of X-linked genes in males. Here we investigate a model of sex chromosome evolution incorporating the coevolution of cis- and trans-regulators of gene expression. We show that the early emergence of dosage compensation favors the maintenance of Y-linked inversions by creating sex-antagonistic regulatory effects. This is followed by degeneration of these non-recombining inversions caused by regulatory divergence between the X and Y chromosomes. In contrast to current theory, the whole process occurs without any selective pressure related to sexual dimorphism.

One Sentence Summary:

Turning sex chromosome theory on its head: early evolution of dosage compensation can maintain successively forming Y chromosome strata that undergo genetic degeneration.

Main Text:

Many species have chromosomal sex determination systems (1). In XX/XY systems, as in mammals, males are heterogametic XY. In ZZ/ZW systems, as in birds, females are heterogametic ZW. We mention only XY systems below, but all arguments apply equally well to ZW systems. Y chromosomes are often non-recombining and have degenerated through the loss of most of the genes present on ancestral autosomes. In several chiasmate species, such as mammals or birds, suppression of recombination involves successive events, each affecting Y sub-regions of different sizes, called strata (2). These strata are detected on the basis of different degrees of sequence divergence from the homologous X regions (2).

Following the establishment of a sex-determining locus on an autosome, current theory (3–5) proposes that Y chromosomes evolve through three steps. First, sex chromosomes evolve recombination suppression because selection favors linkage between sex-determining and sexually

39 antagonistic genes (6–9). These sexually antagonistic genes occur when trait optima differ between
40 the sexes, driving the evolution of sexual dimorphism. In the second step, the absence of
41 recombination reduces the efficacy of natural selection by causing “selective interference”. Such
42 interference leads to an accumulation of deleterious mutations on the Y chromosome and genetic
43 degeneration (10). Finally, dosage compensation evolves to restore optimal gene expression in
44 males, whose Y-linked genes have lowered expression due to degeneration, and possibly in
45 females if dosage compensation mechanisms alter expression in that sex (7, 11, 12). The
46 compensation process involves various mechanisms in different species, and compensation is not
47 always complete for all X-linked genes (13–15).

48 This theory has been explored over the past c.a. 50 years, both empirically and theoretically (3–6,
49 16). Empirical support for the first step is equivocal, where, despite decades of investigation,
50 decisive evidence for a causal role of sexually-antagonistic loci on recombination arrest is lacking
51 (16–19). The second step is difficult to reconcile with the observation of small degenerated strata
52 (16), within which selective interference should be minimal. Lastly, the causal ordering of events
53 has also been challenged by observations of the early evolution of partial dosage compensation in
54 young sex-chromosomes (20–24).

55 Theoretically, each step suffers from limitations (25). However, an important global limitation is
56 that each step has generally been considered independently from the others, resulting in a
57 piecemeal set of models lacking integration. In particular, changes in gene regulation have not
58 been consistently studied throughout sex chromosome evolution. Yet, non-coding mutations can
59 influence the evolution of sex-limited expression, contribute to compensatory adaptive silencing,
60 and are pivotal for the evolution of dosage compensation.

61 We propose here that the joint evolution of regulatory changes and accumulation of
62 deleterious mutations can transform an autosome into a degenerated sex chromosome with dosage-
63 compensation. We use individual-based stochastic simulations assuming a population of N_{pop}
64 diploid individuals, with XY males and XX females (25, fig 1). We consider the evolution of a
65 pair of autosomes carrying hundreds of genes subject to partially recessive deleterious mutations,
66 with one homolog recently acquiring a sex-determining locus. Gene expression is controlled by
67 cis-regulatory sequences (affecting expression only on the same chromosome as themselves)
68 interacting with trans-regulators that can affect the gene copies on both homologs (26). All these
69 elements can mutate. To allow for dosage compensation on a gene-by-gene basis, but keeping the
70 model symmetric for males and females, we assume that each gene is controlled by one male- and
71 one female-expressed trans-regulator (25, fig 1). As in (27), we assume that each gene's overall
72 expression level is under stabilizing selection around an optimal level and that the relative
73 expression of the two copies of each gene determines the dominance level of a deleterious mutation
74 occurring in the coding gene. For instance, a deleterious mutation occurring in a more lowly
75 expressed gene copy is assumed to be less harmful than one in a more highly expressed copy (25).

76 We then assume that mutations occur that suppress recombination on a segment of the Y.
77 We refer to these mutations as inversions for simplicity, although they could correspond to other
78 mechanisms causing recombination arrest (25). Inversions of any size can occur, but we follow
79 only those on the Y that include the sex-determining locus, which will necessarily be confined to
80 males and cause recombination arrest. We assume that inversions can add up, such that new
81 inversions can occur on chromosomes carrying a previous inversion, and thus extend the non-
82 recombining part of the Y. Finally, we assume that reversions restoring recombination can occur,
83 and, for simplicity, that such reversions cancel only the most recent inversion (25).

84 To understand the dynamics of sex chromosome evolution in our model, first consider the
85 case where the cis- and trans-regulators do not mutate. In this case, all inversions on the Y are
86 eventually reversed and lost. This occurs in two steps. First, an inversion appears on a given Y and
87 “freezes” a segment of the chromosome. If, by chance, this Y carries relatively fewer or milder
88 deleterious mutations, this “lucky” inversion has a selective advantage. Consequently, it tends to
89 fix among Y chromosomes, causing recombination suppression in this portion of the sex
90 chromosomes. Larger inversions are overrepresented among these lucky inversions, as they
91 contain more genes and exhibit a larger fitness variance (25, fig S1A). Once fixed, these Y
92 chromosomes start accumulating deleterious mutations due to selective interference. Fitness
93 declines faster for larger inversions due to stronger selective interference (fig S1B). When the
94 inversion’s marginal fitness becomes lower than the fitness of the corresponding chromosomal
95 segment on the X, reversions are selectively favored and spread, which restores recombination.
96 Thus, Y-specific inversions are short-lived and maintained only transiently in the population in the
97 absence of regulatory mutations (fig S1C). These periods of recombination suppression do not last
98 long enough to lead to Y chromosome degeneration.

99 A radically different four-step process emerges when the regulatory sequences can mutate
100 and evolve (fig 2). It starts, as before, with the fixation of a lucky inversion on the Y. However,
101 once the inversion stops recombination, X and Y cis-regulators start evolving independently (step
102 2). This creates a positive feedback loop that causes rapid degeneration of Y-linked alleles (27):
103 by chance, some genes on the Y become slightly less expressed than their X-linked allelic
104 counterparts and accumulate more deleterious mutations (because lower expression makes
105 mutations more recessive), selecting for a further reduction of expression of these Y linked genes.
106 This process can work on individual genes, irrespective of the size of the non-recombining region
107 created by the inversion (27), and the subsequent degeneration does not involve selective
108 interference. However, like in the absence of regulator evolution, recombination arrest also triggers
109 the accumulation of deleterious mutations by selective interference, especially if the inversion
110 includes many genes.

111 The key step is the third in which inversions are stabilized in the long term, even when they
112 become entirely degenerated (fig. 3, S5). Cis-regulator divergence and degeneration in step 2 cause
113 a departure from optimal expression levels in males. Assuming that gene expression is under
114 stabilizing selection, this causes divergence in sex-specific trans-regulators, which evolve to
115 maintain optimal expression in both sexes. For instance, if a Y cis-regulator mutates, causing lower
116 expression, this will favor a stronger allele of the male trans-regulator, to maintain optimal
117 expression levels. The divergence of X- and Y-linked cis-regulators, and the divergence of sex-
118 limited trans-regulators, automatically generate sexually-antagonistic fitness effects: X cis-
119 regulators that recombine onto the Y would result in overexpression in males (due to mismatches
120 with male trans-regulators), and similarly Y cis-regulators recombined onto the X would cause
121 under-expression in females. Hence, if a reversion occurs, the reestablished recombination
122 between X and Y would likely reduce offspring fitness by creating a mismatch between cis and
123 trans-regulators. This sexually antagonistic effect caused by nascent dosage compensation protects
124 diverging inversions from reversion. This is the ultimate cause of Y recombination suppression in
125 our model (25). However, suppose dosage compensation does not evolve quickly enough. In such
126 a case, recombination can be restored: after a reversion, a new recombinant Y can be produced
127 that carries a non-degenerated part of the X without causing strong cis and trans-regulator
128 mismatch in males. This new Y can then replace the previous non-recombining degenerated Y,
129 which restores recombination on the part of the Y derived from the reversion.

130 Of course, only a minority of inversions evolve this nascent dosage compensation within a
131 fast enough timeframe, relative to the speed of degeneration, to remain immune to reversion
132 (meaning that they remain, at all time, unlikely to be selectively outcompeted by recombinant
133 chromosomes arising following a reversion). However, a positive feedback loop is also operating
134 here. Namely, when an inversion starts evolving dosage compensation it becomes relatively
135 immune to reversion and is maintained longer in the population, giving it more time to evolve
136 further dosage compensation. The inversion becomes eventually completely degenerated with
137 complete dosage compensation (for dosage-sensitive genes). This leads to very strong sexually-
138 antagonistic regulatory effects, which effectively make the inversion irreversibly immune to
139 reversions.

140 In our model, recombination suppression evolves along with regulatory evolution, but,
141 paradoxically, is opposed by selective interference. The evolution of nascent dosage compensation
142 involves the fixation of compensatory mutations and is partly adaptive. However, if selective
143 interference is too strong, inversions accumulate deleterious mutations too fast and are quickly
144 replaced by reversions. Accordingly, stabilized inversions tend to be strongly biased towards small
145 sizes, though less so when the population size is larger (fig S2C). In large populations,
146 recombination suppression and degeneration evolve more quickly, since more inversions occur
147 and selective interference (whose effect is stronger in smaller populations) is relatively less
148 efficient at removing large inversions (fig S2). Finally, as expected, this overall process is faster
149 when the intensity of stabilizing selection on gene expression levels is strong. This is because
150 selection on dosage fosters the evolution of dosage compensation and concurrently protects
151 partially degenerated inversions from reversions (fig S3).

152 Thus, our model suggests that the Y chromosome is entangled in a regulatory trap leading
153 to recombination arrest and degeneration, even in the absence of selective pressures related to
154 sexual dimorphism. Indeed, unlike previous theories (6–9), our model only includes genes with
155 the same optimal expression level in males and females and deleterious mutations that have the
156 same effect in both sexes. This process is inherently stochastic, as it involves the rare stabilization
157 of a handful of inversions, and is highly variable (fig S4). However, it works faster in larger
158 populations, as selective interference opposes recombination arrest and the stabilization of large
159 strata.

