Article

Chromosomal inversion polymorphisms shape human brain morphology

Hao Wang,[1] Carolina Makowski,[1] Yanxiao Zhang,[2,3,4] Anna Qi,[1] Tobias Kaufmann,[5,6] Olav B. Smeland,[6] Mark Fiecas,[7] Jian Yang,[3,4] Peter M. Visscher,[8] and Chi-Hua Chen[1,9,*]

[1]Center for Multimodal Imaging and Genetics, University of California San Diego, La Jolla, CA 92093, USA
[2]Ludwig Institute for Cancer Research, La Jolla, CA 92093, USA
[3]School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China
[4]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310024, China
[5]Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University of Tübingen, 72076 Tübingen, Germany
[6]Norwegian Centre for Mental Disorders Research, Oslo University Hospital and University of Oslo, 0450 Oslo, Norway
[7]Division of Biostatistics, University of Minnesota School of Public Health, Minneapolis, MN 55455, USA
[8]Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia
[9]Lead contact
*Correspondence: chc101@ucsd.edu
https://doi.org/10.1016/j.celrep.2023.112896

SUMMARY

The impact of chromosomal inversions on human brain morphology remains underexplored. We studied 35 common inversions classified from genotypes of 33,018 adults with European ancestry. The inversions at 2p22.3, 16p11.2, and 17q21.31 reach genome-wide significance, followed by 8p23.1 and 6p21.33, in their association with cortical and subcortical morphology. The 17q21.31, 8p23.1, and 16p11.2 regions comprise the *LRRC37*, *OR7E*, and *NPIP* duplicated gene families. We find the 17q21.31 *MAPT* inversion region, known for harboring neurological risk, to be the most salient locus among common variants for shaping and patterning the cortex. Overall, we observe the inverted orientations decreasing brain size, with the exception that the 2p22.3 inversion is associated with increased subcortical volume and the 8p23.1 inversion is associated with increased motor cortex. These significant inversions are in the genomic hotspots of neuropsychiatric loci. Our findings are generalizable to 3,472 children and demonstrate inversions as essential genetic variation to understand human brain phenotypes.

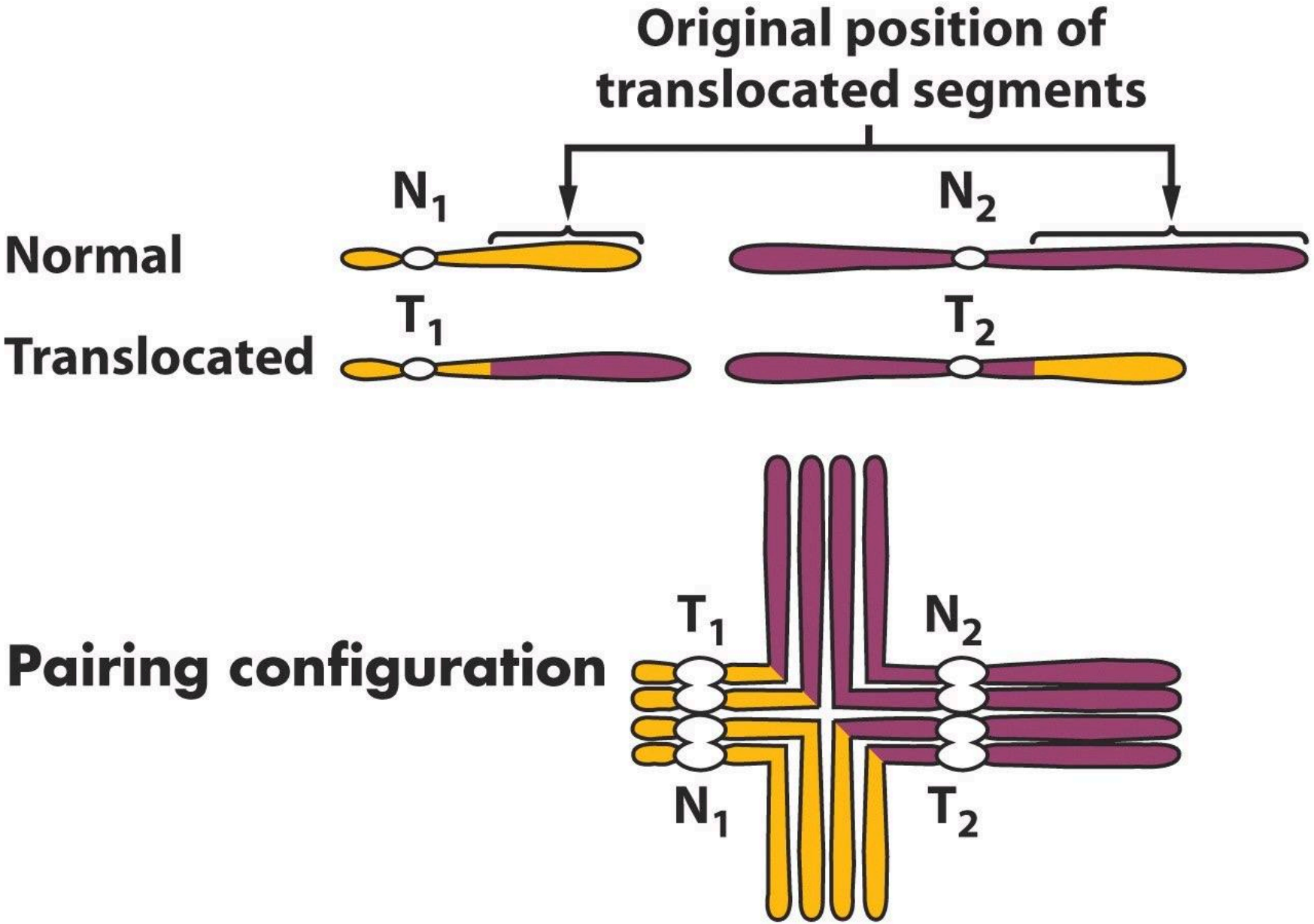
INTRODUCTION

Polymorphic chromosomal inversions are common structural variations in which a chromosomal segment is in the reversed orientation.[1,2] Recent work identified 729 inversions in humans and an average of 117–156 inversions per human genome.[3–5] The inversions included in this study span from less than 1 Kbps to 4 Mbps, and their reported frequencies may range

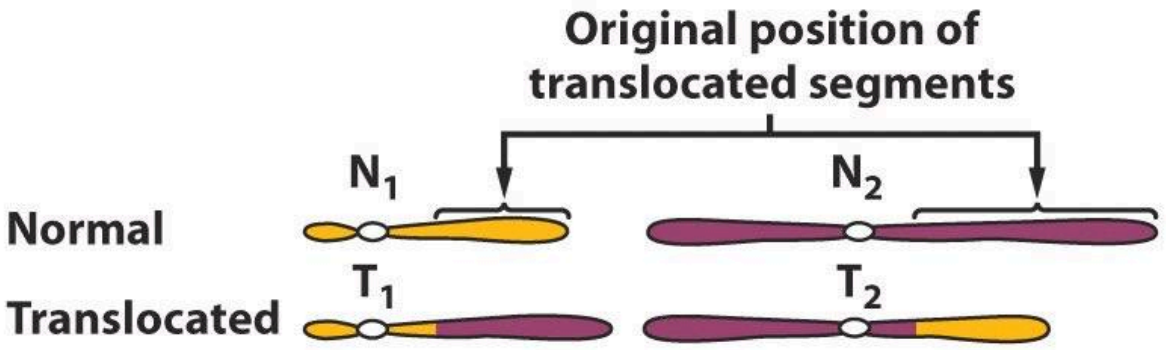
In this study, we use the term "inversion" to refer to an inversion region, an inverted variant/allele (haploid), or an inverted genotype (diploid).

Large inversions are often flanked by segmental duplications (SDs) (repeats with ≥90% sequence identity) or mobile elements (e.g., *LINE1* or *Alu*) at inversion breakpoints.[3,10,11] The hominid lineage has experienced a burst of SDs, which constitute ~7% of the human genome. SDs contribute disproportion-

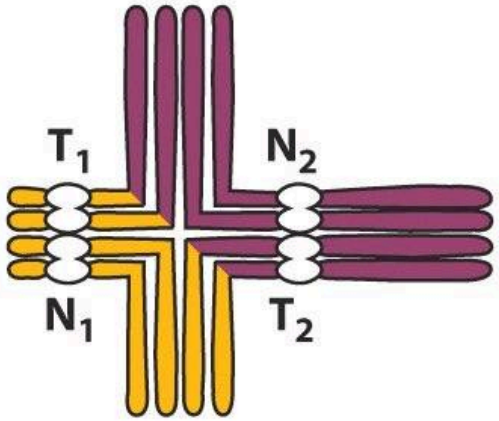
Translocation heterozygote



Translocation heterozygote



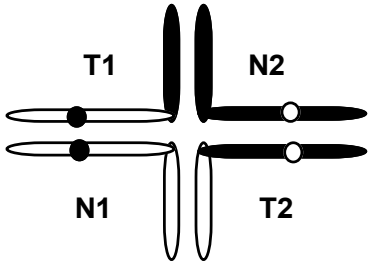
Pairing configuration



Two types of segregations:

Adjacent-1		Final meiotic products	
Up	$T_1 + N_2$	Duplication of purple, deletion of orange translocated segment	Often inviable
Down	$N_1 + T_2$	Duplication of orange, deletion of purple translocated segment	
Alternate			
Up	$T_1 + T_2$	Translocation genotype	Both complete and viable
Down	$N_1 + N_2$	Normal	

In a cross between translocation heterozygotes, only a small fraction of the zygotes are viable. In plants most translocation heterozygotes are viable, while in Drosophila they are lethal. There is a simple explanation for this. In plants translocations must be transmitted through a haploid multicellular gametophyte to be transmitted at all. Those that would be lethal in a diploid homozygote are also lethal to the haploid gametophyte. In flies, which have a very short haploid stage that is unicellular, the gametes are fertile and viable even when the chromosome constitution is grossly abnormal. Thus they can be transmitted even if they would be lethal in the homozygous state.



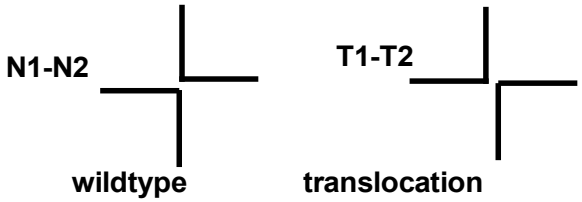
A cross between translocation heterozygotes, <u>in animals</u>, = 6/16.

	T1-N2	T2-N1	T1-T2	N1-N2
T1-N2	–	+	–	–
T2-N1	+	–	–	–
T1-T2	–	–	+	+
N1-N2	–	–	+	+

Plants 1: 2: 1 wildtype: het: transl. homozygote

<u>Adjacent-1 segregation</u>
Dupication + deficiency

<u>Alternate segregation</u>
Viable



<u>Adjacent-2 segregation</u>
Homologous centromeres go to the same pole. These are rare and inviable.

In plants the pattern simplifies because unbalanced translocation gametes are inviable in the gametophyte stage. Thus certainzygote genotypes are not produced.

One common way in which chromosome rearrangements contribute to human disease occurs when genes that promote cancer come under the control of powerful promoters that drive their expression in specific cells of the immune system. This can lead to leukemia and lymphoma, forms of cancer

Deletions
Inversions
Translocations

move genes to new locations

This may alter the protein product or the pattern of gene expression.

An example from human biology is the Philadelphia chromosome

9: 22 translocation → Leukemia Inappropriate activation of a gene, c-abl

Burkitt's lymphoma 8: 14 translocation → Inappropriate activation of c-myc

Changes in chromosome number

Ploidy: **the number of chromosome sets**

Monoploid: **Organism with one chromosome set (described for organisms from essentially diploid taxa).**

Euploid: **An organism that contains some multiple of the basic chromosome set.**

Polyploid: **An organism that contains more than 2 chromosome sets.**

The basic chromosome number, X. This is the number of distinct chromosomes that an organism contains. In a normal diploid x = n, the haploid chromosome number. However, some organisms such as wheat are hexaploid. Wheat has 42 chromosomes, but the basic set is 42/6 = 7 = X. In contrast, n = 42/2 = 21.

Haploid number, n. n is the number of chromosomes in gametes.

Autopolyploids: **Polyploid organisms created by chromosome duplication within a species.**

Allopolyploids: **polyploids created by hybridization between different (but usually related) species. In this situation the related chromosomes are usually described as being homeologous, rather than homologous.**

Polyploidy Generalities

de novo

- **rare in most animal species,**
- **known in lizards, fish and amphibians,**
- **fairly common in plants,**
- **odd numbers of ploidy are not usually maintained,**
  3n, 5n, etc
  rarely found in organisms that rely on sexual propagation.

Monoploidy

…a haploid of a diploid is monoploid,

…has one chromosome set

- **male wasps, bees and ants have only 1 haploid genome,**
  - **males develop from unfertilized eggs,**
    - **gametes are formed by mitosis.**

Monoploids.

In some organisms, such as bees and wasps, the males are monoploid. They develop from an unfertilized egg. Being monoploid should present a problem during meiosis since the individual chromosomes have no pairing partners. Males in these species get around this problem by using mitosis rather than meiosis during gamete production.

If a monoploid cell (or as we will get to in a moment, any cell with an odd number of homologs that must be dealt with at meiosis) does undergo meiosis there is a problem. The movement towards one pole or the other is essentially random and independent for each group of homologs. In a monoploid some cells get none of each chromosome while others get one. Similarly, in a triploid some cells would get one homolog while other cells would get two. In either situation this creates a genetic imbalance in the gamete, and thus almost certainly a genetic imbalance in the fertilized zygote.

$$\left(\frac{1}{2}\right)^{x-1}$$

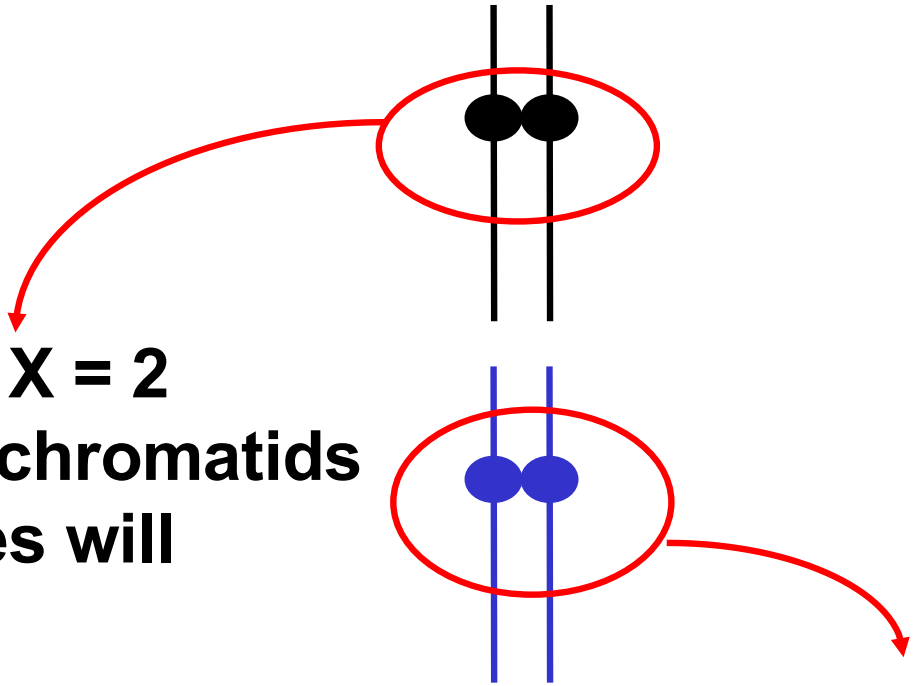
X = the number of chromosomes (in monoploids), or X, the basic chromosome set number, in polyploids such as triploids.