160 Our model also reverses the causality proposed by previous theory by showing that dosage
161 compensation can cause recombination suppression, rather than being a consequence of
162 degeneration after such suppression. Sexually-antagonistic effects are involved in the evolution of
163 suppressed recombination. However, they result from the fact that one sex is heterogametic, not
164 from males and females having divergent sex-specific optima for reproductive traits or expression
165 levels. All genes whose dosage affects fitness can contribute to the process, not just a subset of
166 sexually-antagonistic loci. The potential sexually-antagonistic effect of dosage compensation has
167 long been appreciated (7, 12, 28–30). However, its potential role in recombination arrest has not
168 been previously recognized, as it is usually thought to occur late in the degeneration process. Once
169 recombination has stopped, sexually-antagonistic alleles can arise and be maintained (9, 31), but
170 as shown here, they are not required for recombination arrest.

171 We showed that the emergence of non-recombining and degenerated sex chromosomes in
172 diploid organisms requires very few ingredients: genetic sex determination, deleterious mutations,
173 inversions, sex-specific trans-regulators, and stabilizing selection on gene expression levels. This
174 theory includes all steps (fig 4, fig S8) in a single set of assumptions and is compatible with current
175 data on sex chromosome evolution in chiasmate species (25). It predicts the occurrence of strata,

176 including small ones (reviewed by 16) and the occurrence of early regulatory changes in young
177 sex-chromosomes (20–24). It also accounts for the lack of decisive evidence for a causal role of
178 sexually-antagonistic loci on recombination arrest (16–19). Overall, this theory explains the rapid
179 expansion, degeneration, and dosage compensation of the non-recombining region of sex
180 chromosomes without requiring preexisting selection pressures favoring sexual dimorphism.

181 **References and Notes:**

- 182 1. D. Bachtrog *et al.*, *PLoS Biol.* **12**, e1001899 (2014).
- 183 2. D. Charlesworth, B. Charlesworth, G. Marais, *Heredity*. **95**, 118–128 (2005).
- 184 3. J. K. Abbott, A. K. Nordén, B. Hansson, *Proc. R. Soc. B Biol. Sci.* **284**, 1–9 (2017).
- 185 4. L. W. Beukeboom, N. Perrin, *The Evolution of Sex Determination* (Oxford University
186 Press, 2014).
- 187 5. D. Bachtrog, *Nat. Rev. Genet.* **14**, 113–124 (2013).
- 188 6. J. J. Bull, *Evolution of sex determining mechanisms* (Benjamin Cummings, Menlo Park,
189 CA, 1983).
- 190 7. B. Charlesworth, *Proc. Natl. Acad. Sci. U. S. A.* **75**, 5618–5622 (1978).
- 191 8. D. Charlesworth, B. Charlesworth, *Genet. Res.* **35**, 205–214 (1980).
- 192 9. W. Rice, *Evolution*. **41**, 911–914 (1987).
- 193 10. B. Charlesworth, D. Charlesworth, *Philos. Trans. R. Soc. B Biol. Sci.* **355**, 1563–1572
194 (2000).
- 195 11. J. Engelstädter, *Genetics*. **180**, 957–967 (2008).
- 196 12. B. Charlesworth, *Curr. Biol.* **6**, 149–162 (1996).
- 197 13. L. Gu, J. R. Walters, *Genome Biol. Evol.* **9**, 2461–2476 (2017).
- 198 14. A. Muyle, R. Shearn, G. A. Marais, *Genome Biol. Evol.* **9**, 627–645 (2017).
- 199 15. J. E. Mank, *Trends Genet.* **29**, 677–683 (2013).
- 200 16. D. Charlesworth, *Evolution*. **75**, 569–581 (2021).
- 201 17. D. Charlesworth, *Phil. Trans. R. Soc. B.* **372**, 20160456 (2017).
- 202 18. J. E. E. Ironside, *BioEssays*. **32**, 718–726 (2010).
- 203 19. S. Ponnikas, H. Sigeman, J. K. Abbott, B. Hansson, *Trends Genet.* **34**, 492–503 (2018).
- 204 20. P. Veltsos *et al.*, *Genetics*. **212**, 815–835 (2019).
- 205 21. K. H. C. Wei, D. Bachtrog, *PLoS Genet.*, e1008502 (2019).
- 206 22. J. Hough, J. D. Hollister, W. Wang, S. C. H. Barrett, S. I. Wright, *Proc. Natl. Acad. Sci.*
207 *U. S. A.* **111**, 7713–8 (2014).
- 208 23. H. Martin *et al.*, *Genome Biol. Evol.* **11**, 350–361 (2019).
- 209 24. A. Muyle *et al.*, *PLoS Biol.* **10**, e1001308 (2012).
- 210 25. see supplementary materials.

- 211 26. M. S. Hill, P. Vande Zande, P. J. Wittkopp, *Nat. Rev. Genet.* **22**, 203–215 (2021).
- 212 27. T. Lenormand, F. Fyon, E. Sun, D. Roze, *Curr. Biol.* **30**, 3001–3006.e5 (2020).
- 213 28. S. Ohno, *Sex Chromosomes and Sex-Linked Genes* (Springer Berlin Heidelberg, Berlin,
214 Heidelberg, 1967), vol. 1 of *Monographs on Endocrinology*.
- 215 29. C. Mullon, A. E. Wright, M. Reuter, A. Pomiankowski, J. E. Mank, *Nat. Commun.* **6**, 1–
216 10 (2015).
- 217 30. J. Engelstädter, D. Haig, *Evolution.* **62**, 2097–2104 (2008).
- 218 31. W. R. Rice, *Evolution.* **38**, 735–742 (1984).
- 219 32. code available on zenodo, DOI 10.5281/zenodo.5504423.
- 220 33. F. Manna, G. Martin, T. Lenormand, *Genetics.* **189** (2011),
221 doi:10.1534/genetics.111.132944.
- 222 34. L. Rowe, S. F. Chenoweth, A. F. Agrawal, *Am. Nat.* **192**, 274–286 (2018).
- 223 35. R. Bonduriansky, S. F. Chenoweth, *Trends Ecol. Evol.* **24**, 280–288 (2009).
- 224 36. G. S. Van Doorn, M. Kirkpatrick, *Nature.* **449**, 909–912 (2007).
- 225 37. H. Ellegren, J. Parsch, *Nat. Rev. Genet.* **8**, 689–698 (2007).
- 226 38. C. D. Meiklejohn, J. Parsch, J. M. Ranz, D. L. Hartl, *Proc. Natl. Acad. Sci. U. S. A.* **100**,
227 9894–9899 (2003).
- 228 39. P. Milesi *et al.*, *Mol. Ecol.* **27**, 493–507 (2018).
- 229 40. D. Bachtrog, *Genetics.* **179**, 1513–1525 (2008).
- 230 41. M. Nei, K. I. Kojima, H. E. Schaffer, *Genetics.* **57**, 741–750 (1967).
- 231 42. J. Maynard Smith, *The evolution of sex* (1978).
- 232 43. S. Ahmed *et al.*, *Curr. Biol.* **24**, 1945–1957 (2014).
- 233 44. S. B. Carey *et al.*, *Sci. Adv.* **7**, eabh2488 (2021).
- 234 45. S. Branco *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **114**, 7067–7072 (2017).
- 235 46. L. Boyer, R. Zahab, M. Mosna, C. Haag, T. Lenormand, *Evol. Lett.* **5**, 164–174 (2021).

236 **Acknowledgments:** We thank D. Charlesworth, G. Marais, Y. Michalakis, and four anonymous
237 reviewers for comments and suggestions and K. McKean for editing. We thank MBB cluster from
238 Labex CEMEB, and CNRS ABiMs cluster. **Funding:** This work was supported by grant GenAsex
239 ANR-17-CE02-0016-01. **Author contributions:** Original idea TL, DR; Model conception TL, DR;
240 Code DR, TL; Simulations TL; Data analyses TL; Interpretation TL, DR; First draft, editing and
241 revisions TL, DR; Project management and funding TL. **Competing interests:** the authors declare
242 no conflict of interest. **Data and materials availability:** simulation code is available on Zenodo
243 (32)

244 **Supplementary material.**

245 Materials and Methods

246 Supplementary Text

247 Figs. S1 to S8

249

250 **Fig. 1. An overview of the simulated genome evolving sex chromosomes.** A chromosome pair
 251 carries the sex locus at one end with two alleles (purple X/ light purple Y) determining two sexes
 252 (XX female, XY male). This chromosome carries 500 coding genes, each with a cis-regulatory
 253 region. Each cis-regulator interacts with a trans-acting factor. This trans-acting factor is not on the
 254 sex chromosomes but is expressed from a pair of autosomal trans-regulators, which differ in males
 255 and females. See text for other assumptions of the model.

256

257 **Fig 2. Example of a typical Y degeneration process.** The Y progressively degenerates by the
 258 accumulation of inversions, which accumulate deleterious mutations, evolve dosage compensation
 259 with sex-antagonistic fitness effects, and become immune to reversions. **(A)** The black stairplot
 260 shows the extension of each successive stratum of the Y (expressed as the fraction of the physical
 261 length of the Y), corresponding to stabilized inversions. Gray dots : average fraction of the physical
 262 length of the non-recombining Y in the population. Red: proportion of Y genes that are silenced
 263 and knock-out (i.e. they accumulated deleterious mutations effects up to the maximal value s_{max} ,
 264 here $s_{max} = 0.3$). At this timescale, silencing and degeneration appear as simultaneous, but silencing
 265 is slightly ahead. **(B)** \log_{10} of the average effect of deleterious mutations carried by inversions
 266 when they first arise in the population (averaged over the different genes within the inversion).
 267 Gray dots: random subsample of inversions that are lost before fixing in the population. Black
 268 dots: inversions that reach fixation, but are lost after the occurrence of a reversion. Red dots:
 269 inversions that reach fixation, and become stabilized strata on the Y. **(C)** Mean dominance of
 270 deleterious mutations on each stabilized inversion (noted h_Y). Initial dominance of deleterious
 271 mutations is assumed to be 0.25 (25). Fig S7 shows the detailed dynamics of h_Y at a smaller time
 272 scale. **(D)** Accumulation of deleterious mutation on each stabilized inversion (the maximum effect
 273 s_{max} , is set to 0.3 for all genes). **(E)** Fitness that the Y carrying the stabilized inversions would have
 274 on average, if expressed in a female (relative to the actual average fitness of males). The different
 275 colors highlight the occurrence of the successive strata. The average fitness of males that would
 276 carry two X chromosomes at that time is indicated in gray, but yields very similar values and is
 277 therefore almost undistinguishable. This simulation considers a population of $N_{pop} = 10^4$
 278 individuals, an intensity of stabilizing selection on dosage $I = 0.1$, and a mean effect of deleterious
 279 mutation $s_{mean} = 0.05$. See (25) for other parameter values.

280

281

282 **Fig 3.** Fitness trajectories of stabilized and lost inversions. x -axis, inversion age: the number of
 283 generations since the appearance of the inversion (in log-scale). y -axis : marginal fitness of the
 284 inversion relative to the same chromosomal segment on the X if it was in a male, noted W_{margX}
 285 (25). After fixation, this measures the sexually-antagonistic effect of nascent dosage
 286 compensation. The marginal fitness of the inversion relative to the same chromosomal segment
 287 among Y chromosomes not carrying the inversion, noted W_{margY} (25), yields undistinguishable
 288 results before the inversion fixes (note that W_{margY} cannot be computed after the inversion fixes as
 289 all Y chromosomes carry the inversion). In gray, individual trajectories, in black average values.
 290 **(A)** Inversions that are stabilized as first Y strata, collected over 10 evolutionary replicates after 1
 291 million generation. Their fixation date is indicated by a star at the bottom. **(B)** Inversions that are
 292 in the top 15 longest lived ones before a first stratum is stabilized, collected over 10 evolutionary
 293 replicates simulated over 1 million generations. Their extinction date is indicated by a gray disk at
 294 the bottom (and the average extinction date by the black disk). The time averaged fitness at time t

295 (in black) is computed over all inversions, counting their last achieved fitness if they are extinct at
296 t . The dashed line indicates value 1.

297

298 **Fig 4.** Steps involved in the evolution of a Y non-recombining stratum. The process involves four
299 steps, as explained in the text, and is briefly described by captions on the figure. Only the first
300 stratum is illustrated, but steps 1-4 are repeated for strata extending the non-recombining portion
301 of the Y until the whole chromosome is degenerated, silenced and dosage compensated.

302

303

304

305



Supplementary Materials for

Y recombination arrest and degeneration in the absence of sexual dimorphism

Thomas Lenormand*, Denis Roze

*Correspondence to: thomas.lenormand@cefe.cnrs.fr

This PDF file includes:

Materials and Methods

Supplementary Text 1 to 3

Figs. S1 to S8

Other supplementary material for this manuscript includes:

MDAR Reproducibility Checklist (PDF)

Materials and Methods

Genome

We use a simplified sex chromosome model similar to (27) that includes a sex determining locus at one end and $n_L = 500$ coding genes G . The expression of each gene is controlled by a cis-regulator C , and *two* trans-regulators T_m and T_f each of which is expressed either in males or in females. We assume that G and C sequences are uniformly spaced along the sex chromosome, with adjacent genes G recombining at a rate R_g (initially in both sexes), and each C regulator being closer to the G gene it regulates (at recombination rate R_c , $R_c < R_g$, fig. 1). Our simulations use $R_g = 0.0005$ (resulting in an initial overall map length of 25 cM), and $R_c = R_g/10$. Trans-regulators, such as transcription factors, are unlinked to their target genes, and influence expression on both homologs, whereas cis-regulators, such as enhancers, affect expression only of the gene carried on the same chromosome as themselves (26). Trans-regulators with sex-limited effects are necessary for dosage compensation to evolve (27). For simplicity, we assume that these T loci are autosomal.

Inversions and reversions

Contrary to model described in (27) where X-Y recombination is arbitrarily stopped during the simulations, we assume that inversions occur on the Y at a rate U_{inv} . We only consider inversions that include the sex locus (other inversions are not relevant to the topic investigated here as they are not confined to males). We denote z the non-recombining fraction of the Y (z is therefore comprised between 0 and 1). This variable is also used to measure the endpoint of each inversion on the map. When $z = 0$, X and Y chromosomes recombine freely, but otherwise X-Y recombination only occurs within the chromosomal segment $[z, 1]$. When $z = 1$, the X and Y do not recombine at all. When a new inversion occurs, its size is drawn as a uniform fraction of the non-recombining part of the Y. Specifically, on a Y where recombination is already stopped between 0 and z_i , the arrest of recombination will extend, after the new inversion $i+1$ to $z_{i+1} = z_i + (1 - z_i)u$, where u is a uniform deviate between 0 and 1. Finally, we assume that reversions can also occur, at a rate U_{rev} , removing the last inversion on the non-recombining part of the Y (we use $U_{inv} = U_{rev} = 10^{-5}$).

Regulatory traits

The effects of alleles at the cis- (C) and trans-regulators (T_m, T_f) are modelled as quantitative traits, with Gaussian mutations, denoted by c, t_m, t_f , respectively. These regulators control allele-specific expression as well as the overall level of expression Q of each gene. Mutations in cis and trans regulators are assumed to occur at rates U_c and U_t , respectively, and add a Gaussian deviate to allelic values for these traits ($c + dc \sim N(0, \sigma_c), t + dt \sim N(0, \sigma_t)$). We use $\sigma_c = \sigma_t = 0.2$ and $U_t = U_c/2 = 10^{-4}$. Negative trait values are counted as zero. These values are used to compute the total and allele-specific expression values for each coding gene G , as explained below. Note that we introduced two trans-regulators per gene, one expressed in males, and the other in females. We could instead assume a single trans-regulator determining two independent traits in males and females, which would be equivalent. The key is that sex-specific trans-regulation is possible, so that dosage-compensation can occur. Indeed with only cis-regulators, it would not be possible to maintain constant expression levels in both males and females if cis-regulators carried by X and Y start diverging. For instance, a stronger X cis-regulator could compensate for a weak Y cis-regulator in males, but it would cause overexpression if homozygous in a female. Dosage

compensation may also be achieved with either male or female trans-regulators, and both may not be required. However, we opt for a completely male-female symmetrical model with both, to ensure that the results do not depend on any male-female initial asymmetry, beyond the presence of the sex-determining locus.

Some genes may be unable to quickly evolve sex-specific regulation. If included, they would not diverge between X and Y, and they would therefore not contribute to recombination arrest. We do not investigate how chromosome-level mechanisms of dosage compensation evolve once gene-by-gene dosage compensation is achieved. The model does not require such global mechanism, but does not exclude that they could evolve later on.

Allele-specific expression and dominance

Arbitrarily denoting with a 1 or 2 subscript two alleles at a gene locus G , we assume that the fraction of the protein expressed from allele 1 is $\phi_{1,i} = c_{1,i}/(c_{1,i} + c_{2,i})$. This ratio measures the degree of allele-specific expression. With $\phi = 1/2$ (i.e. with equally strong cis-regulators on both homologs), alleles are co-expressed, while a departure from $1/2$ indicates that one allele is relatively more expressed than the other. The dominance of deleterious mutations occurring on the gene G depends on this allele-specific expression and is given by

$$h_{1,i} = \phi_{1,i}^{-\ln(h)/\ln(2)} \quad (\text{Eq. 1})$$

where h is a parameter measuring the dominance of the fitness effect of deleterious mutations in a heterozygote when both alleles are equally expressed (with $\phi_{1,i} = 1/2$, we have $h_{1,i} = h$). We set $h = 0.25$, corresponding to the average value observed across species (33). Fig. S6 shows results for other baseline dominance levels. Recombination suppression and degeneration is faster for higher level of baseline dominance. Indeed, as one would expect, Y silencing would likely play no role if deleterious mutations are already recessive to begin with.

Deleterious mutations on genes

Deleterious mutations occur on genes G at a rate U_G per gene. Their fitness effect s is drawn from an exponential distribution with mean s_{mean} . We use $s_{mean} = 0.05$. The effects of multiple mutations in the same gene are assumed to be additive, but can be cumulated only up to a maximum effect per gene, s_{max} (measuring the fitness effect of the gene knock-out). Their dominance depends on the strength of their associated cis-regulator. The more they are expressed (relative to the other allele), the larger their effective dominance, as explained above. The fitness effect resulting from the presence of deleterious mutations in gene i is

$$W_i^G = 1 - s_{2,i} - h_{1,i}(s_{1,i} - s_{2,i}) \quad (\text{Eq. 2})$$

where, by convention subscript 1 denotes the most deleterious allele of the two present in a given individual for that gene i .

Stabilizing selection on expression levels

We assume that the overall expression level of coding genes is under stabilizing selection with an optimum value Q_{opt} . In males, the total expression level Q_i equals $(c_{1,i} + c_{2,i})\bar{t}_{m,i}$, where $\bar{t}_{m,i}$ is the average strength of the *trans*-regulators expressed in males, which assumes that both *cis*- and *trans*-regulators are essential for proper expression (neither can be zero). Symmetrically, it is $(c_{1,i} + c_{2,i})\bar{t}_{f,i}$ in females. We assume that $\ln(Q_i)$ is under Gaussian stabilizing selection around $\ln(Q_{opt})$ (with $Q_{opt} = 2$). We use a log-scale to ensure that, irrespective of the intensity of stabilizing selection, the fitness effect of complete regulatory silencing ($Q_i = 0$) would be s_{max} , the maximum possible fitness effect of deleterious alleles on the coding gene, which we assume to be the same as the effect of a gene knock-out. Denoting by I the intensity of stabilizing selection on the expression level, the fitness resulting from the departure from optimal dosage W_i^Q is

$$W_i^Q = 1 - s_{max} \left(1 - e^{-I(\ln Q_i - \ln Q_{opt})^2} \right) \quad (\text{Eq. 3})$$

This function is equivalent to assuming that fold-changes in expression levels are under symmetric stabilizing selection, while selection on expression levels Q_i is asymmetric. Unless otherwise stated, we use $I = 0.1$.

Individual fitness

Individual fitness is contributed by two components: the fitness consequences of carrying deleterious mutations in the coding gene (whose dominance depends on allele-specific expression), W_i^G , and stabilizing selection on overall expression level W_i^Q . The overall fitness of an individual is computed as the product over all genes i of $W_i^G W_i^Q$.

Sexual dimorphism

It is important to note that our model does not include traits selected to be different in males and females. All the genes we consider have exactly the same expression optimum in males and females and we only consider deleterious mutations on genes that have the exact same effect in males and females. The presence of a sex-determining locus and sex-specific *trans*-regulators implies the existence of some sexually dimorphic traits coded somewhere in the genome, but these traits are absent from the simulations and therefore play no role in the results.