Triploids have problems during meiosis that lead to sterility

Triploids arise from a cross between a tetraploid and a diploid. Diploid and haploid gametes fuse to produce a triploid embryo. The triploid offspring can develop. However, during meiosis a similar problem occurs to that present in a monoploid (figure below). It is very unlikely [p = (1/2) x-1] that a similar number of the different chromosomes will all go to the same gamete during meiosis. This leads to genome imbalance.

Take as an example the banana. It has 11 chromosomes and is a triploid (3x = 33 chromosomes). This means that at meiosis there is a 1/1024 chance of producing a gamete with a full haploid or diploid complement of chromosomes. The frequency of two such gametes getting together to produce a viable seed is extremely low. Triploidy is used to create seedless versions of fruits such as bananas and watermelon. So, you might ask, if they are seedless, how do they propagate. This is done artificially, using cuttings, grafts or bulbs. Essentially, a piece of the original triploid is used to start a new plant.

In meiosis 1 in a monoploid the duplicated chromatids go one way or the other. They have no pairing partner. In the example, X = 2 There is a 50% chance pairs of chromatids from two different chromosomes will go to the same side.



Monoploids.

In some organisms, such as bees and wasps, the males are monoploid. They develop from an unfertilized egg. Being monoploid should present a problem during meiosis since the individual chromosomes have no pairing partners. Males in these species get around this problem by using mitosis rather than meiosis during gamete production.

If a monoploid cell (or as we will get to in a moment, any cell with an odd number of homologs that must be dealt with at meiosis) does undergo meiosis there is a problem. The movement towards one pole or the other is essentially random and independent for each group of homologs. In a monoploid some cells get none of each chromosome while others get one. Similarly, in a triploid some cells would get one homolog while other cells would get two. In either situation this creates a genetic imbalance in the gamete, and thus almost certainly a genetic imbalance in the fertilized zygote.

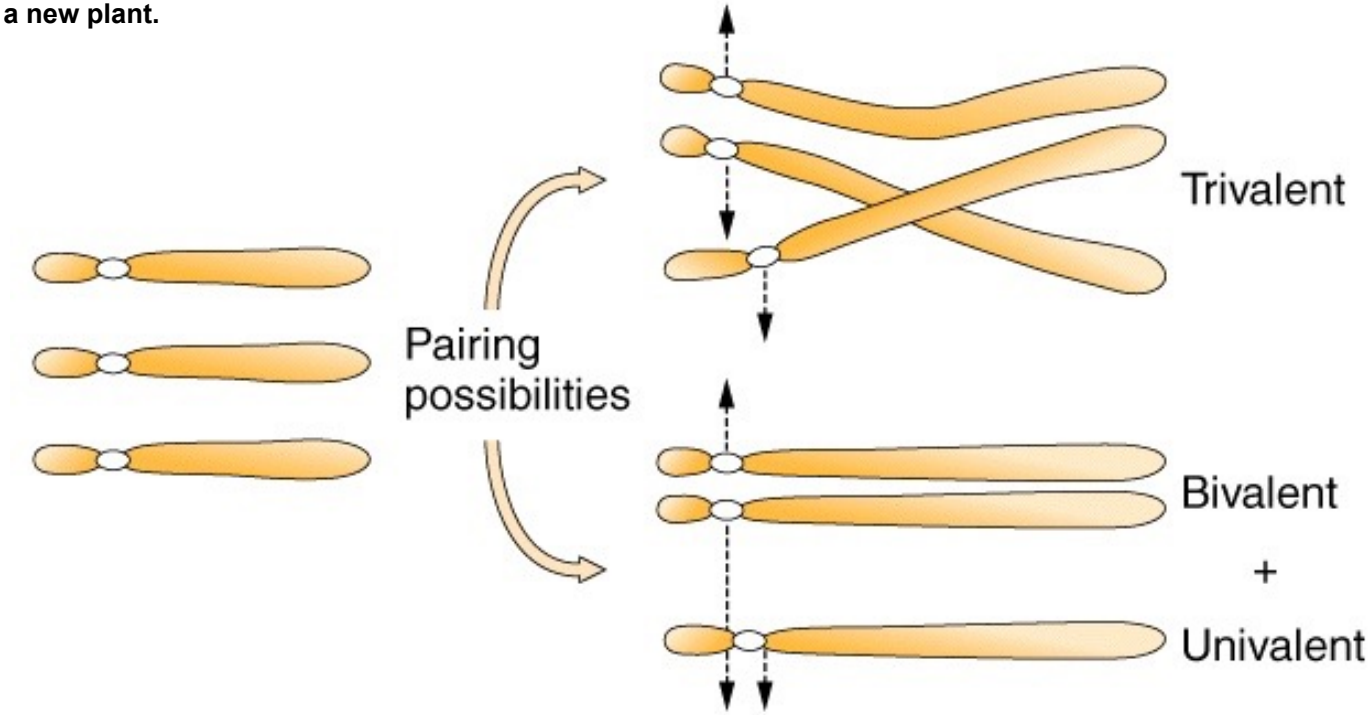
$$\left(\frac{1}{2}\right)^{x-1}$$

X = the number of chromosomes (in monoploids), or X, the basic chromosome set number, in polyploids such as triploids.

Triploids have problems during meiosis that lead to sterility

Triploids arise from a cross between a tetraploid and a diploid. Diploid and haploid gametes fuse to produce a triploid embryo. The triploid offspring can develop. However, during meiosis a similar problem occurs to that present in a monoploid (figure below). It is very unlikely [p = (1/2) x-1] that a similar number of the different chromosomes will all go to the same gamete during meiosis. This leads to genome imbalance.

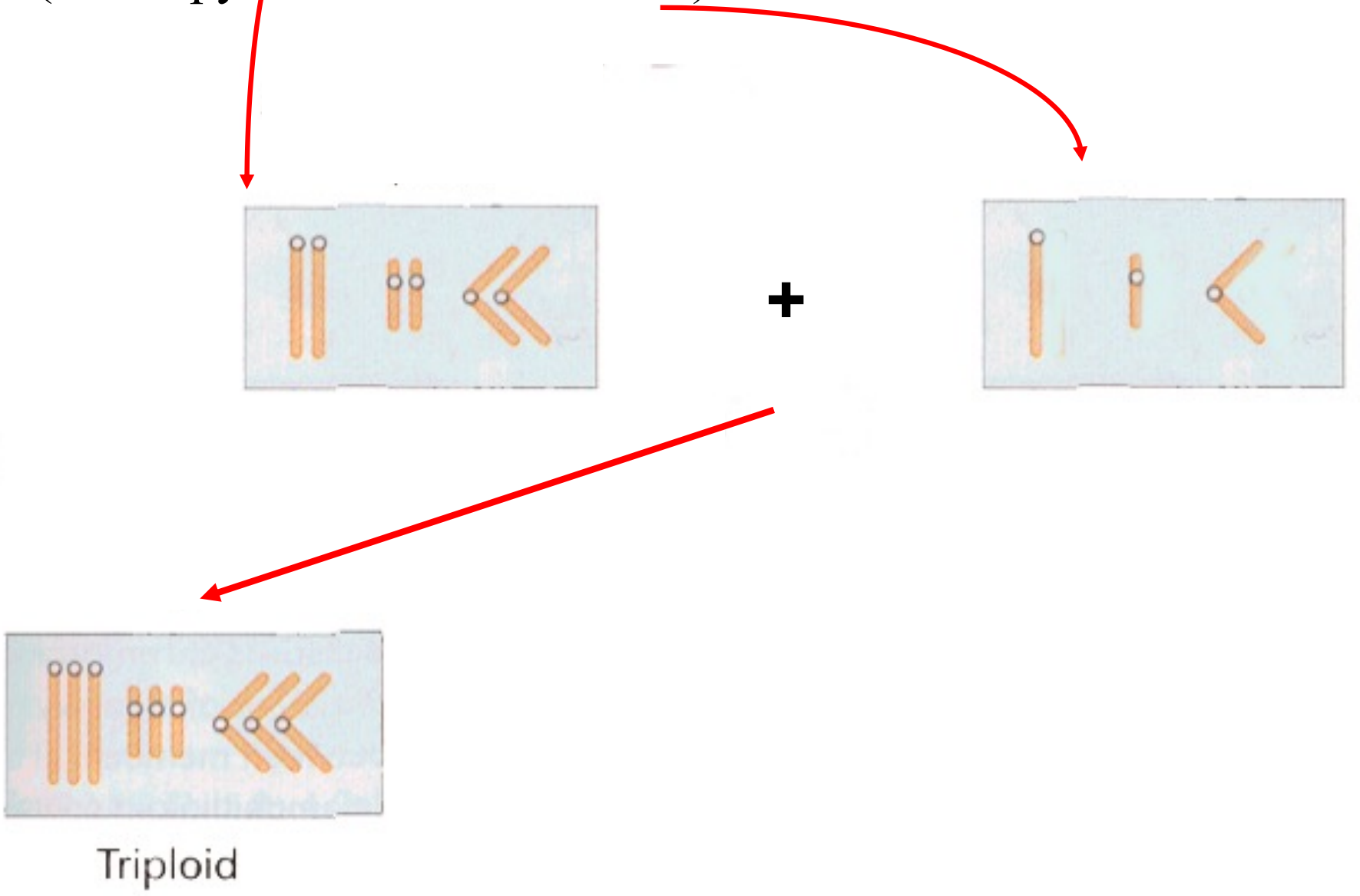
Take as an example the banana. It has 11 chromosomes and is a triploid (3x = 33 chromosomes). This means that at meiosis there is a 1/1024 chance of producing a gamete with a full haploid or diploid complement of chromosomes. The frequency of two such gametes getting together to produce a viable seed is extremely low. Triploidy is used to create seedless versions of fruits such as bananas and watermelon. So, you might ask, if they are seedless, how do they propagate. This is done artificially, using cuttings, grafts or bulbs. Essentially, a piece of the original triploid is used to start a new plant.



Triploids are sterile.

Bananas, seedless watermelons are seedless because they are triploid

Triploids are generated by combining gametes from a tetraploid (2 copies of each chromosome) with those of a diploid (one copy of each chromosome).



Production of Trout Offspring from Triploid Salmon Parents

Tomoyuki Okutsu,[1] Shinya Shikina,[1] Megumi Kanno,[1] Yutaka Takeuchi,[1] Goro Yoshizaki[1,2]*

In recent decades, the number of salmonid species has declined markedly, and several species have become extinct or endangered. Because cryopreservation of fish eggs is difficult due to their large size and high fat content, we investigated the potential of surrogate broodstock technologies as a new method of genetic resource preservation for fish. Surrogate broodstock technologies involve the transplantation of primordial germ cells (PGCs) (*1*) or spermatogonia (*2*) from a target fish species into a related species for which rearing techniques are well developed. In doing so, the recipient species can produce sperm and eggs of the target species (*3*). Furthermore, because PGCs and spermatogonia are sufficiently small for cryopreservation, animals can be generated via the transplantation of thawed PGCs or spermatogonia into recipients, even if the target species becomes extinct. In prior work, we demplanting spermatogonia into sterile xenogeneic recipients.

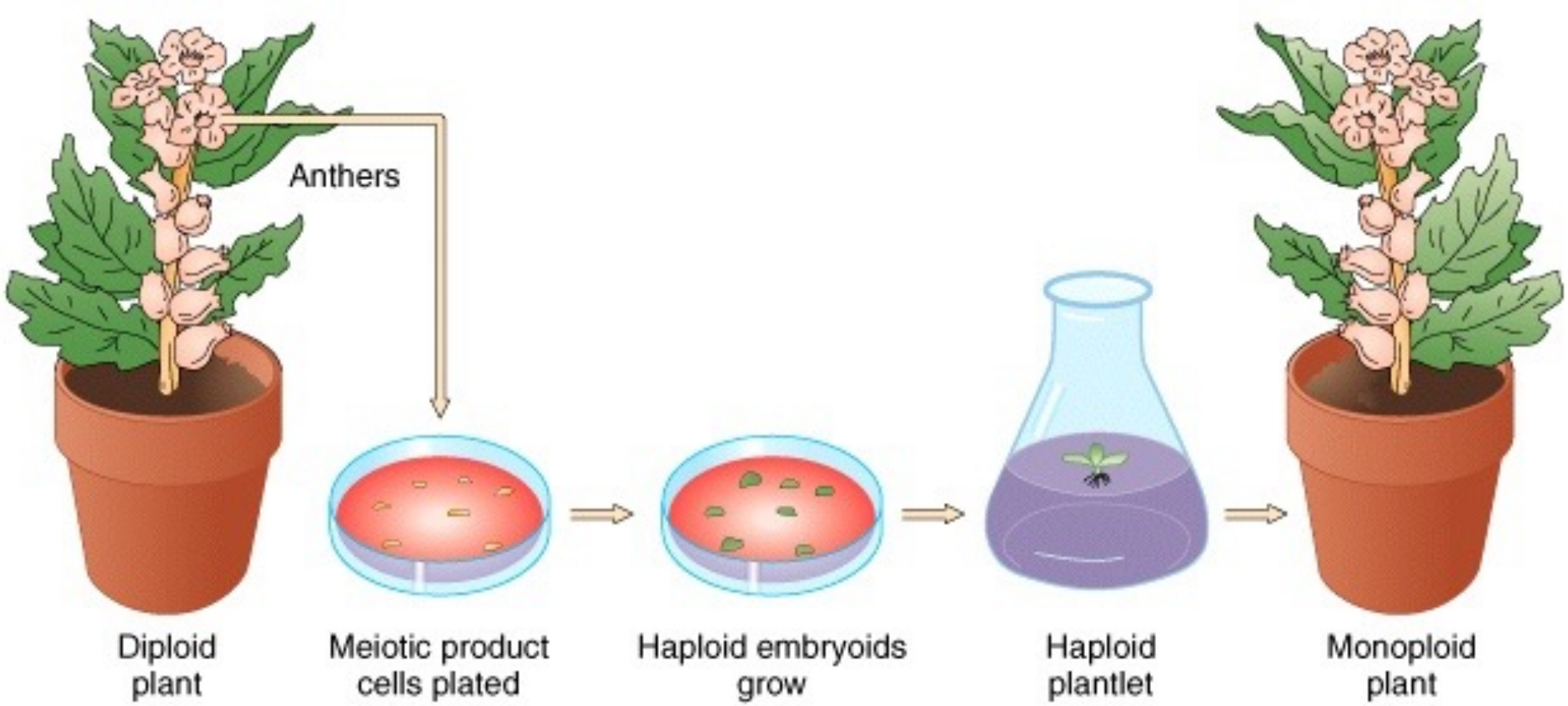
In this study, spermatogonia of p*vasa-Gfp* (where Gfp represents green fluorescent protein) hemizygous (p*vasa-Gfp*/–) and dominant orange-colored mutant heterozygous [*OR/wild type (WT)*] adult rainbow trout (*Oncorhynchus mykiss*) were intraperitoneally microinjected into newly hatched embryos of triploid sterile masu salmon (*O. masou*). Hybrids of these two species do not survive. Histological examination showed that, whereas the testes of 2-year-old triploid salmon in the control group (no transplantation) were immature and contained mostly spermatogonia, testes of recipients appeared normal (Fig. 1A). Ten of the 29 male triploid salmon recipients produced milt. Offspring produced with milt from these 10 recipients and wild-type trout eggs developed normally (fig. S1 and table S1). Five F_1 progeny were collected from

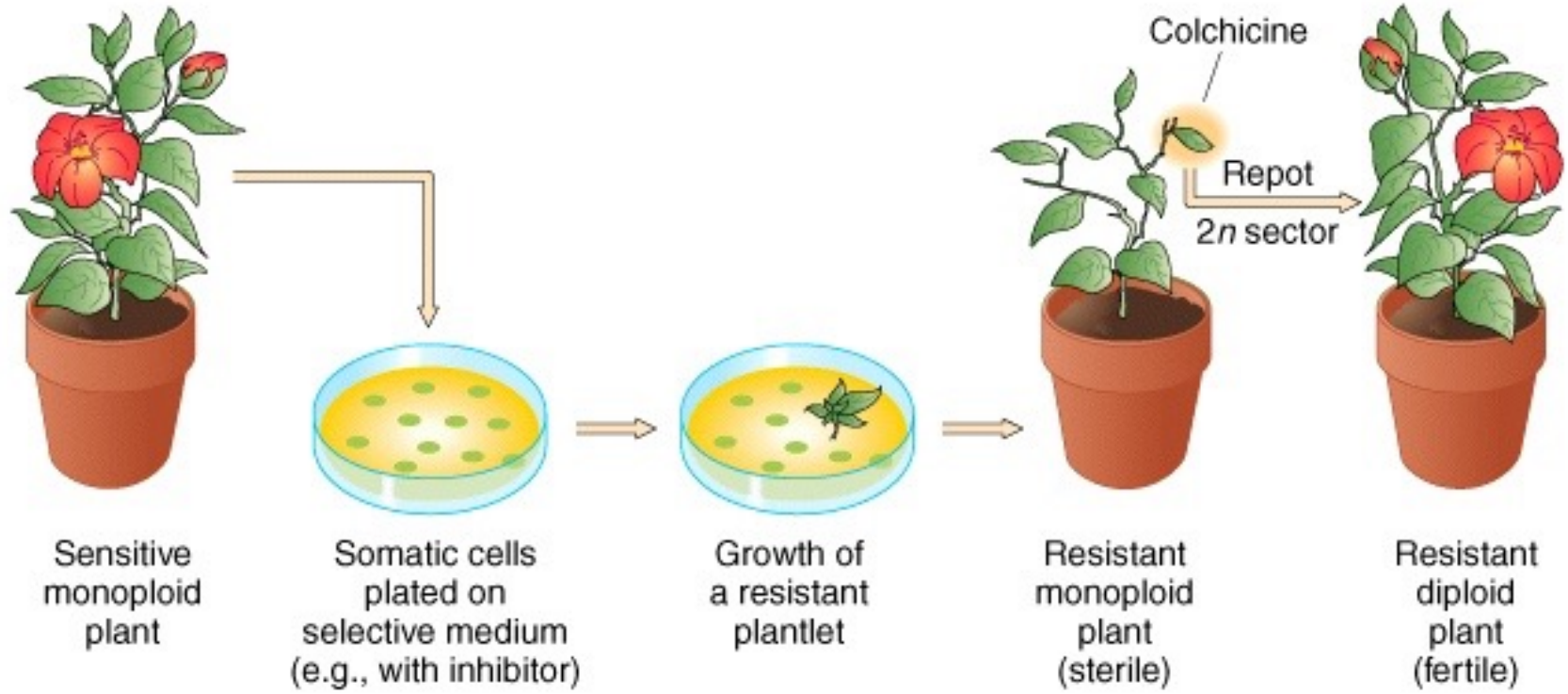
Monoploid plants in approaches to plant breeding

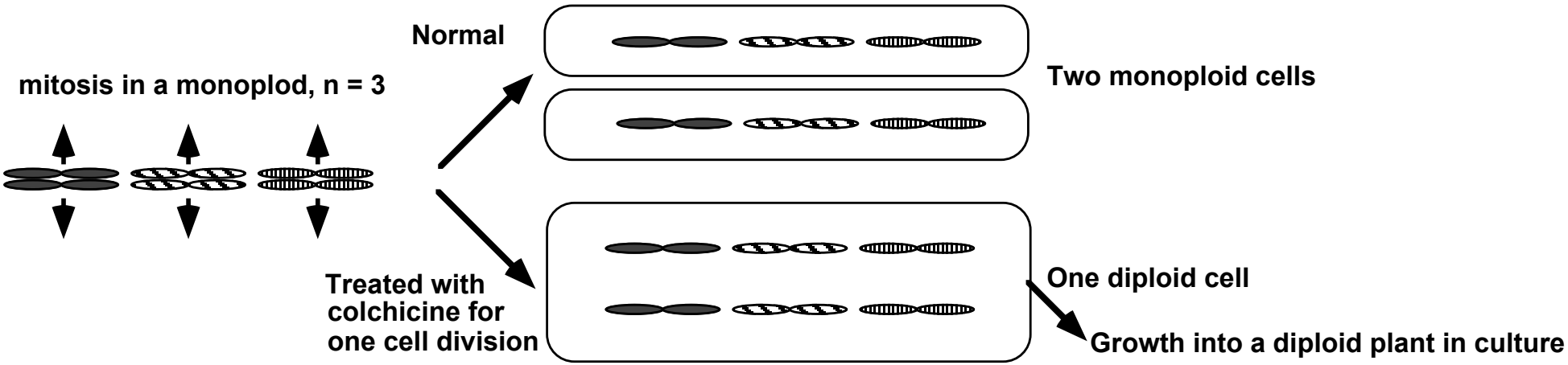
Diploidy is a pain in the ass if one is trying to create organisms that have particular genetic characteristics through mutagenesis. This is because in order to generate organisms that are homozygous recessive for some trait it is necessary to go through multiple generations in order to mutagenize, generate F1s, and then self these F1s and test them in the next generation for the desired characteristic. See for example the mutagenesis screen described at the end of lecture 2 in which we were trying to carry out a screen for plants with white flowers.

1. Cells from a plant can be dissociated, placed in culture, and used to generate an entirely new plant. Thus essentially all cells in a plant are pluripotent (able to contribute to all cell types).

2. The ploidy state of a plant can be easily manipulated in tissue culture.
A. Monoploid plants can be created from gametes grown in culture.
B. Plants with increased ploidy can be generated by preventing the segregation of sister chromatids during mitosis using the microtubule disrupting drug colchicine.




Colchicine
Repot
2n sector
Sensitive monoploid plant
Somatic cells plated on selective medium (e.g., with inhibitor)
Growth of a resistant plantlet
Resistant monoploid plant (sterile)
Resistant diploid plant (fertile)

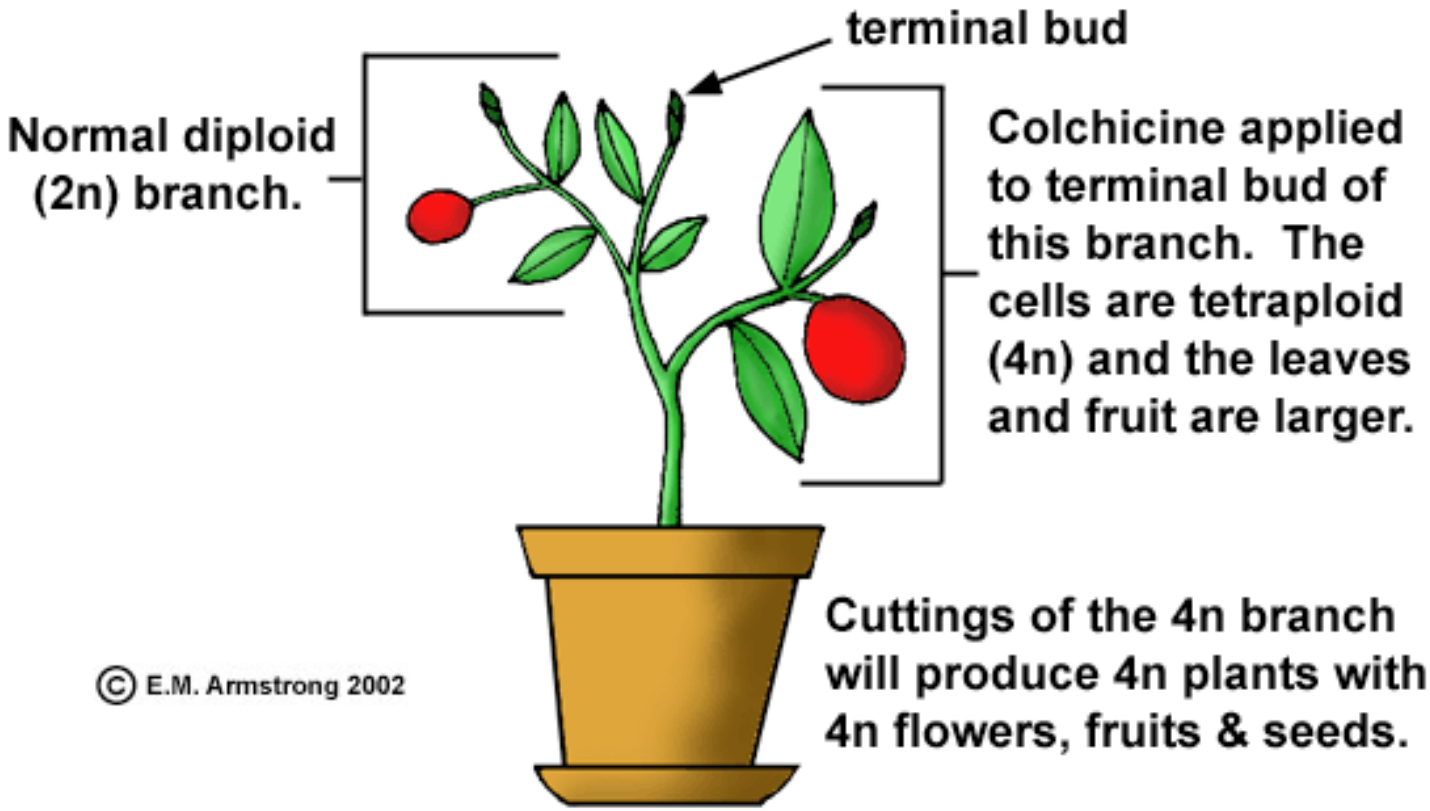

mitosis in a monoplod, n = 3
Normal
Two monoploid cells
Treated with colchicine for one cell division
One diploid cell
Growth into a diploid plant in culture

Autopolyploidy

...polyploidy resulting from the replication of one or more sets of chromosomes,

…the additional set of chromosomes is identical to the normal haploid complement of that species.

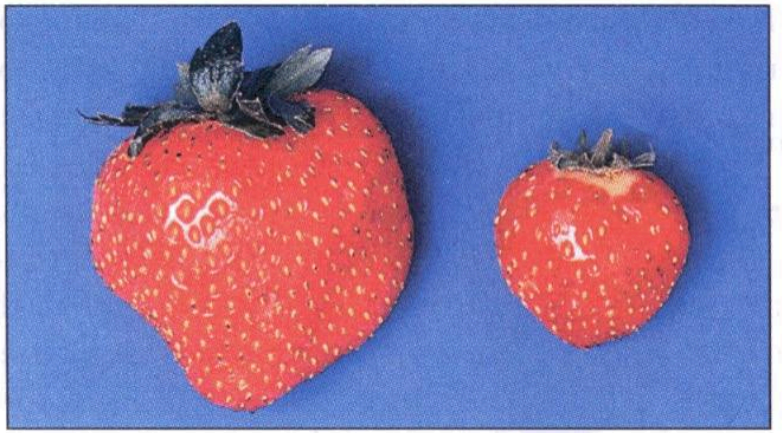
Can artificially use colchicine to increase ploidy.