Life cycle and simulations

The different events of the life cycle occur in the following order: diploid selection, meiosis with recombination, mutation, and syngamy. Simulations are initialized with N_{pop} individuals, no polymorphism present, fully recombining sex-chromosomes and optimal gene expression levels (no deleterious allele, all c and t_m, t_f alleles fixed to 1, all Y chromosome with $z = 0$). After a burn-in phase, mutations producing inversions and reversions are turned on and we follow the dynamics of the system. Unless otherwise stated, we use $N_{pop} = 10^4$.

Measures

At regular time steps, we record average regulatory trait values in the population, mean fitness of males and females, average effect of deleterious mutations on the Y (s_Y), average dominance of

mutations on the Y (h_Y), P_{silent} the probability that $\phi_{Y,i}$ becomes close to zero (below 0.01) so that alleles on the Y become nearly entirely recessive (averaged over all genes), P_{dead} the probability that deleterious mutations on Y gene copy have reduced fitness by an amount s_{max} , indicating that the gene has entirely degenerated on the Y (averaged over all genes), and the average length of the non-recombining portion of the Y (z). We also record average “sex-switched” fitness, i.e. the average fitness of females computed replacing one of their X by a randomly drawn Y, and the average fitness of males computed replacing their Y by a randomly drawn X (relative to female and male average fitnesses, respectively).

We record all inversions occurring in the population (time of occurrence, start and end point, s_Y , h_Y , frequency, average regulatory traits, marginal fitnesses). We compute the marginal fitnesses of inversions (denoted W_{margY}) as the product of $W_i^G W_i^Q$ for all genes carried by this inversion (averaged over all Y chromosomes carrying this inversion), relative to the same product computed over all Y chromosomes not carrying this inversion. This quantity indicates whether inversions involve segments of the Y chromosome that present a higher or lower fitness effect compared to the equivalent non-inverted Y segment in the population. We also compute this marginal fitness relative to the average fitness effect of the same chromosomal segment sampled from an X chromosome, and placed in a male carrying the inversion (we use 1000 such samples to compute this value). Indeed, when a reversion occurs, followed by recombination events, it creates new Y chromosomes including (part of) this homologous X segment. We denote this quantity W_{margX} . It measures the average fitness of recombinant Y, relative to the actual Y, and therefore whether reversions could be selectively favored (if $W_{margX} < 1$). When $W_{margX} \gg 1$, it indicates that reversions and recombinant Y have a much lower fitness than current Y in the population, which is the signature of ‘stabilized inversions’.

Supplementary Text 1: Some limitations of current models

Recombination arrest

Current theory for Y recombination arrest is mostly based on the idea that the evolution of sexually dimorphic traits involves sexually antagonistic loci (SA-loci), where mutations occur that are beneficial in one sex, but deleterious in the other (6–9). If SA-loci are widespread, some would inevitably occur on the sex-chromosomes, which would then favor the evolution of tight linkage with the sex-determining locus. There is ample evidence for sexual conflict (34, 35), making this assumption plausible. However, the role of SA-loci in Y recombination arrest is not demonstrated empirically (16–19). This demonstration is inherently difficult to make because SA-loci are difficult to detect and because they can occur after the recombination arrest (18, 19, 31). Another issue is that models of evolutionary transitions between sex-determining systems frequently involve linkage between new sex-determining genes and SA alleles (36). Hence, whether this theory explains why sex chromosomes stop recombining is still entirely open.

One limitation of this hypothesis is that sex-linkage is only a particular way to solve intralocus sexual conflicts. Other resolutions based on regulatory evolution can occur as well (34, 35, 37), and the question is really whether these resolutions are more or less likely to evolve than sex-linkage (9, 18). (a) In the simplest case, the sexual conflict is on the level of expression of a given protein. In this case, sex-specific trans-regulators can easily adjust optimal expression in both sexes, solving the conflict. (b) In a second case, a particular protein modification is favourable in

one sex but needs not to be expressed in the other. This locus can evolve to be expressed only in the sex where it is favourable, by changing either cis or sex-specific trans-regulators, solving the conflict. (c) In a third case, two versions of a protein are favourable, each in one sex, but this protein needs to be expressed in both sexes. Here, the conflict can be solved by evolving an heterogeneous gene duplication with each copy becoming expressed in the sex where it is favourable (“subfunctionalization” scenario), or sex-specific alternative splicing. These genetic changes can evolve rapidly. Evolution of sex-specific expression is rapid (37, 38) and evidence for SA-polymorphism is limited compared to the signature of resolved expression differences (34). Even the most constrained case (c) may evolve rapidly, as demonstrated by the astonishing diversity of newly emerging heterogeneous gene duplication that have been documented at extremely short time scales in some cases (39).

Our model also suggests other limitations to the sexual antagonism theory. It is possible to propose a variant of our model where the initial fixation of inversions is promoted by the presence of a SA-locus. Whether this mechanism is more or less likely than the occurrence and spread of a “lucky” chromosomal segments with fewer deleterious mutations is a matter of relative mutation rates and effects of SA alleles and deleterious mutations. However, as we show (Fig. 3), such SA-loci are not necessary, as regulatory sex-antagonistic effects will inevitably emerge if early dosage compensation is allowed. Moreover, even if recombination suppression is caused by the presence of a SA locus, it may not be robust to the occurrence of reversions. The protection of the SA locus against reversion would last only to a point: until the fitness decay caused by degeneration becomes larger than this SA effect. This limitation is not operating with the regulatory model we propose as the protection grows with time (by the accumulation of dosage compensation and its growing “protecting” sex-antagonistic effect).

Degeneration

Models about sex-chromosome degeneration are based on selective interference once recombination is stopped. A limitation of these models is that they tend to be inefficient in large populations or on small non-recombining regions, and especially if only deleterious mutations are considered (11, 12, 40). Selective interference is inevitable in absence of recombination, but it may not be the only process causing degeneration. We previously showed that cis-regulatory divergence after recombination arrest could efficiently lead to Y silencing and degeneration, in absence of selective interference (27). This process works faster and can explain degeneration of very small non-recombining regions. Here too, there is ample evidence that selective interference occurs, but there is currently no evidence that it is the only mechanism at work, especially in species with large population sizes and small degenerated strata.

Dosage compensation

Dosage compensation (DC) tends to be considered separately from the process of degeneration and always assumed to evolve late. This claim is based on the causal chain of events: in current theory, DC results from degeneration, not the other way around (2, 4, 28, 30, 5–7, 10, 12–15). In fact, degeneration and DC are likely to be nearly simultaneous, at least for dosage sensitive genes, in any viable theory, as otherwise males would suffer an unbearable fitness decline relative to females during Y degeneration. However, models incorporating degeneration and DC tend to be currently missing, with some exceptions (11, 30). We recently showed that silencing and DC could theoretically occur slightly before and cause rather than result from degeneration (27).

2. Supplementary Text: Inversion and reversion assumptions

Inversions

The model assumes Y-carried mutations that stop recombination on a portion of the Y (and that this non-recombining region can be later extended or reduced). It helps to think that these mutations correspond to inversions, but there is really nothing ‘inversion-specific’ in the model. Chromosome collinearity is not altered in the code, and there is no specific assumptions determining how recombination proceeds on non-perfectly collinear chromosomes. The non-recombining region is determined using the variable z . Recombination does not occur between 0 and z and occurs between z and 1. “Inversion” mutations only alter this z value, but do not specify the exact mechanism by which this recombination arrest actually takes place. Hence, the results presented extend to other mechanisms of recombination arrest, as long as they involve a genetic change occurring on the Y (e.g. heterochromatinisation, hotspot presence etc.). Inversion fixing on the X could also cause recombination suppression. We expect the dynamics to be very similar in this case, although not exactly identical since, for instance, X and Y have different effective population sizes. In addition, the mutations considered here may not represent “real” inversions. Large real inversions do not necessarily create complete linkage across the inverted region, as double crossovers may be possible. Thus, the inversion that we consider would correspond to the non-recombining part of real inversions. This issue is however minimized by the fact that large inversions do not contribute to recombination arrest in our model (figs S2C and S3C). Note that “lucky” inversions can also fix on autosomes (41), but it does not lead to recombination suppression in this case, as they become homozygous in the population. Finally, although not considered here, it is also likely that degeneration may eventually lead to a loss of homology between the Y and X, contributing to the irreversibility of recombination suppression on stabilized inversion.

Reversion

We introduced reversions to make sure recombination arrest is not unescapable. Without this, inversions would inevitably occur and fix quickly, and no specific explanation would be required for Y recombination arrest. Such a model would be rather superficial, since the outcome would directly result from this hypothetical constraint. A similar mistake was made in early sex-chromosome models, which assumed that deleterious mutations could only occur on the Y and then concluded that the Y would inevitably degenerate (7). As we show, “lucky” inversions fix at a much shorter time-scale than the evolution of Y silencing (fig S7). Hence, in the absence of reversions, all fixed inversions would cause recombination arrest, irrespectively of their subsequent degeneration, and irrespectively of regulatory evolution. However, we also show that in absence of regulator evolution, reversions eventually eliminate all inversions, showing that recombination suppression is not selectively favored and cannot in fact be maintained (fig S1). Hence, not considering reversion makes the very strong assumption that recombination can only evolve one way. Models making this assumption will therefore inevitably fail to explain the long term maintenance of recombination suppression.

We considered another model for reversion, where each reversion was restoring recombination on all the Y, not just canceling the last inversion. This other model led to similar results suggesting that it does not matter much how precisely recombination can be restored, as long as it can. This other model involved however more complex dynamics. One of its specificity is to make

reversions less and less likely as the Y degenerates, by introducing a strong dependence between successive strata. In particular, a new inversion can be protected from reversions by the sex-antagonistic effects accumulated on other previous strata. This model also introduces coupled reversion-inversion dynamics, since a reversion can only persist if it is immediately followed by an inversion that stops recombination on the part of the Y that is already degenerated.

Other model of reversions could be imagined, but most other solutions would be computationally challenging to perform. For instance, reversion breakpoints could be randomly sampled anywhere in the non-recombining region. Such an approach would create each time a new inversion (with a new starting and ending point). Since we need to keep track of all inversions, this adds an important computational burden, without adding any interesting process. Another, very complicated way to allow for recombination restoration, would involve inversion on the X. If an inversion is present on the Y, but an inversion also occurs on the X, it could partially restore recombination, but this model would be extremely complex to run and analyze as it would require to keep track of X and Y collinearity and introduce complex additional assumptions to decide how recombination proceeds on non-perfectly collinear chromosomes. It would also require doing very complex bookkeeping of all possible chromosome orderings, including inversions that do not include the sex-determining locus.