Autopolyploidy Applications

- Treating a plant with colchicine often produces autopolyploidy, resulting in plants with larger flowers and/or fruit,

4n *2n*



8n *2n*

…And of course generating adults that can be crossed with diploids to generate sterile triploid progeny

(a)

(b)

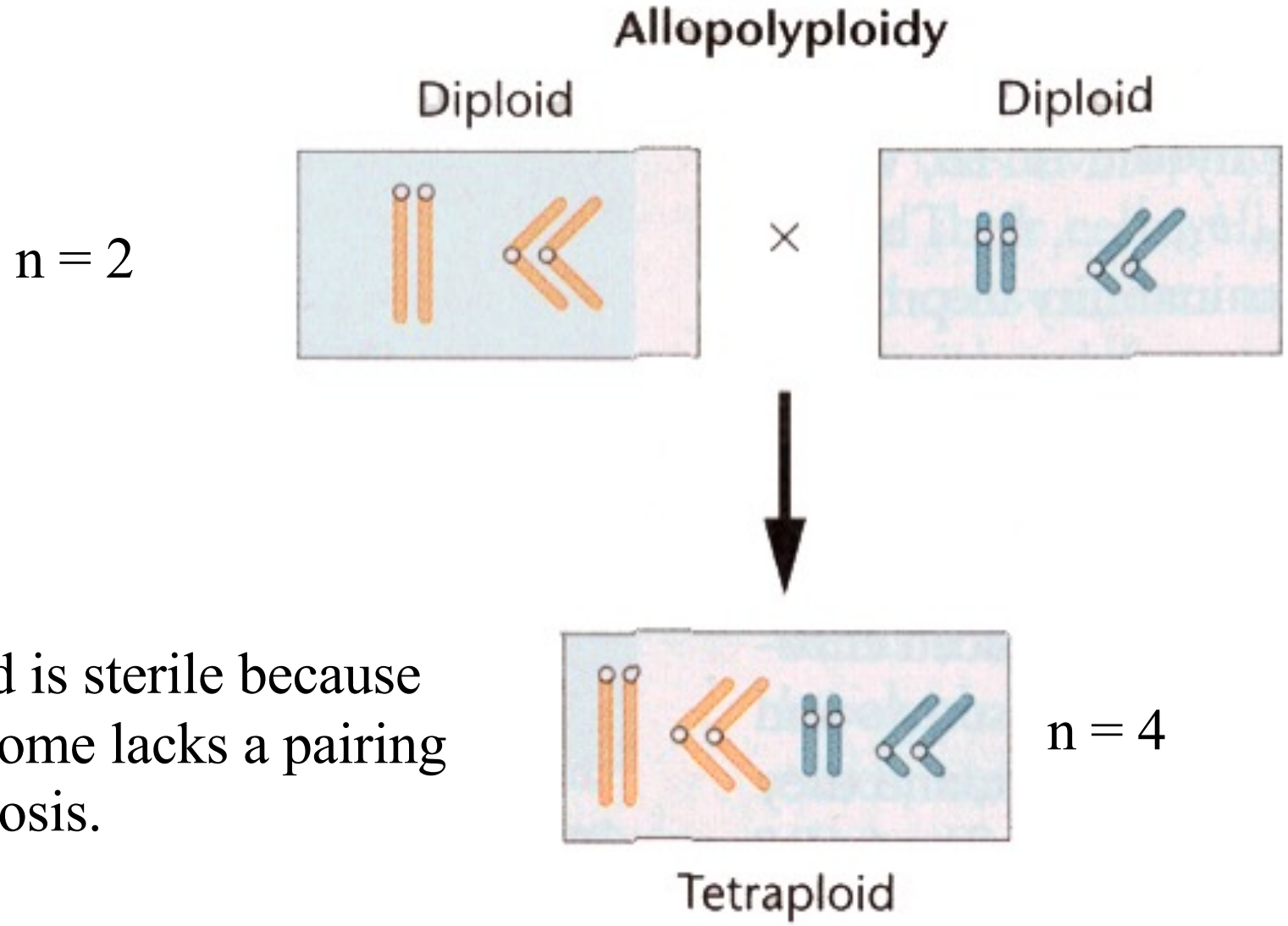
Diploid & Tetraploid *Hyla*

These are essentially the same frog - it is the *balance* between the number of gene copies that is important for proper development, not necessarily the *total number* of gene copies

Up until now we have been dealing with autopolyploids, which contain multiple sets of chromosomes from the same species. Allopolyploids are produced from sets of chromosomes of closely related species.

Closely related chromosomes from different species are not homolgous, but are referred to as homeologous.

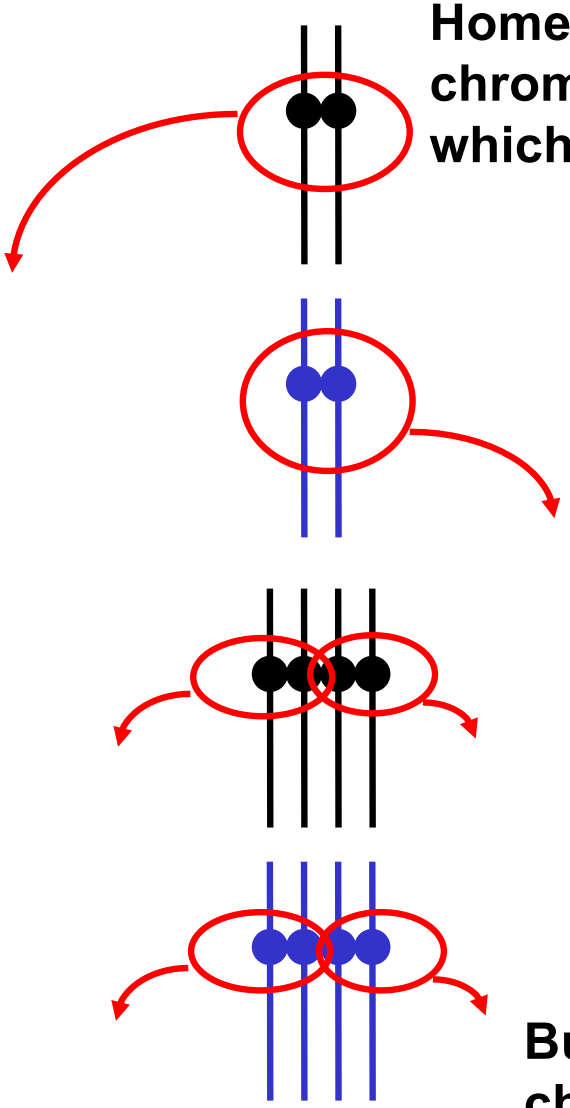
Homeologous chromosomes do not pair during meiosis and this can lead to problems.



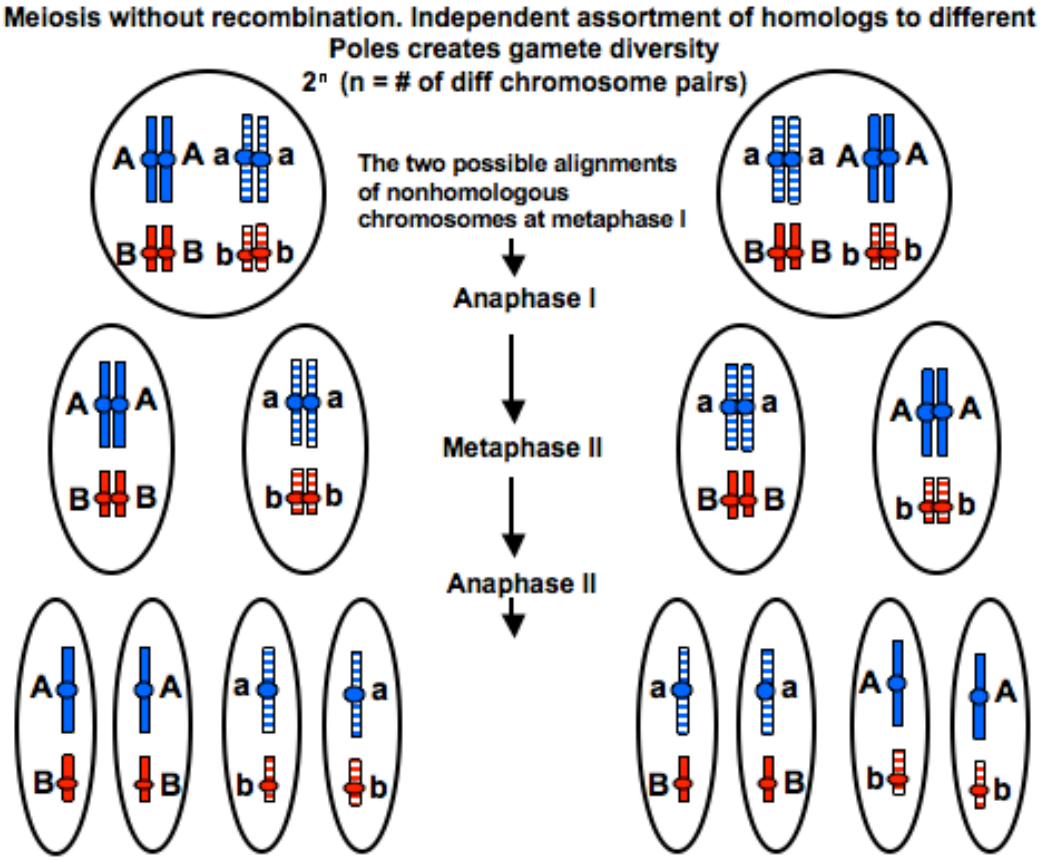
This tetraploid is sterile because each chromosome lacks a pairing partner at meiosis.

Solution: Duplicate each chromosome by using colchicine to double the number of chromosomes, thereby giving each chromosome a pairing partner.

If cross two related species, ok until meiosis.



Homeologous chromosomes don't pair, which leads to sterility.



But by doubling the chromosomes, generate an allopolyploid (amphidiploid) that is fertile

The Sad Tail of *Raphanobrassica*

Cast

Dr. G. Karpechenko, Russian Plant breeder,

Brassica oleracea (2*n* = 18),

- cabbage, large above ground food stock, worthless root,

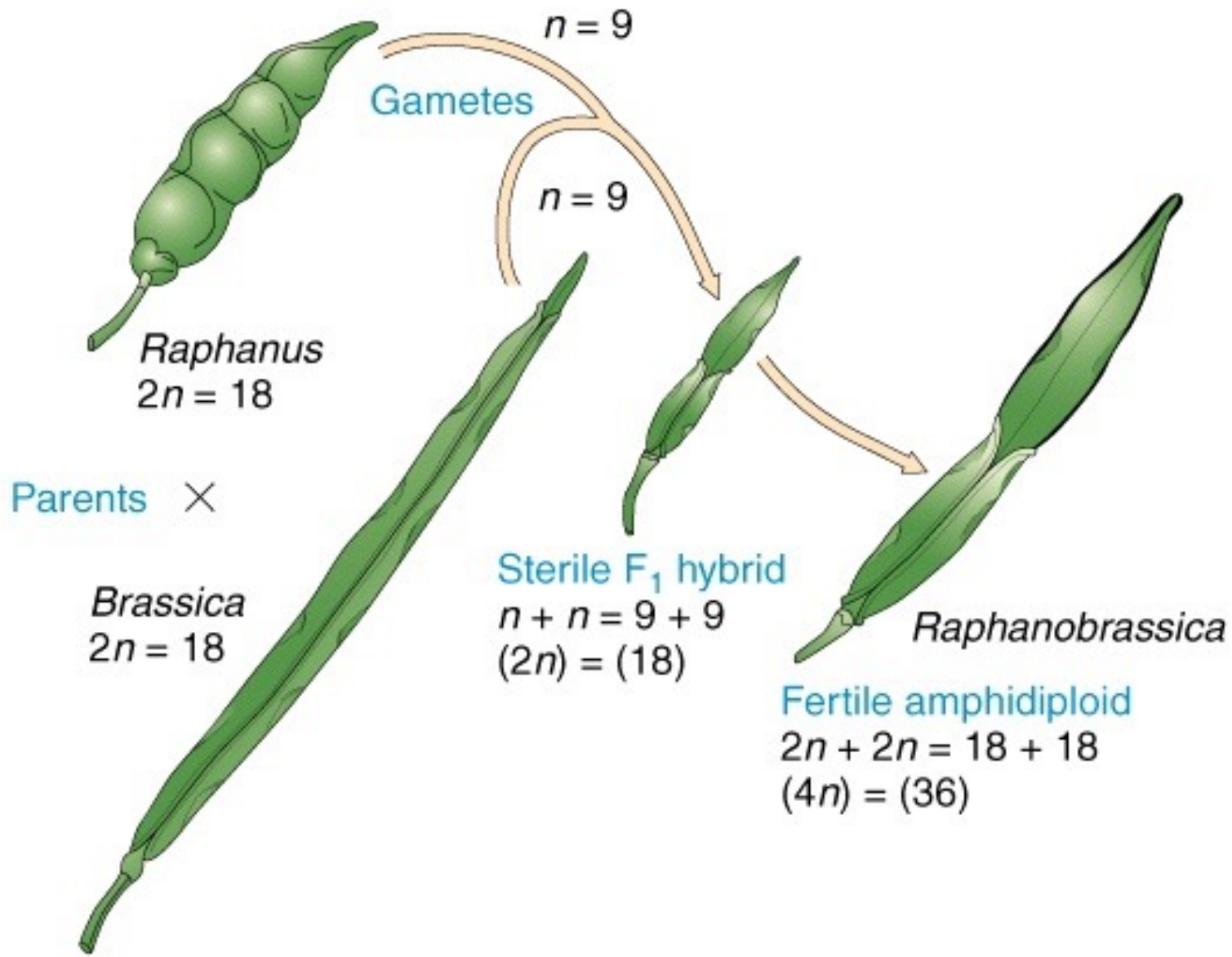
Raphanus sativus (2n = 18),

- the common radish, large root food stock, worthless leaves.

Georgi Karpechenko

wanted to generate a plant that had the leaves of a cabbage and the roots of a radish.

cabbage + radish = cabrad?


n = 9
Gametes
n = 9
Raphanus
2n = 18
Parents ×
Brassica
2n = 18
Sterile F1 hybrid
n + n = 9 + 9
(2n) = (18)
Raphanobrassica
Fertile amphidiploid
2n + 2n = 18 + 18
(4n) = (36)

and produced a …radcab!

leaves of a radish, roots of a cabbage. No good for people food. But.....