3. Supplementary Text : Comparison to current theory and empirical observations

Comparisons of different theories can be made on several grounds that we can broadly categorize in terms of plausibility, parsimony and predictive power.

Plausibility

The model we propose involves a very general process with almost no specific assumption. All ingredients are basic genetic features of eukaryotes, such as deleterious mutations, *cis* and *trans*-regulators, inversions, stabilizing selection on dosage, and partial dominance. All these features have been extensively demonstrated. It also includes more specific assumptions, such as the occurrence of genetic mutations that can alter recombination rate on the Y, in a reversible manner. There is ample evidence for genetic variation in recombination rates, and this is an ingredient that must be present in any theory aiming at explaining Y recombination arrest. As discussed above, our way to represent reversion is certainly a drastic simplification, yet other models of Y recombination arrest do not even include the possibility of a restoration of recombination, especially once degeneration has started to occur. As we show in the model without regulatory evolution, if this possibility is included, degeneration would not occur on the long run, as reversions allow eliminating partially degenerated Y from the population. The presence of a sex-antagonistic polymorphic locus is also unlikely to be sufficient to counteract this process. The model we propose naturally generates ever increasing sex-antagonistic effects, which accumulate over all dosage sensitive genes, and naturally increase with the degree of degeneration. Hence, this mechanism is more likely to consistently favor the maintenance of recombination suppression than the occurrence of a handful of isolated loci with sex-antagonistic effects. Models allowing for dynamical accumulation of sex-antagonistic alleles might be possible, but they have not been worked out, and probably demand that an unrealistically large fraction of genes is involved in sexual dimorphism. Our model distinguish a proximal and ultimate cause for recombination arrest. The origin and maintenance of recombination suppression have distinct causes. Initially, there is no selection against recombination. However, a fortuitous but selective phenomenon (lucky

inversion spread and fixation) causes a selection pressure against recombination that was previously absent (the quick emergence of regulatory sex-antagonistic effects), and which protect this inversion from being subsequently lost. Distinguishing the problem of origin and the maintenance of a trait has often been considered in other contexts, notably the evolution of sex (42).

Parsimony

The theory we propose tends to be more parsimonious than current theory. It makes loci with sex-antagonistic effects superfluous, while all the ingredients present in our model are required in any global theory of sex chromosomes. For instance, deleterious mutations are necessary for degeneration, sex-specific regulatory changes are necessary for dosage compensation, and mutation altering recombination are necessary to explain recombination suppression. It is also parsimonious as it explains all the process of Y recombination arrest, degeneration and dosage compensation in a single model where all steps are integrated and work consistently within the same set of assumptions. In contrast, current theory is mostly made of series of models addressing each step separately, with different sets of assumptions.

Predictive power

Compared to current theory, our theory explains the same global pattern seen across many eukaryotes. Y or W chromosomes are often non-recombining, degenerated and at least partially dosage-compensated (2, 4–6, 12, 16, 17). Even if the causal explanation for each of these steps differs between current theory and the theory presented in this paper, observations distinguishing them may not be available yet. The established model (3–5) currently lacks decisive empirical support. In particular, there is no firm evidence that sex-antagonistic loci cause recombination suppression (16–19). However, absence of evidence is not evidence of absence, and the mechanism works in principle. There is ample evidence for the existence of sexually antagonistic selection (34, 35, 37), indicating that this explanation could work. In comparison, there is no indication that early dosage compensation on dosage-sensitive genes generates sex-antagonistic effects on young sex chromosome, but to our knowledge this has not been examined. There are indications of early evolution of dosage compensation or regulatory evolution in some species (20–24). This is an essential piece of information, consistent with our theory, but not proving that recombination arrest is caused by these modifications. Current theory does not predict a particular size for Y chromosome strata. It is however difficult to explain that small strata degenerate by selective interference if they contain only few genes. Our theory tends to indicate that rather small strata are involved in recombination arrest, although the exact size depends on details that we did not investigate here. For instance, if there is an important heterogeneity for dosage sensitivity among genes, strata size could be partly dictated by the chance localization of genes that are strongly dosage sensitive. In any case, our theory can certainly explain better why small strata can occur and degenerate. It may also easily explain cases where divergence appears nearly continuous along the X-Y chromosome pair, as if many small strata accumulated. Conversely, if several small strata occur on a short time interval, it may look as if a single large stratum evolved (see fig. S4 for the heterogeneity of simulation replicates). Several but not all strata result from inversions. Our model is presented using the term “inversion”, but as explained above there is nothing inversion-specific in our model, and other genetic modification suppressing recombination would work.

Empirically, the relative timing of degeneration and dosage compensation is not easily established. Dosage insensitive genes will almost never evolve dosage compensation, by definition, unless they are caught in a chromosomal level mechanism. Hence seeing that some genes are degenerated but not dosage compensated is not very informative. It may simply indicate that they are dosage insensitive, not that dosage compensation evolves after degeneration for dosage sensitive genes. Hence, observing that degeneration is more advanced than dosage compensation is not refuting our theory. Similarly, our model does not require complete early and full dosage compensation of all genes in a stratum. Only dosage sensitive genes are expected to evolve quick and early compensation, so observing partial or gene specific compensation is not refuting what we propose. The evolution of dosage compensation in neo-Y systems, after the fusion of an autosome to an already existing Y might be special cases, where a chromosomal-level dosage compensation mechanisms is just extended, without having to evolve from scratch. Such chromosomal-level mechanism would involve all genes (dosage sensitive or not) and therefore lead to a quite specific patterns. Other observations consistent with our theory include the absence of degenerated sex chromosomes in species where sex chromosomes are expressed in haploids, as in *Ectocarpus* algae (43) or *Ceratodon* moss (44). The presence of evolutionary strata on young mating-type chromosomes in species entirely lacking sexual dimorphism (45). could be explained by regulatory evolution if mating-type specific expression could evolve. However, recombination may evolve for other reasons in species with specific reproductive modes (46).

Our model predicts that Y recombination arrest and degeneration should be quicker in large populations, everything else being constant. This pattern may not hold, or may be saturating for larger population sizes than the ones we considered. This is open for investigation, but would require very long computation time. Empirically, we do not have clear indication of the effect of population sizes on the patterns of sex-chromosome evolution. There are many confounding factors that would need to be taken into account while attempting to test this prediction (shared ancestral sex determination systems across species, different ages of sex chromosomes, the initial recombination rate around the sex-determining locus, and the particular case of achiasmate species).

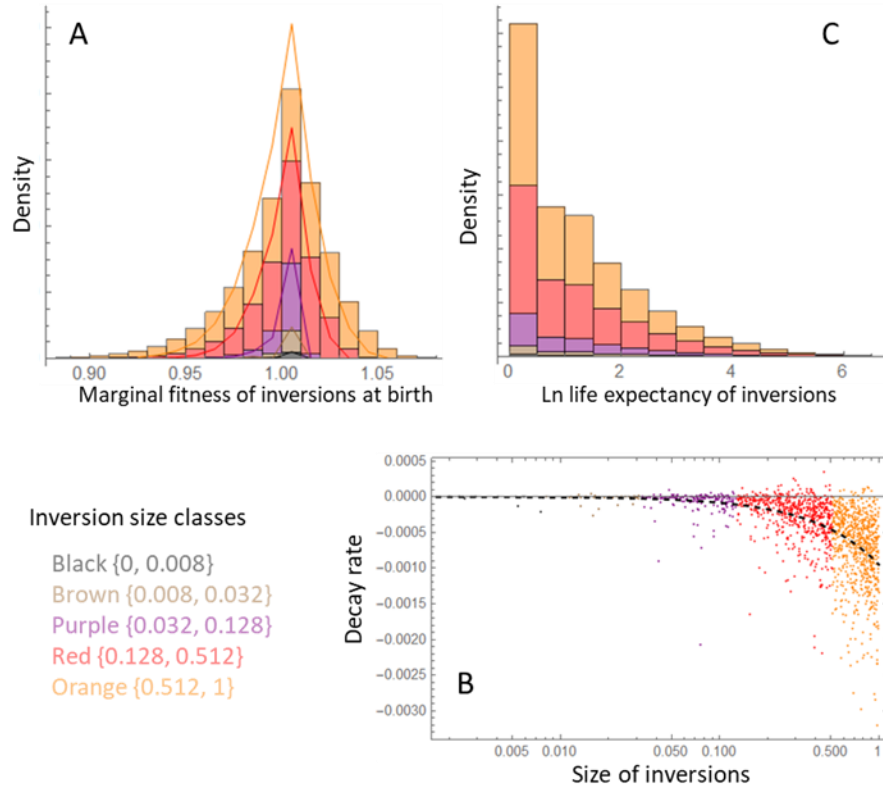


Fig S1. Characteristics of inversions in the absence of regulatory evolution. (A) Cumulated distributions of the marginal fitness of newly arising inversions for different size classes of inversions (the different colors, legend at bottom left; the whole chromosome has size 1 on this scale, the sex-determining locus being located at one end of the chromosome). When an inversion arises, the corresponding chromosomal segment carries by chance more or less deleterious mutations than average, resulting in a variance in marginal fitness. Lines: expectations computed assuming a Poisson distribution of the number of mutations per inversion, each with a fixed effect $s = s_{mean}$. (B) Marginal fitness decay rate per generation for inversions (y-axis), as a function of their size (x-axis, log scale). Inversions accumulate deleterious mutations because of selective interference. This decay rate is computed over the first 50 generations on the relative marginal fitness of newly arising inversions. Only inversions lasting at least 50 generations are therefore represented. The dashed line is the least square fit of a power law yielding $y = -0.00095x^{1.06}$, indicating that this decay rate varies approximately linearly with inversion size. Parameter values are described in methods ($N_{pop} = 10^4$, $I = 0.1$), but with no mutation on regulators ($U_c = U_t = 0$). (C) Distribution of the log time before inversions become extinct for the different size classes (in number of generations). These distributions have approximately the same mean values (about 3 generations).