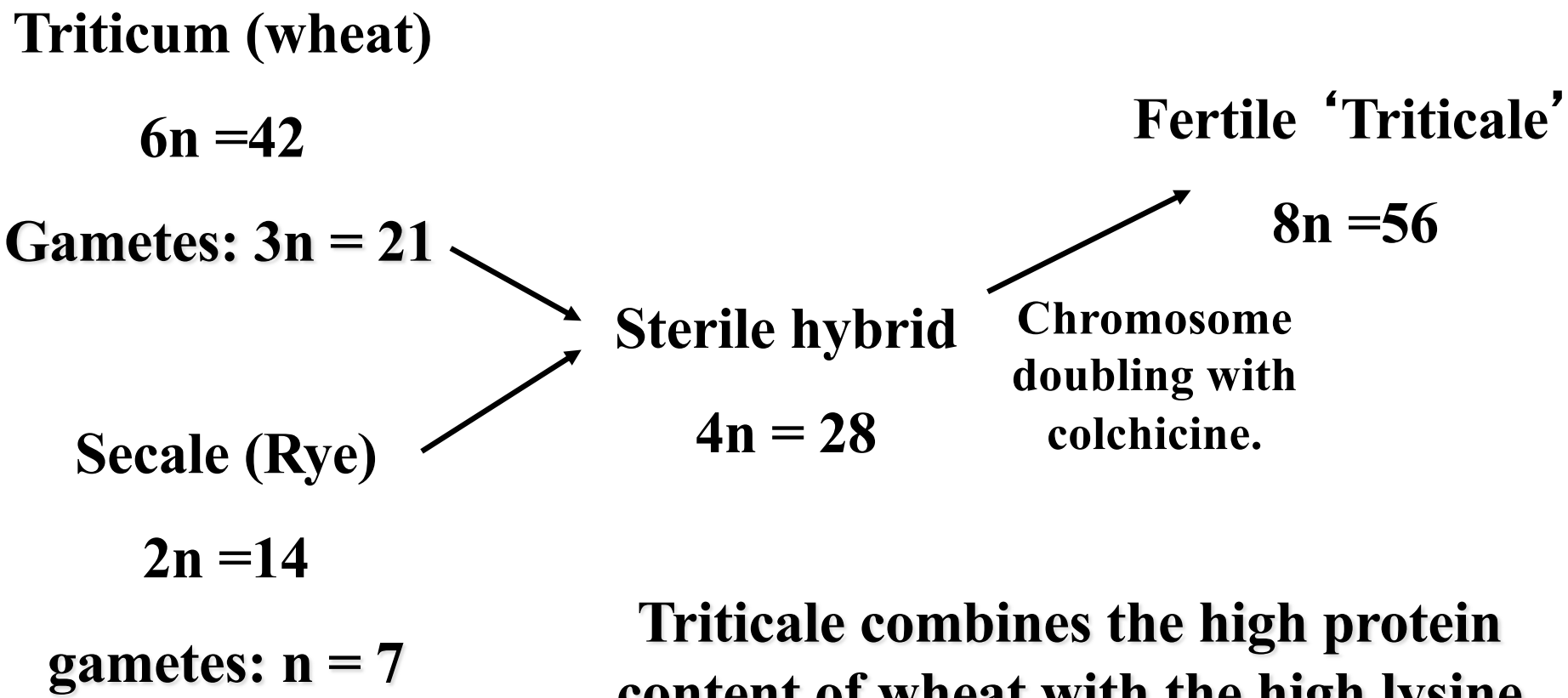
Revived by the Scottish Plant Breeding Station in Dundee, Scotland,

Raphanobrassica's high dry matter content makes it an excellent fodder for sheep and cattle.

Cannot breed with either parent = a new species!!!

While it is possible to produce allopolyploids, the outcomes are unpredictable.

One successful allopolyploid is Triticale



Triticale combines the high protein content of wheat with the high lysine levels and ability to adapt to marginal environments of rye.

Polyploidy Summary

- More than 2 whole sets of chromosomes,

- Autopolyploidy,
  - from the same genome,
  - naturally occurring, or induced,
  - often results in larger varieties,

- Allopolyploidy,
  - from different genomes,
  - naturally occurring, or induced,
  - often results in larger varieties,

- Autotriploids,
  - most often sterile
  - can produce beneficial traits.

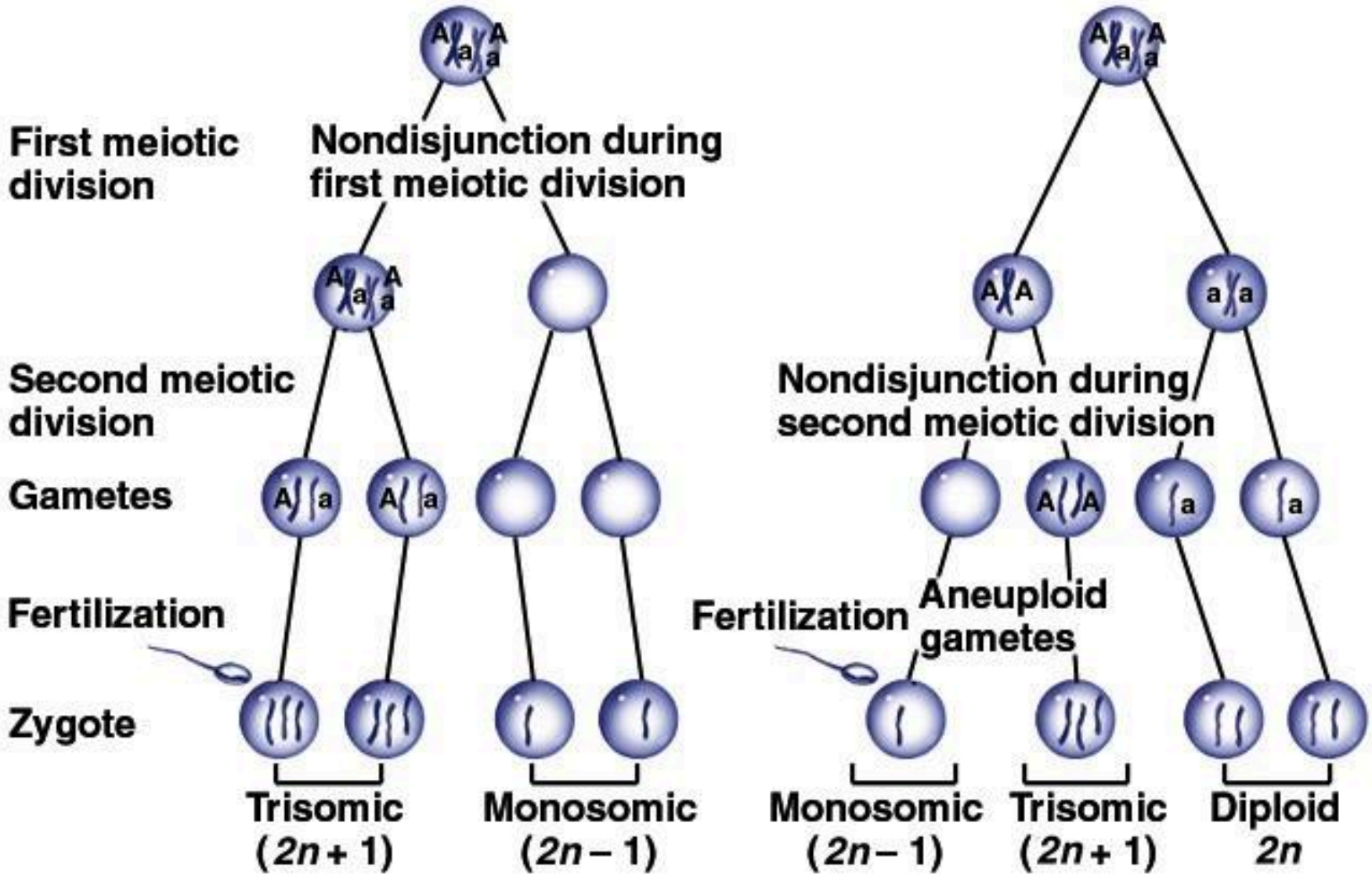
Aneuploidy

Nondisjunction in meiosis I or II results in gametes with an extra or missing chromosome.

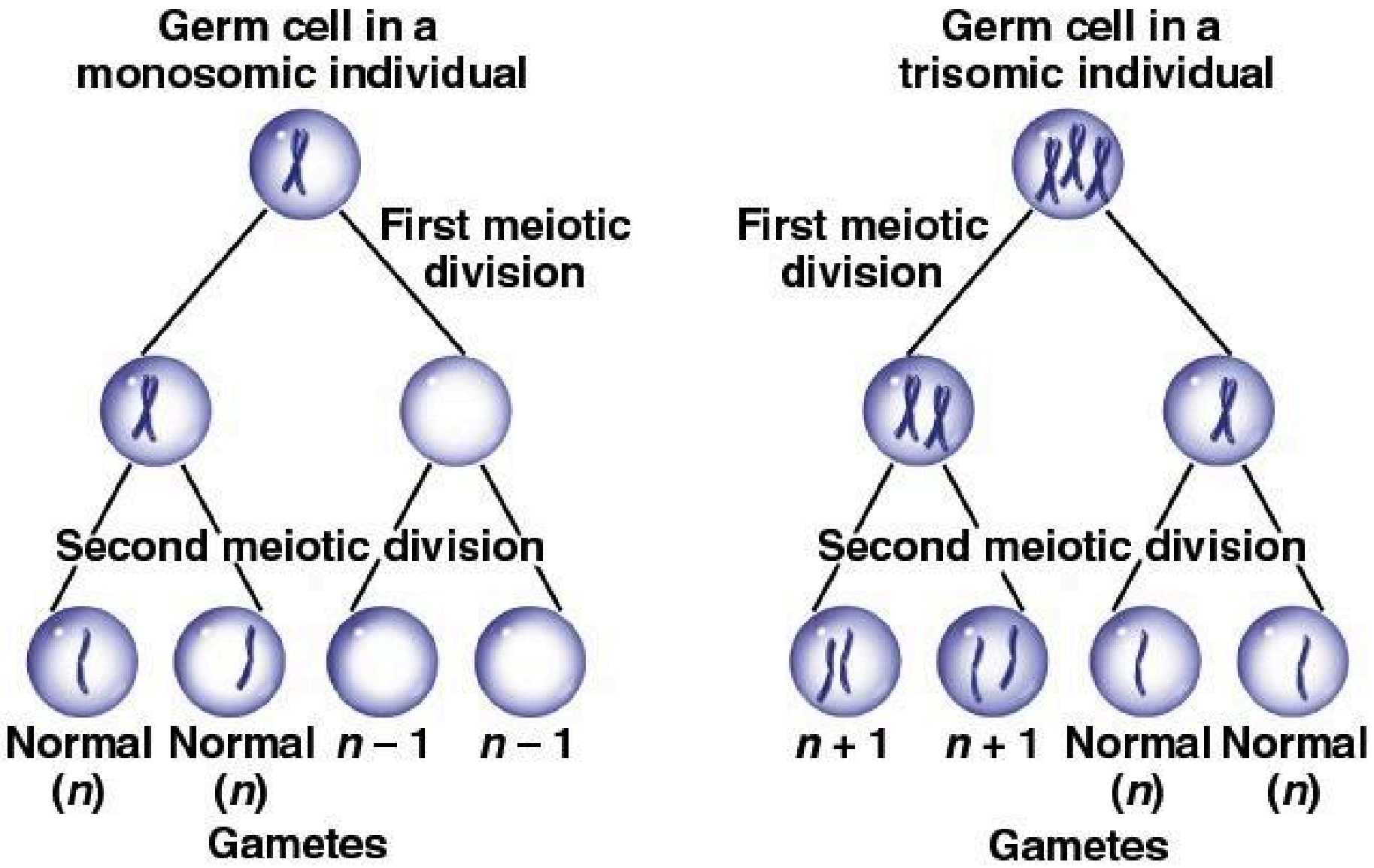
When these gametes fuse, the fusion results in zygotes with an extra or missing chromosome, a situation termed aneuploidy!

What are the consequences of aneuploidy?