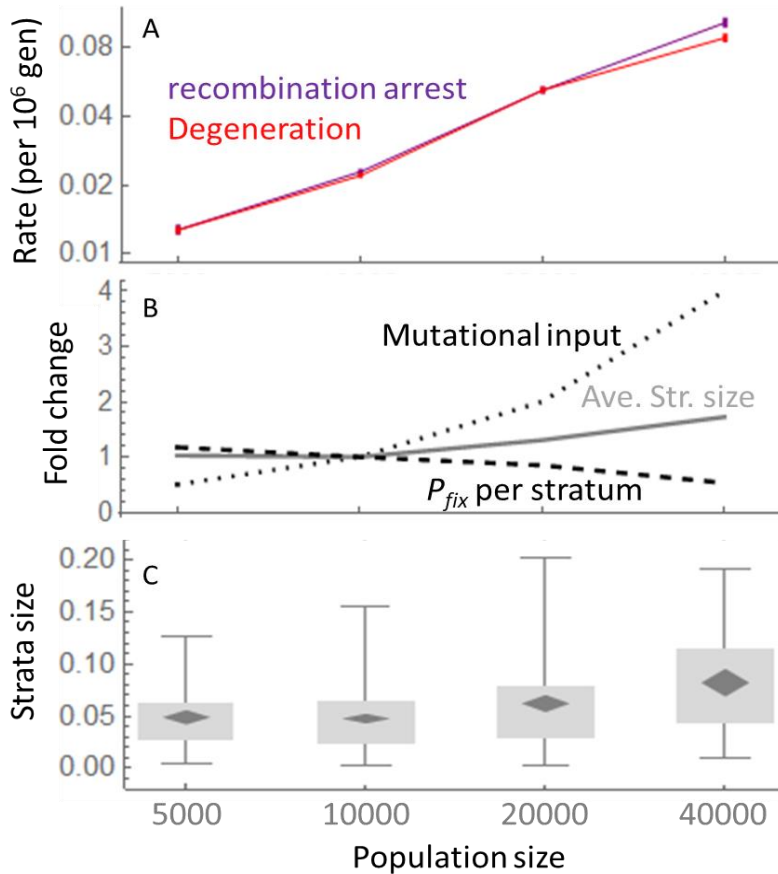


Fig S2. Effect of population size on Y recombination arrest and degeneration.

(A) Rate of recombination arrest (fraction of Y becoming non-recombining, purple) or degeneration (proportion of gene knocked-out, red) per million generations, for different population sizes (x -axes). These processes are approximately linear in time (excluding the initial approach to the steady state, during the ca first million generations, where these rates are higher, and the final approach to full Y recombination arrest, where rates necessarily saturate). Vertical bars indicate confidence intervals. (B) Contribution of the different factors to the variation in the rate of recombination arrest. The absolute number of inversions arising is proportional to population size (dotted curve, scaled relative to the value at $N = 10,000$). Stabilized inversions tend to be larger in larger populations (gray line, scaled relative to the value at $N = 10,000$). The probability that an inversion fixes and becomes stabilized decreases with population size (dashed line, scaled relative to the value at $N = 10,000$). The contribution of these three factors explain the difference in rates shown on panel A. (C) Distribution of stabilized inversions sizes for different population sizes. Black diamonds show means and confidence intervals; gray boxes show limits of 25% and 75% quantiles; whiskers show 5% and 95% quantiles. In panels A and B, lines connect data points for better visualization, but we only investigated the parameter values indicated on the x -axis.

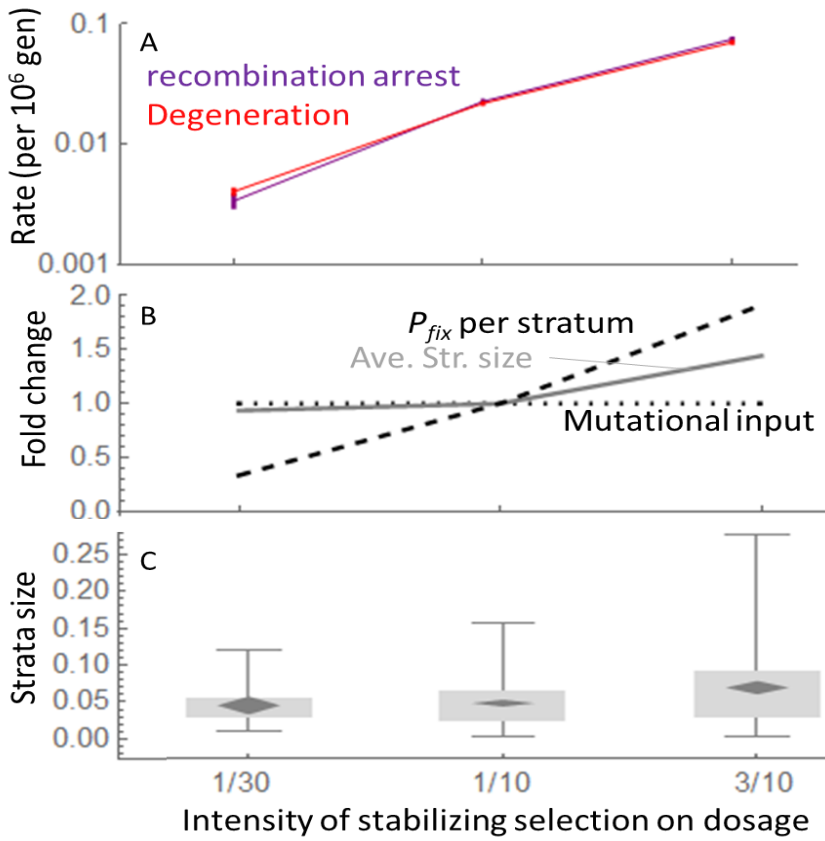


Fig S3. Effect of the intensity of stabilizing selection on gene expression levels, on Y recombination arrest and degeneration.

(A) Rate of recombination arrest (fraction of Y becoming non-recombining, purple) or degeneration (proportion of gene knocked-out, red) per million generations, for different intensities of stabilizing selection (I) on dosage (x -axes). These processes are approximately linear in time (excluding the initial approach to the steady state, during the ca first million generations, where these rates are higher, and the final approach to full Y recombination arrest, where rates necessarily saturate). Vertical bars indicate confidence intervals. (B) Contribution of the different factors to the variation in the rate of recombination arrest. The absolute number of inversions arising is identical for the different values of I and correspond to the value at $N = 10,000$ (dotted line). The probability that an inversion fixes and becomes stabilized increases with I (dashed line, scaled to the value at $I = 0.1$). Stronger stabilizing selection increases the sex-antagonistic fitness effect of nascent dosage compensation, which increase the chance that an inversion is stabilized and that large inversions escape reversion. The contribution of these three factors explain the difference in rates shown on panel A. (C) Distribution of stabilized inversions sizes for different values of I . Black diamonds show means and confidence intervals; gray boxes show limits of 25% and 75% quantiles; whiskers show 5% and 95% quantiles. In panels A and B, lines connect data points for better visualization, but we only investigated the parameter values indicated on the x -axis.

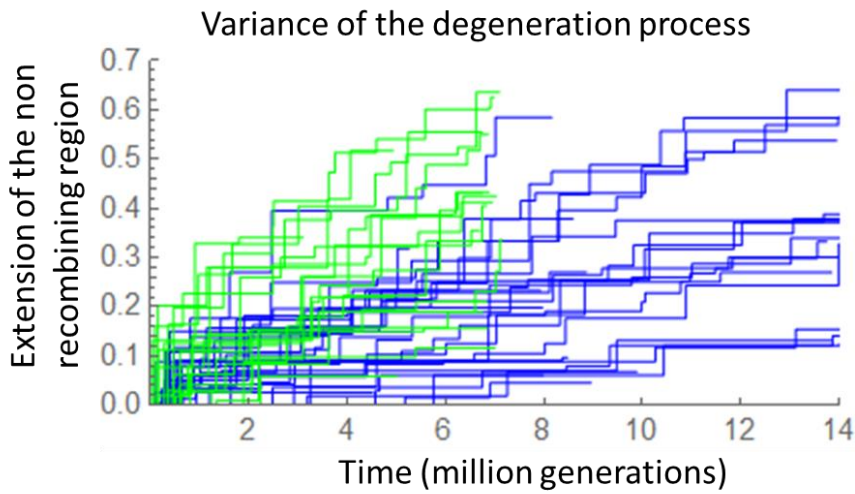


Fig S4. Stairplot representing the occurrence of stabilized strata on Y in different evolutionary replicates.

Each line corresponds to the stairplot illustrated in black in Fig. 2A for one replicate. Parameters as in Fig. 2, except for population size: blue $N_{pop}=10,000$, green $N_{pop}=20,000$. Note that runs with $N_{pop} = 10,000$ and $20,000$ were stopped at 7 million generations and 14 million, respectively. In both cases, some runs were slower and were interrupted before reaching this limit, because of computation time limits. The figure shows that the process of recombination arrest is highly stochastic.

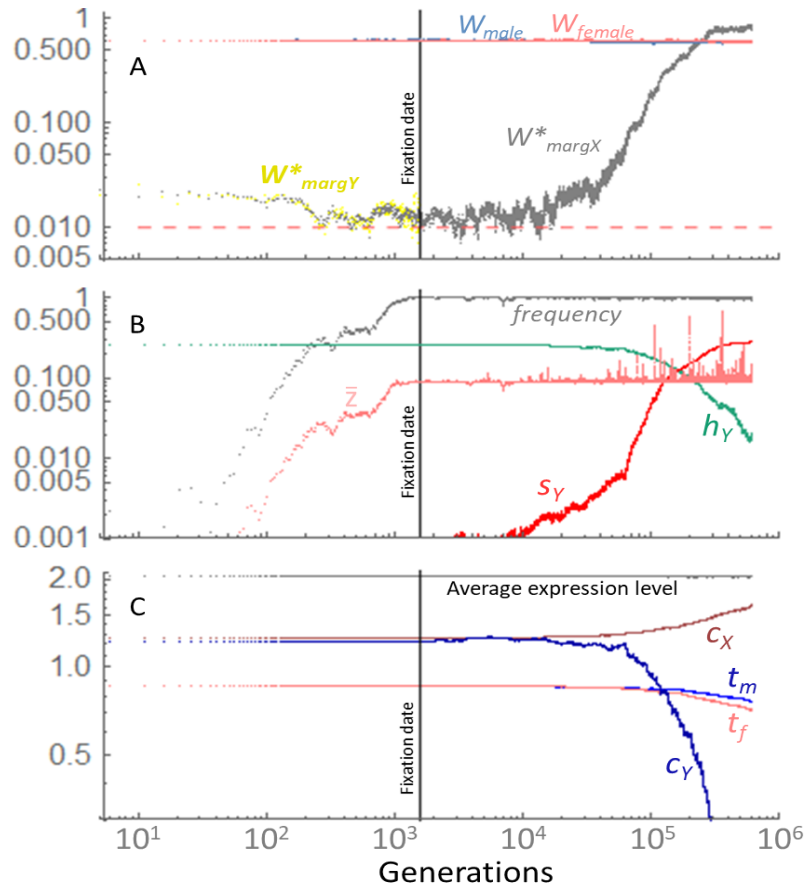


Fig S5. Details of the fixation and stabilization of a first stratum.