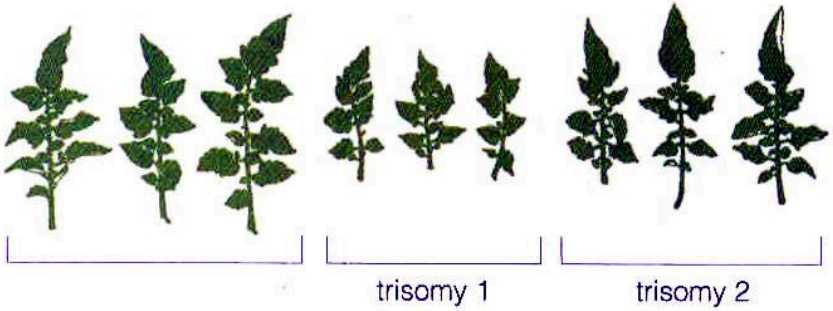
Aneuploidy

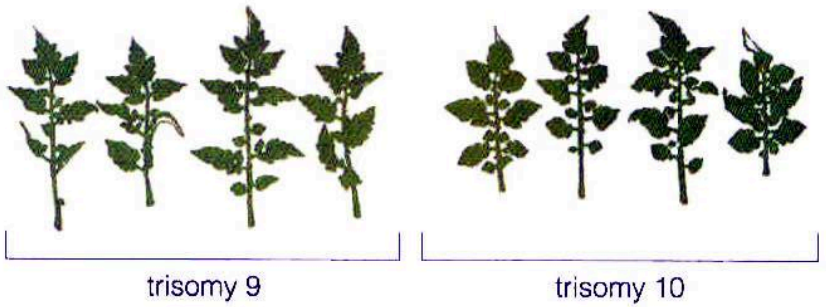


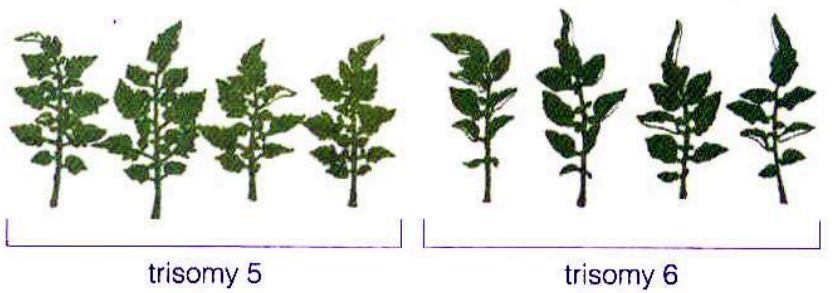
Aneuploidy

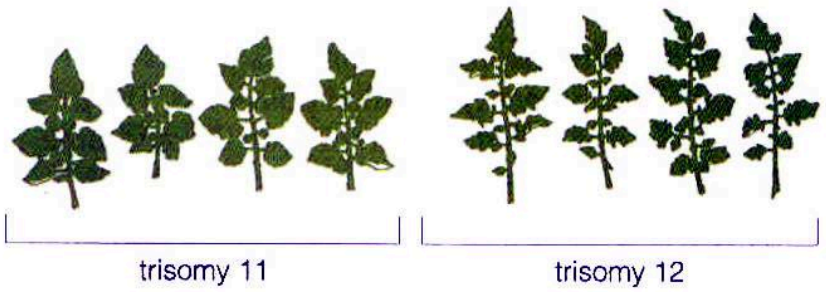


Plants just don't care that much: the tomato









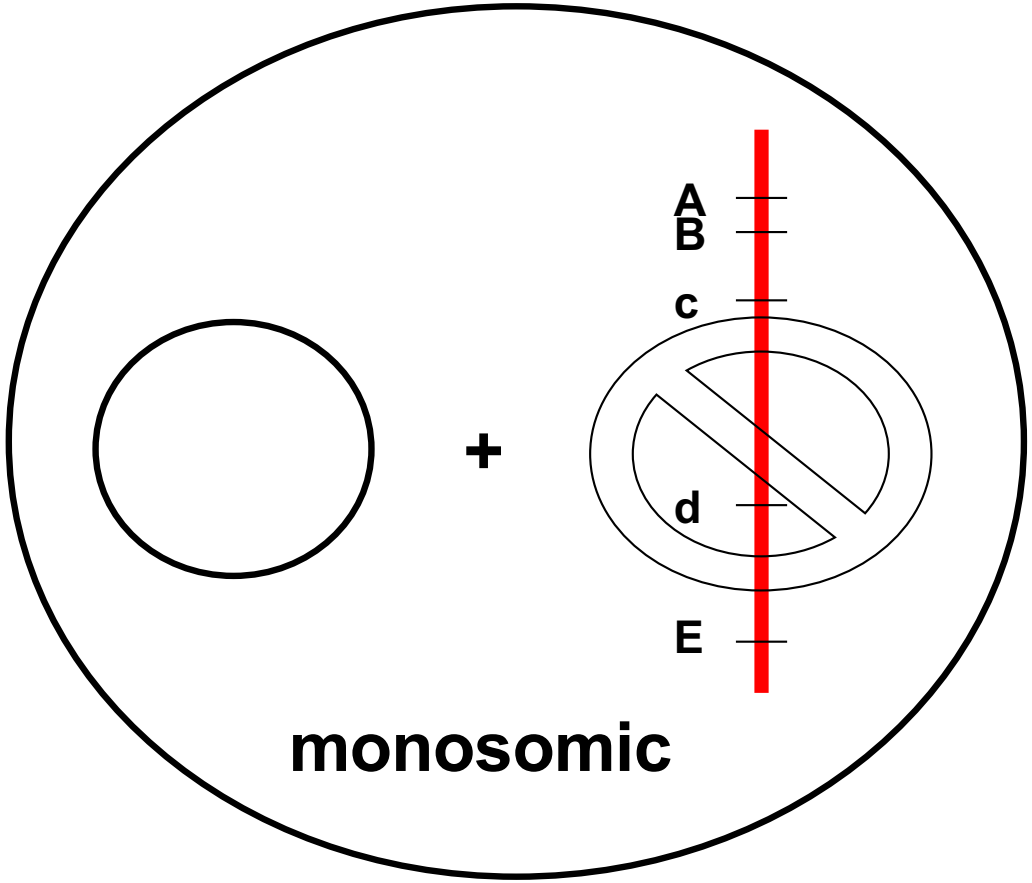
Human Autosomes first!

All monosomic (43 autosomes; missing an autosome) spontaneously abort.

Almost all trisomic (45 autosomes; an extra autosome) fetuses spontaneously abort!
There are three exceptions.

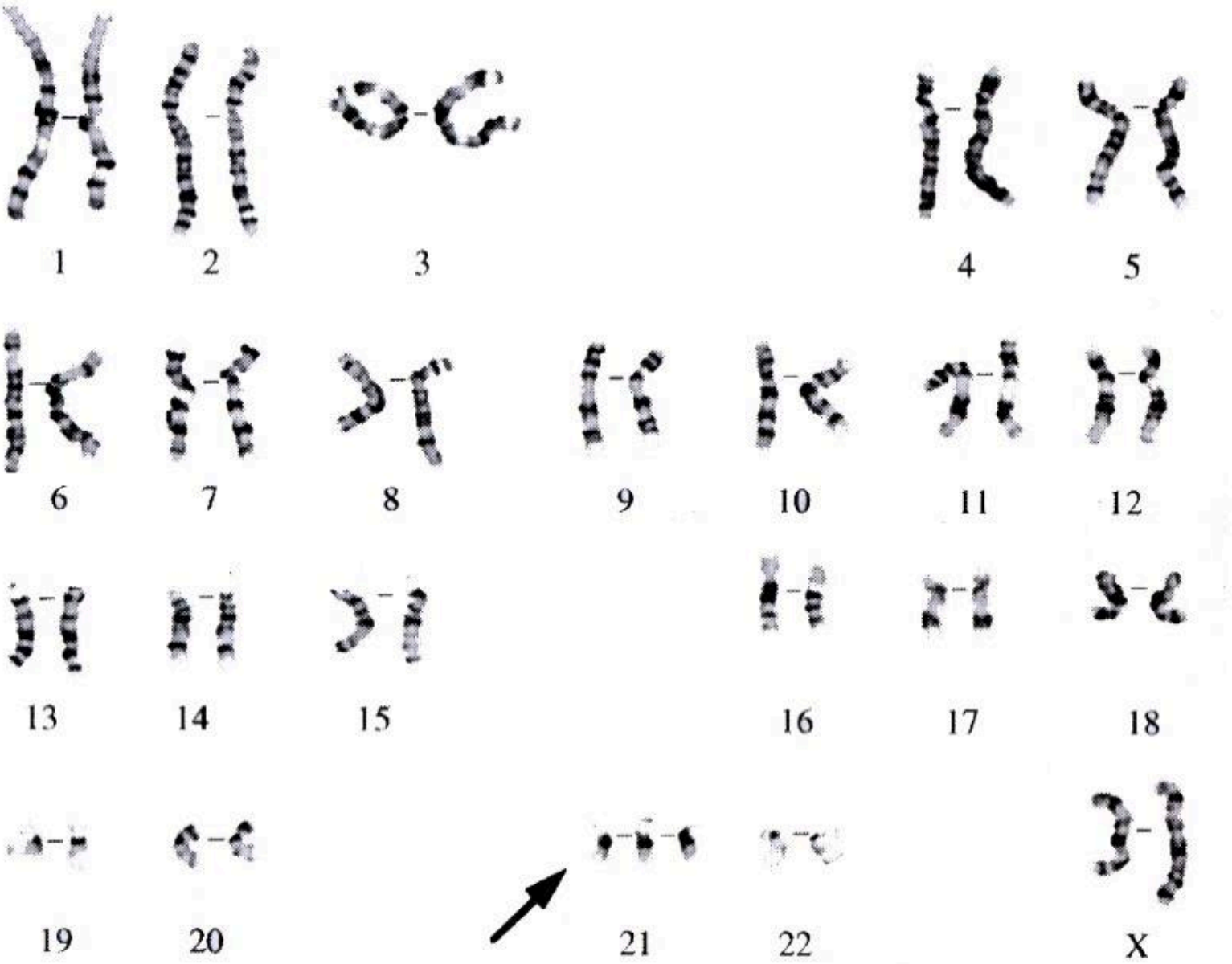
Autosomal Monosomy

- nearly always deleterious, usually lethal,
  - recessive lethal alleles are expressed,



c and d are recessive lethal alleles. Therefore, a monosomic individual that carries this chromosome is dead.

Down Syndrome (47, +21)



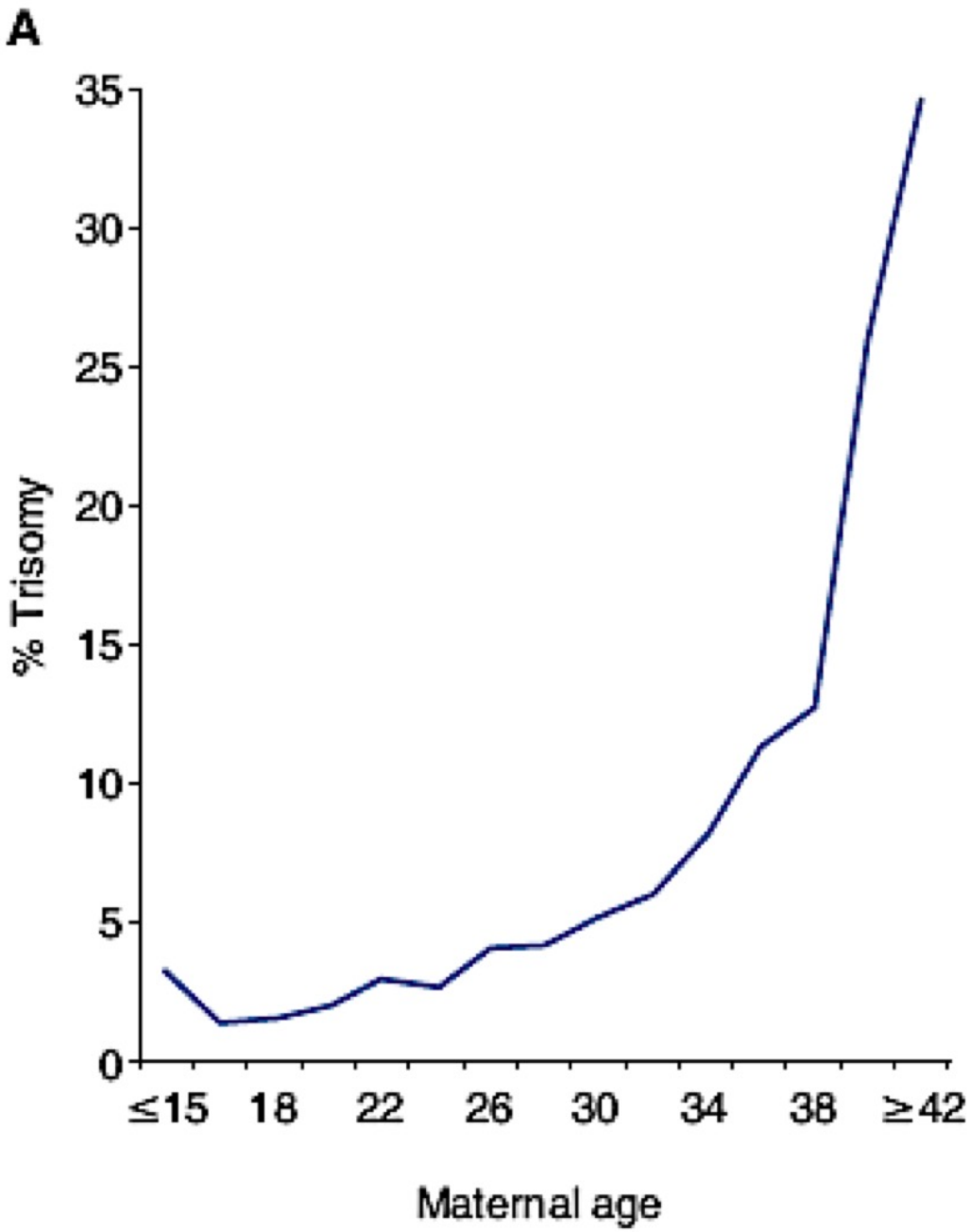
Trisomy in Humans

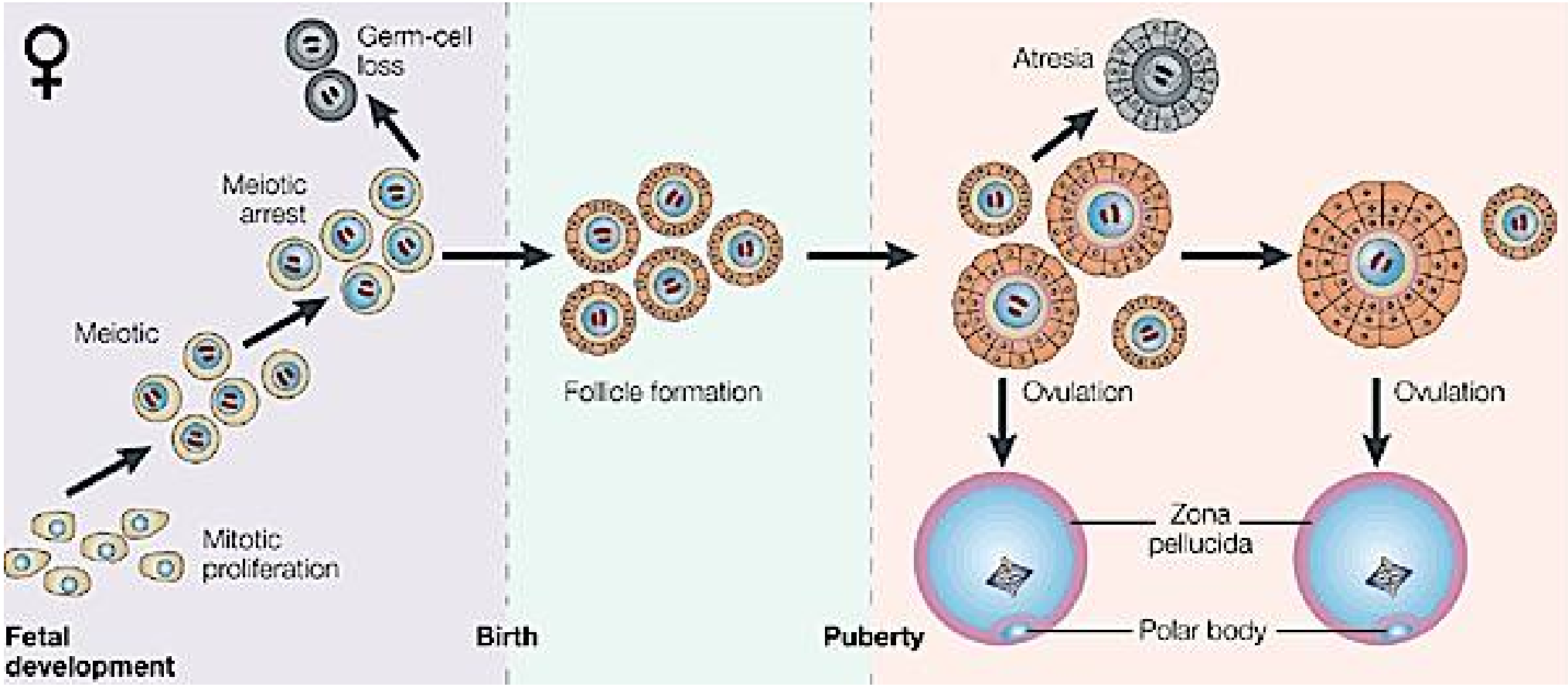
Down Syndrome

(47, +21)

- occurs at a frequency of about 0.15%,
- mental retardation (I.Q. 20 - 50),
- mean life expectancy is about 17 years,
- 47, +21 females are fertile,
- trisomy 21 risk factors, maternal age.