(A) *x*-axis, inversion age: the number of generations since the appearance of the inversion (in log-scale). Gray : marginal fitness of the inversion relative to the same chromosomal segment on the X if it was in a male (W_{margX} , see methods). After fixation, W_{margX} measures the sex-antagonistic effect of nascent dosage compensation. Yellow: marginal fitness of the inversion relative to the same chromosomal segment among Y-chromosomes not carrying the inversion (W_{margY} , see methods). Note that W_{margY} cannot be computed after the inversion fixes as all Y-chromosomes carry the inversion. Both these fitness values are represented minus 0.99 (and noted with a *) to allow for a better visualization on the *y*-axis log-scale. Consequently, the red dashed line at 0.01 represents a marginal relative fitness of 1. Average fitness of males and females in the population are also indicated in blue and pink. (B) Gray: Frequency of the inversion. Pink average fraction of the non-recombining Y in the population (\bar{z}). Green: average dominance of deleterious mutations on the inversion (h_Y). Red: average deleterious effect of mutations among genes on the inversion (s_Y). (C) Regulatory trait variation. Dark blue: average cis-regulatory trait on the inversion. Brown: average cis-regulatory trait on the corresponding X segment. Blue: average trans-regulatory trait associated to genes on the inversion. Pink: average trans-regulatory trait associated to genes on the corresponding X segment. Gray: average total gene expression per genes (undistinguishable in males or females) for genes present on the inversion. Note that between 10^3 and $\sim 5 \cdot 10^4$ generations, there is enough sex-antagonistic effect of nascent dosage compensation to protect the inversion from reversion, as seen on panel A with W_{margX} , while there is still almost undetectable X-Y cis-divergence and male –female trans-regulatory divergence. On all panels, the vertical bar indicates the date at which the inversion fixes for the first time (because of the occurrence of reversions the frequency slightly departs from one afterwards).

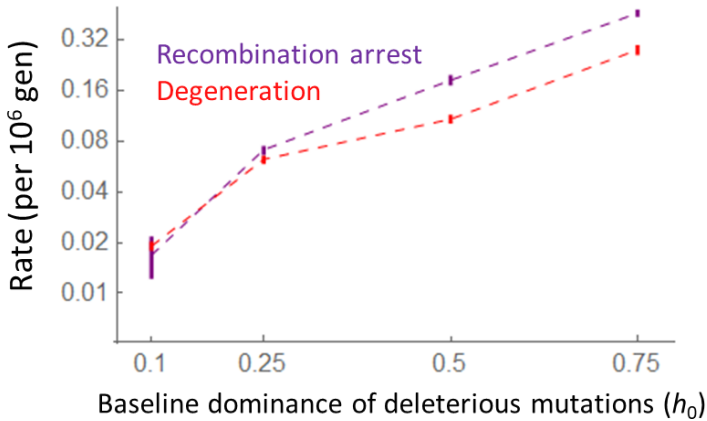


Fig S6. Effect of the baseline dominance level on the rates of recombination arrest and degeneration.

Results shown on other figures correspond to $h = 0.25$, i.e. to partially recessive deleterious mutations, which is the consensus empirical estimate across several eukaryotes (33). The figure shows rates of recombination arrest and degeneration for other values of h (x-axis). Lines connect data points for better visualization, but we only investigated the parameter values indicated on the x-axis. Average rates are computed over 20 replicates and over the first million generations (these initial rates are inflated by a factor ca 2-3 compared to the steady state rates shown on figures S2 and S3). Vertical bars indicate confidence intervals. Rates of recombination arrest and degeneration are higher for higher h . This rate variation is approximately linear around $h = 0.25$, which suggests that variance in baseline dominance will have limited effect on average rates of recombination arrest and degeneration (note that the y-axis is in log-scale). Other parameters are $N_{pop} = 10,000$, $I = 0.1$, as in figs. 2, 3.

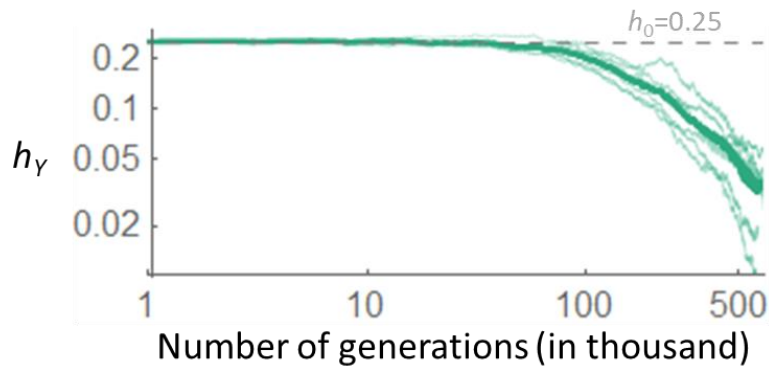


Fig S7. Detailed variation of h_Y for stabilized inversions.

This figure shows how h_Y varies through time, on average for all the stabilized inversions shown on Fig 3A. The x-axis, in log-scale, is the number of generation (in thousands) since the inversions appeared by mutation. The y-axis (in log-scale) indicates the average realized dominance of deleterious mutations on the stabilized inversions. The mean fixation time of inversions that reach fixation is ca 2000 generations (range 600-6000), i.e. well before h_Y starts to change. Light green: individual replicates; Dark green: average across replicates. The gray dashed line indicates the baseline dominance level $h = 0.25$.

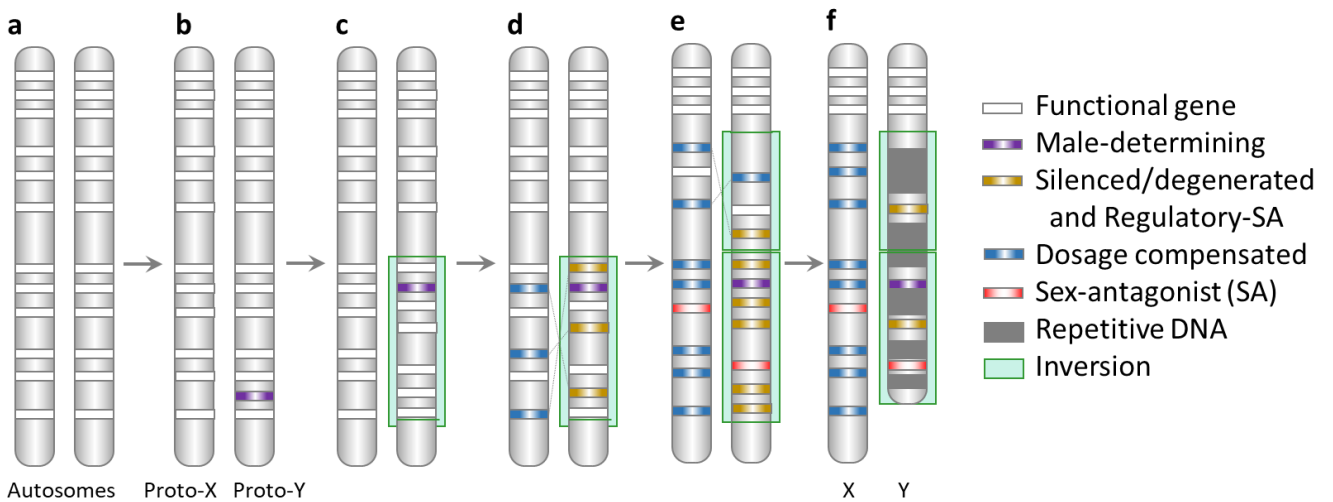


Fig S8. Overview of Y evolution through regulatory evolution.

The figure summarizes the different steps occurring during Y evolution by regulatory evolution. It can be contrasted with Fig. 1 in (5) representing current theory. (a) and (b) A pair of autosomes acquire a sex-determining locus (purple). The theory presented in this paper starts in (b). (c) A lucky inversion carrying fewer or milder deleterious mutations than average selectively fixes in the population (green). (d) Cis- and trans-regulatory divergence causes regulatory sex-antagonistic effects on dosage sensitive genes, which stabilizes the inversion on the long term. Y genes tend to be silenced (yellow) and accumulate deleterious mutation (degeneration by regulatory evolution), while their X copy already show dosage compensation (blue). (e) The process repeats itself with another sex-linked inversion, creating another stratum. Some sex-antagonistic alleles can occur and be maintained on these sex-linked regions, but they were not involved in recombination arrest (red). (f) The absence of recombination leads to the accumulation of repetitive DNA and/or structural changes (deletions).

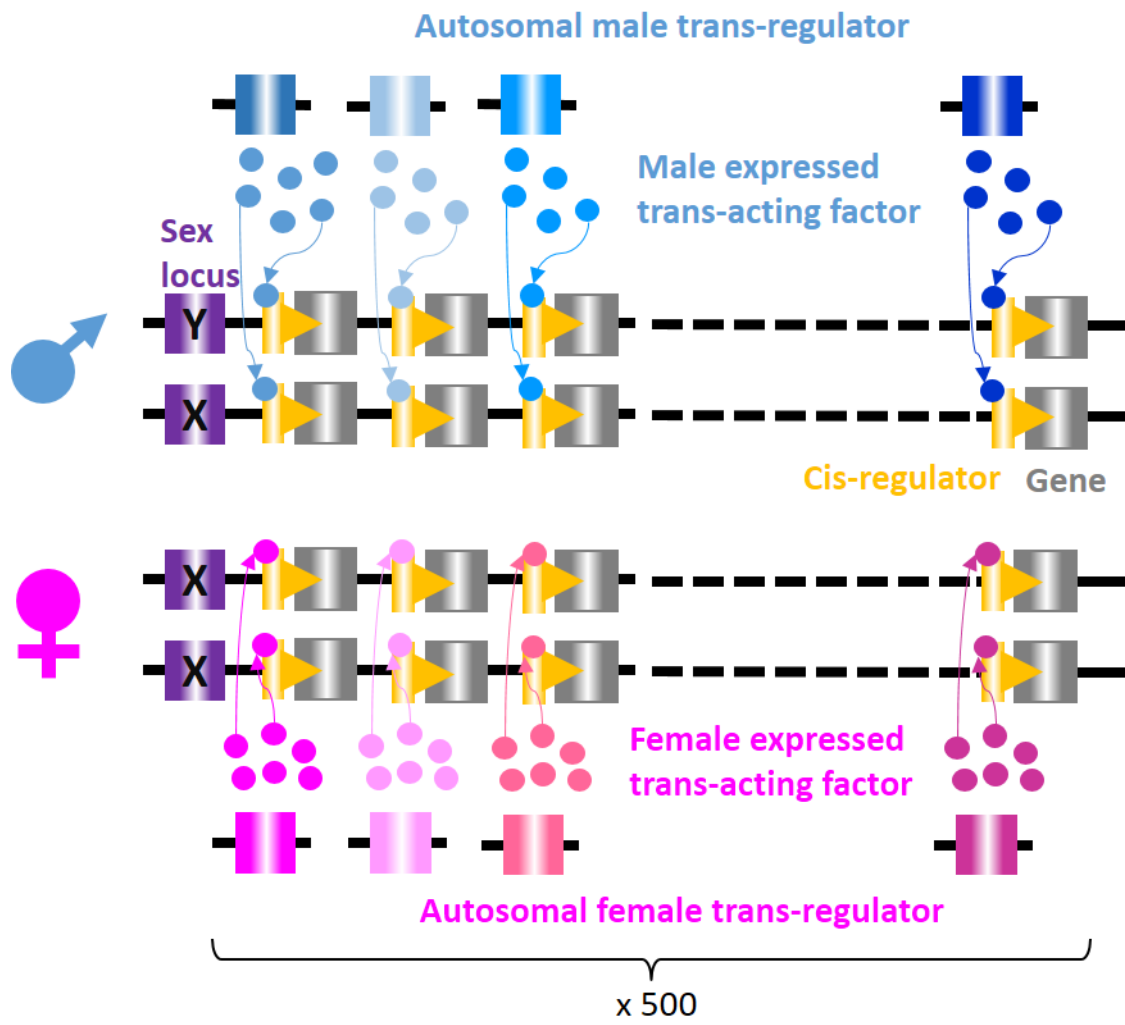


Fig 1

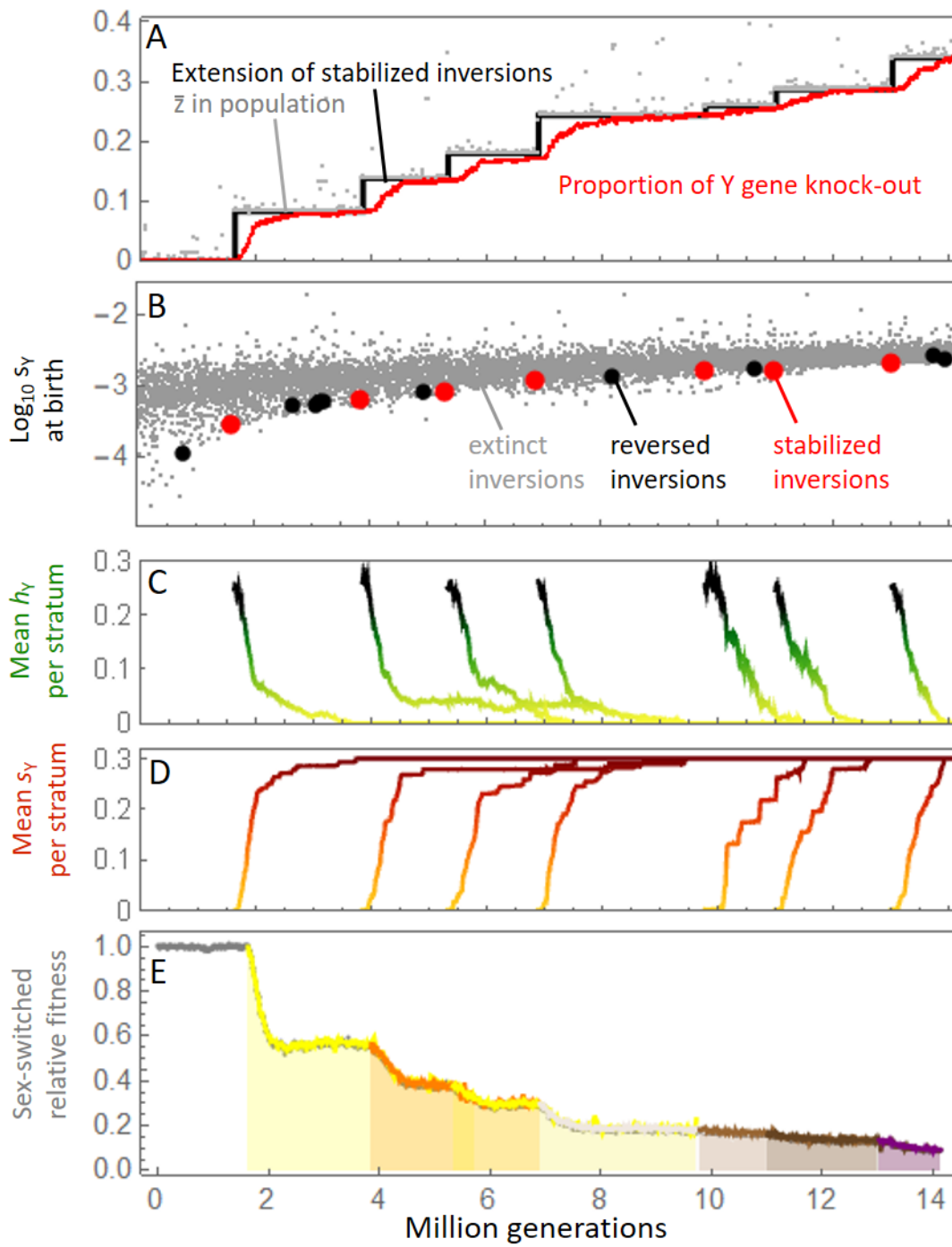


Fig 2

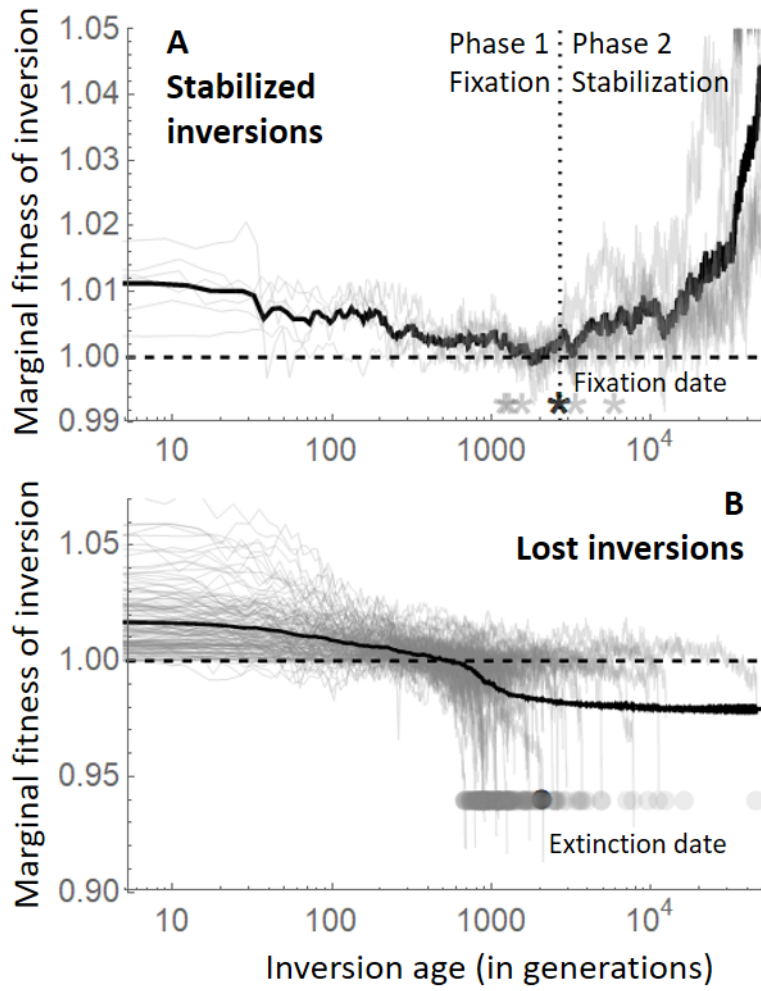


Fig 3

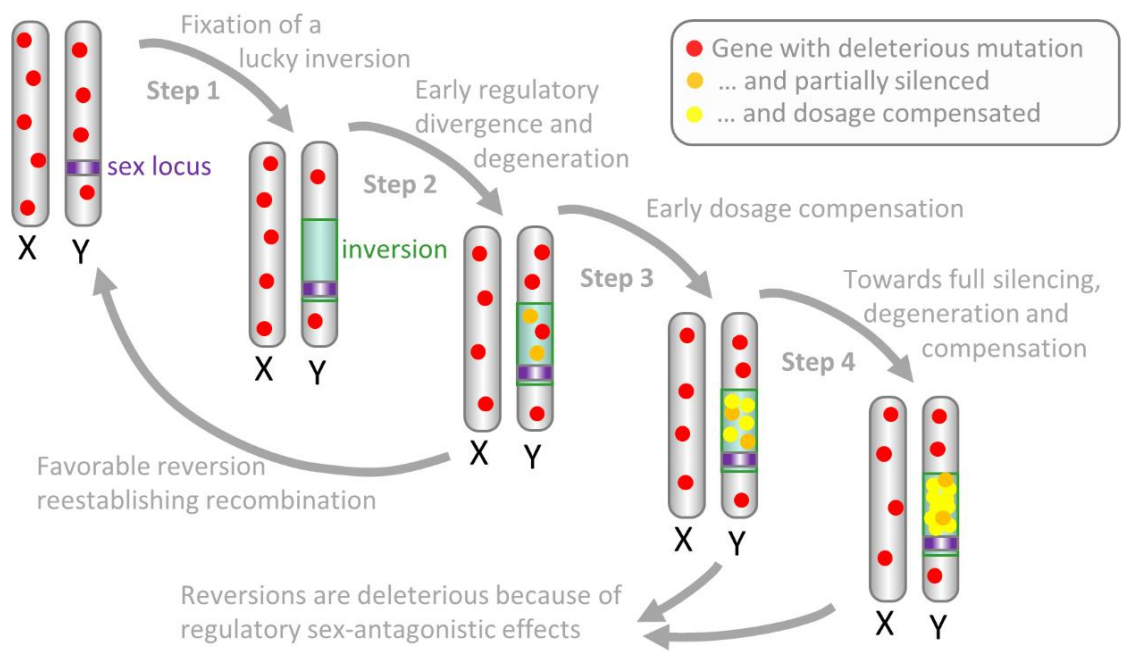


Fig 4