Maternal Age and trisomy

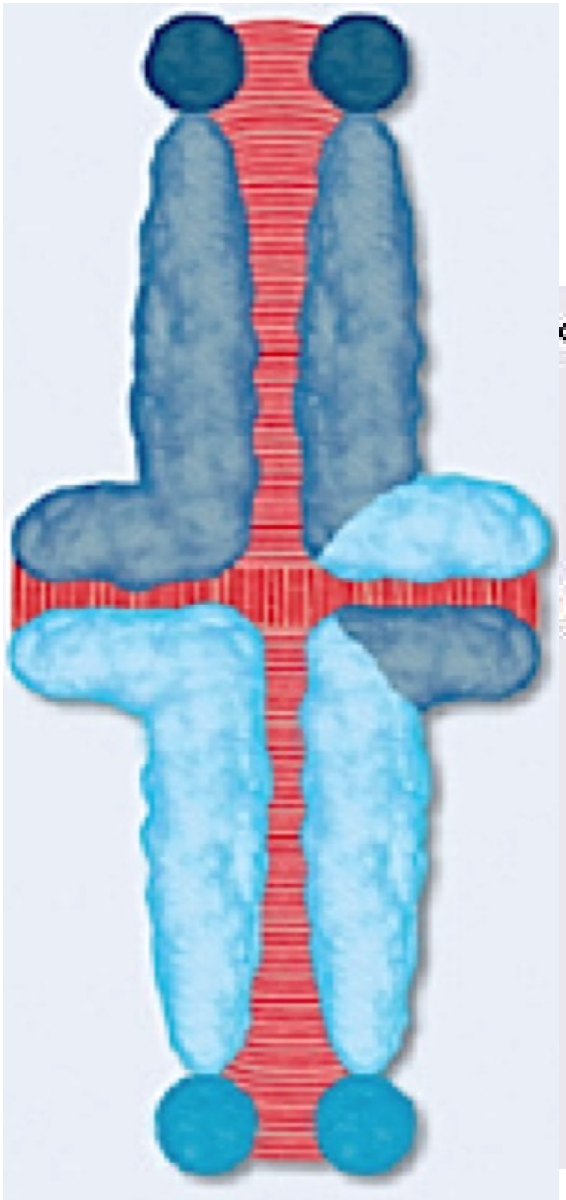




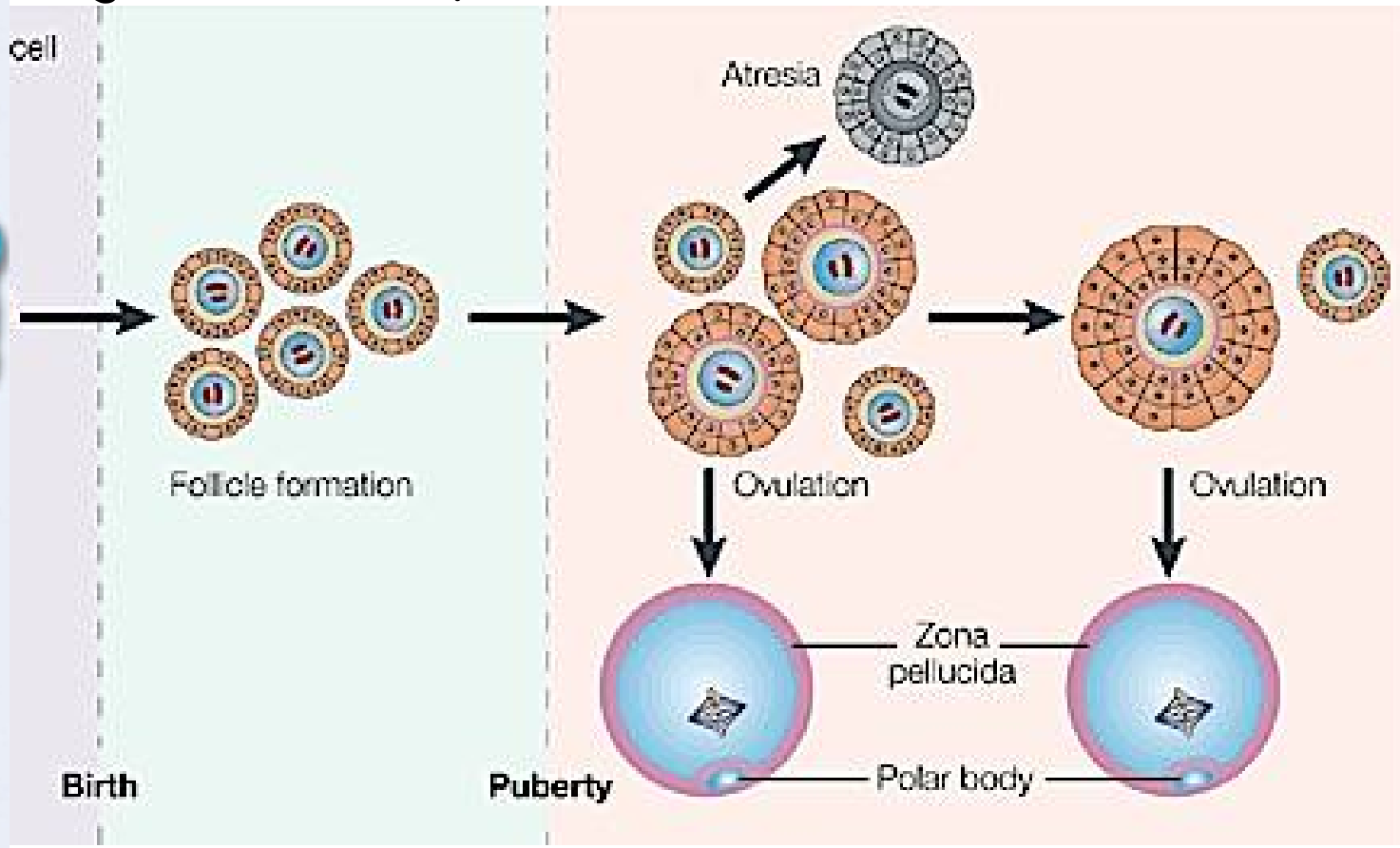
-In female humans, oocytes arrest in Prophase I before birth. Meiosis continues upon menstruation.

-Thus, tetrad formation and cellular function must be maintained for decades.

- The proteins mediating



-The cohesion molecules that maintain alignment of the homologs when you are born are literally the same protein molecules that maintain alignment when you are 40. Errors arise.



-In female humans, oocytes arrest in Prophase I before birth. Meiosis continues upon menstruation.

-Thus, tetrad formation and cellular function must be maintained for decades

Other Human Autosomal Trisomics

- **Edwards Syndrome (47, +18)**
  - **0.0125% live births,**
  - **mental abnormalities,**
  - **“faunlike ears”, “rockerbottom feet”, small jaw,**
  - **nearly all die within several weeks of birth,**

- **Patau Syndrome (47, +13),**
  - **0.005% live births,**
  - **mental abnormalities,**
  - **cleft lip, small mal-formed head, “rockerbottom feet”,**
  - **mean life expectancy is about 130 days.**

Somatic polyploidy

Drosophila salivary glands. 1024-2048 copies

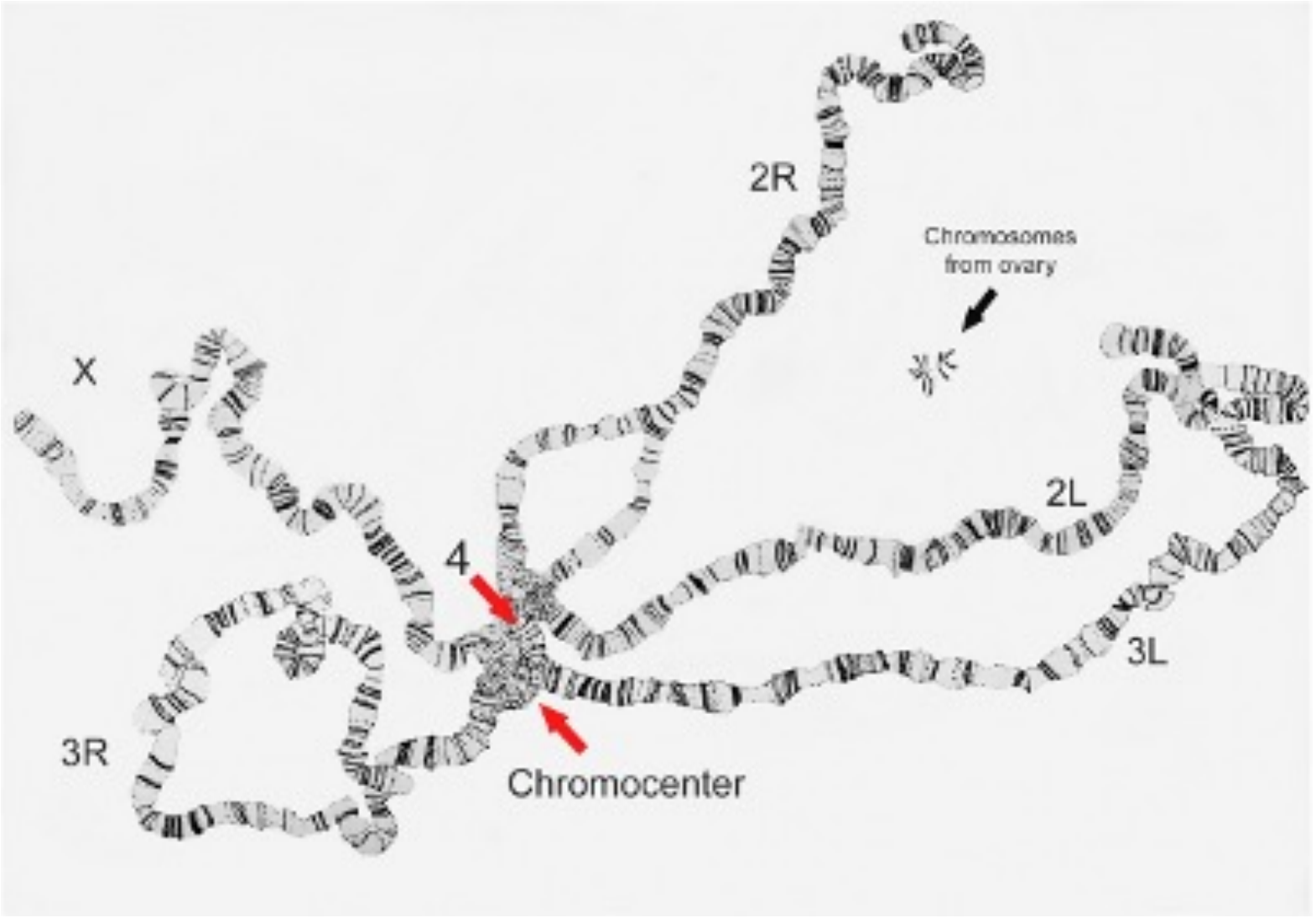
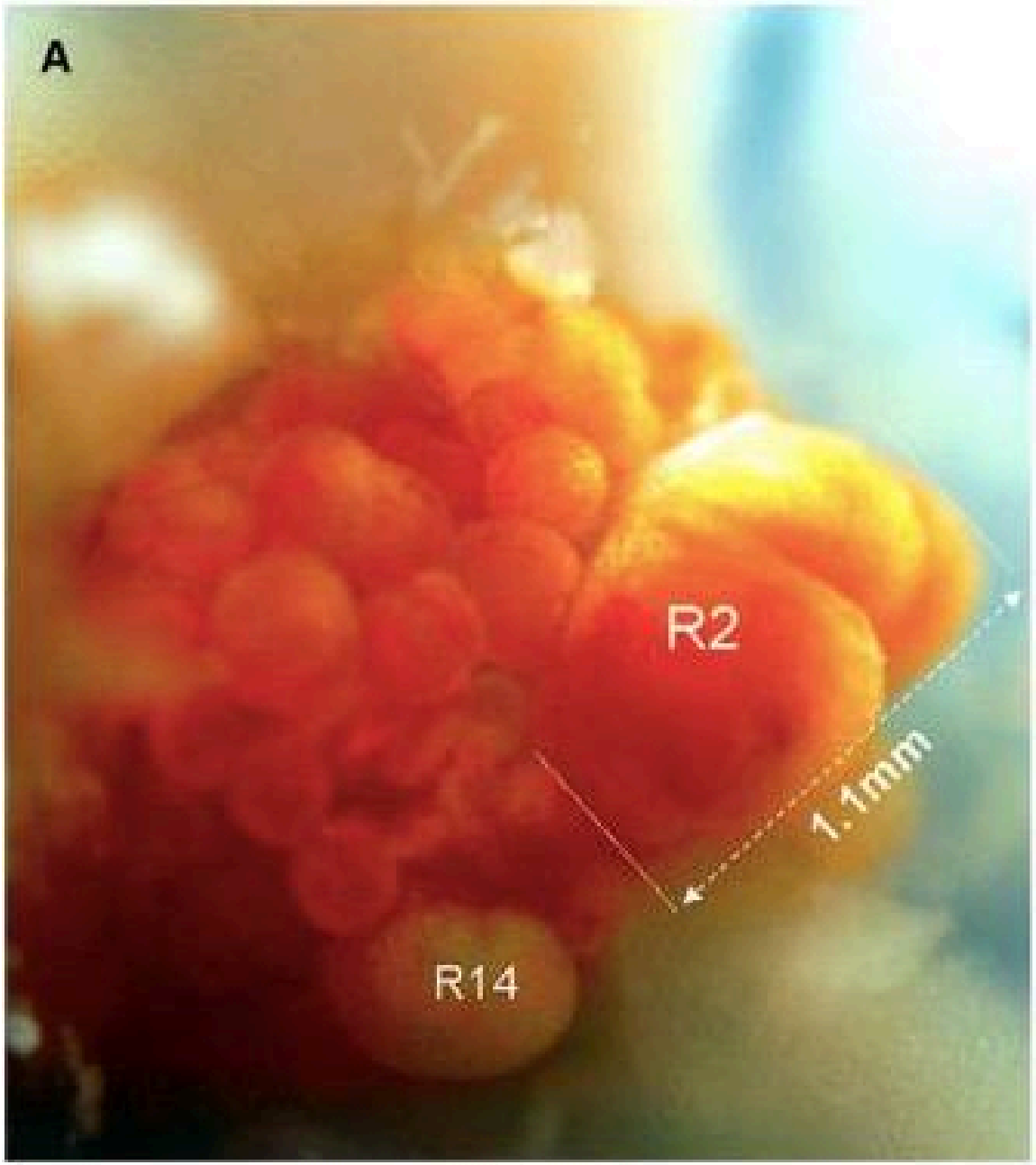


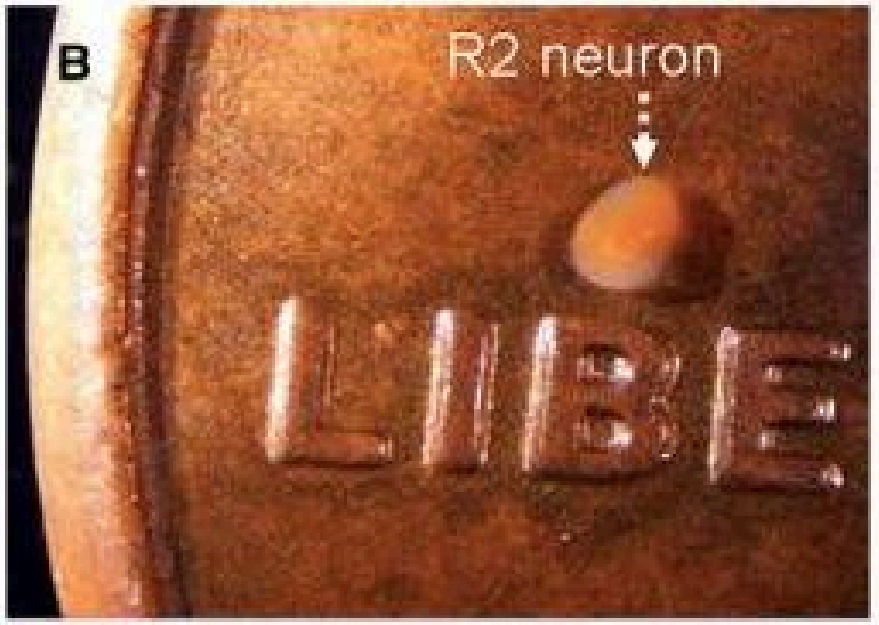
TABLE 1 - *High-level polyteny in animals and plants. C values are based on DNA measurements, n values on nuclear volumetry.*

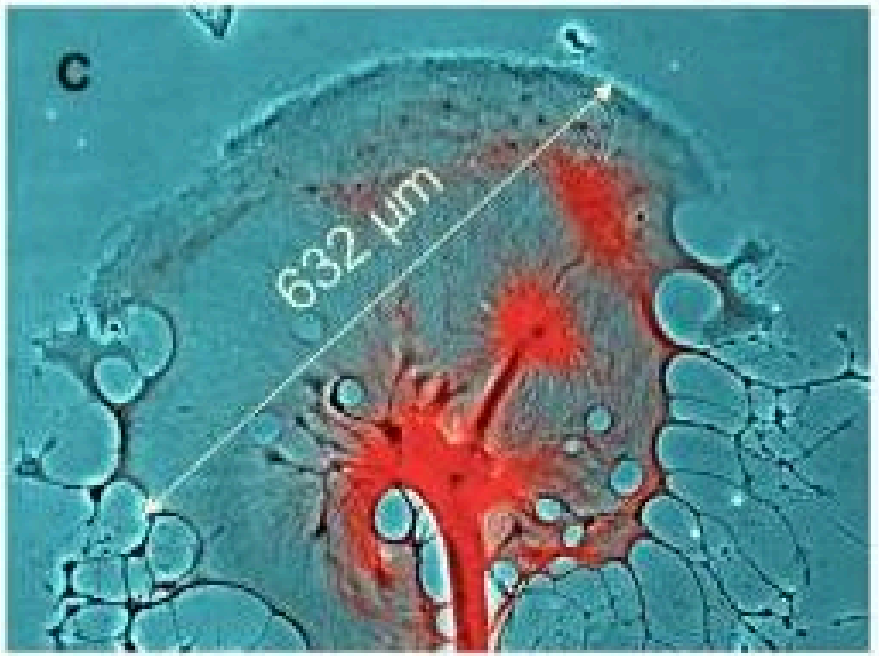
Species	Tissue or cell	Maximum level of polyteny	Reference
ANIMALS			
Aplysia californica	Giant neurons	200,000 C	LASEK and DOWER 1971 [101]
Ascaris suum	Esophageal gland	262,144 C	ANISIMOV 1976 [102]
Bombyx mori	Silk gland	524,288 C	GAGE 1974 [103]
	id, posterior zone	1,048,576 C	PERDRIX-GILOT 1978 [104]
Chironomus tentans	Salivary gland	32,768 n	BEERMANN 1952 [105]
Chironomus thummii	Salivary gland	16,384 C	RASCH 1970 [106]
Drosophila melanogaster	Salivary gland	2,048 C	RASCH 1970 [106]
Microtus arvalis	Trophoblast	2,048 C	ZYBINA *et al.* 1975 [107]
Mus musculus	Trophoblast	1,024 C	BARLOW and SHERMAN 1972 [108]
Rattus rattus	Trophonblast	4,096 C	NAGL 1972 [109]
Sciara coprophila	Salivary gland	8,192 C	RASCH 1970 [106]
PLANTS			
Arum maculatum	Endosperm haustorium *	24,576 n	ERBRICH 1965 [110]
Echinocystis lobata	Endosperm, central	3,072 n	TURALA 1966 [111]
Melampyrum nemorosum	Endosperm haustrium	1,536 n	ERBRICH 1965 [110]
Phaseolus coccineus	Suspensor	8,192 C	BRADY 1973 [112]
Phaseolus vulgaris	Suspensor	4,096 C	NAGL 1974 [113]
Scilla bifolia	Elaiosome	4,096 n	SPETA 1972 [114]
Scilla decidua	Antipodal **	1,024 C	FRISCH and NAGL 1979 [75]

* Endosperm cells are initially triploid ($3n$).
** Antipodal cells are initially haploid (n).

Aplysia R2 neuron

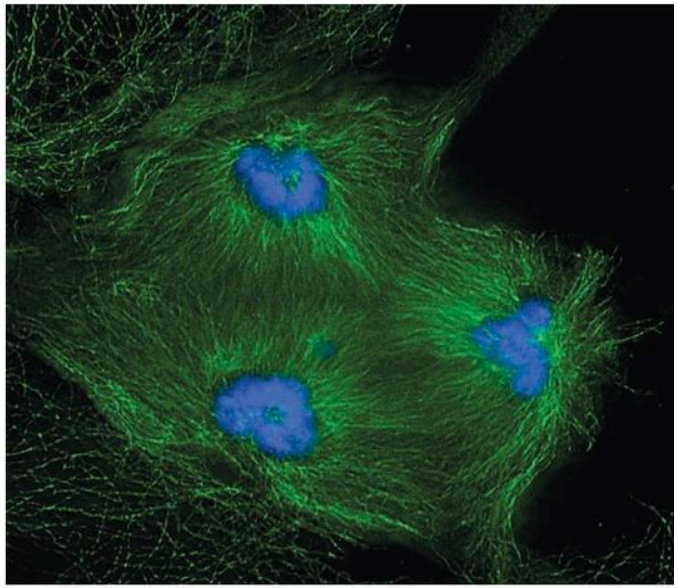






POLYPLOID CELL TYPES IN MAMMALS

CELL TYPE	LOCATION	FUNCTION	NUMBER OF GENOME COPIES
Megakaryocyte	Bone marrow	Producing blood clotting platelets	Up to 128
Hepatocyte	Liver	Detoxification, metabolism	Typically 4 to 16
Trophoblast giant cell	Embryo	Promote implantation	Up to 1000
Cardiomyocyte	Heart	Contraction	Typically 4



Bonus DNA. The polyploid cells in mammalian bodies differ in their location, function, and number of chromosome sets (table). In a liver cell (*right*), the multiple chromosome copies (blue) have sorted into three clusters in preparation for cell division.

Region-specific amplification/underreplication

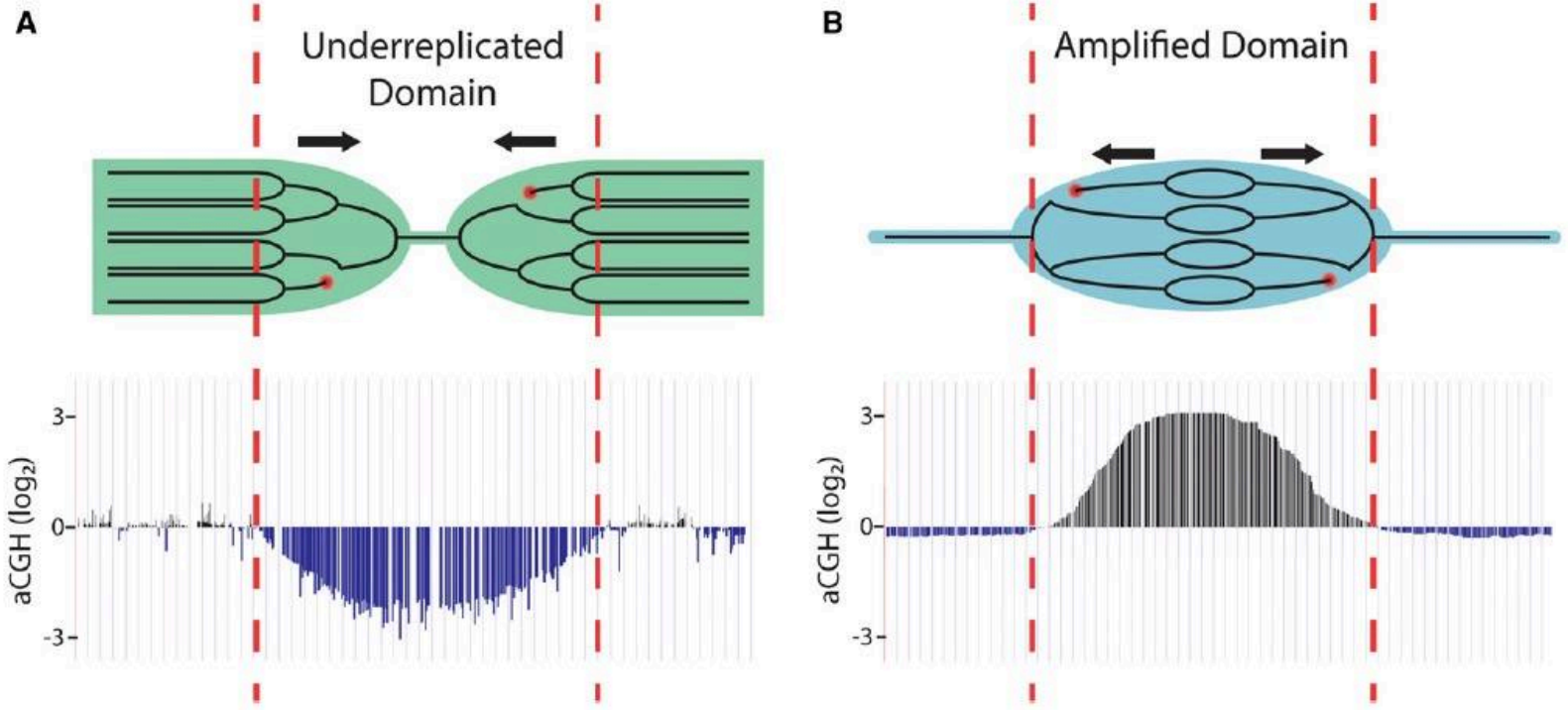


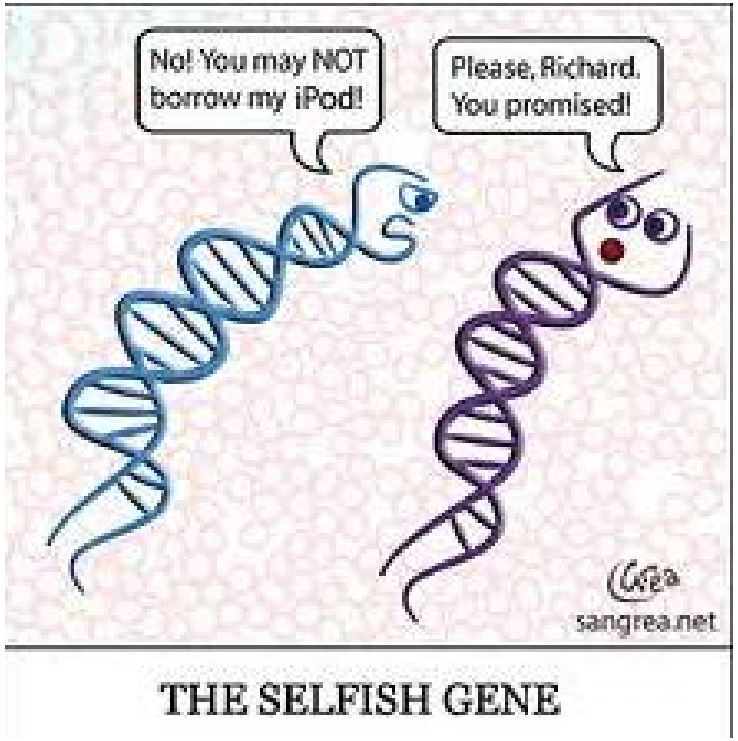
Figure 3 Differential DNA replication. (A) Underreplication results from two effects: absence of replication origins and initiation within a domain coupled with impaired progression of replication forks initiating from flanking origins (arrows indicate direction of fork progression). Replication forks are destabilized within underreplicated regions and can lead to double-stranded DNA breaks (red circles). These can lead to deletions or rearrangements in the region. Array-based comparative genome hybridization (aCGH) analysis of an underreplicated site indicates decreased copy number relative to overall ploidy. aCGH data are modified from Sher *et al.* (2012). (B) In endocycling follicle cells, developmentally programmed gene amplification occurs through repeated replication origin firing followed by bidirectional fork progression away from the origin (arrows indicate direction of fork progression). Rereplication can lead to fork collisions, resulting in double-strand DNA breaks (red circles). aCGH analysis of an amplified site indicates a gradient of increased copy number relative to overall ploidy spanning 100 kb, with the highest copy number at the origin of replication. aCGH data are modified from Kim *et al.* (2011).

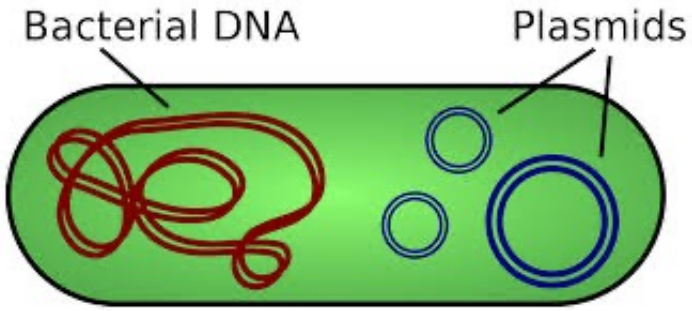
Table 2 *Drosophila* amplicons in follicle cells

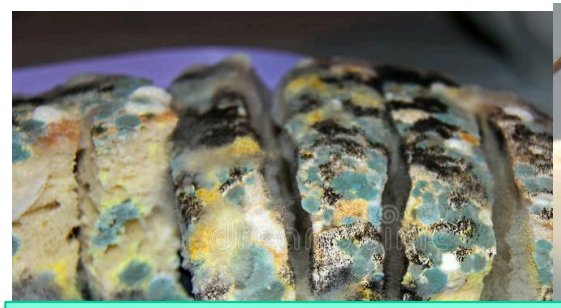
Cytological location	Max fold amplification	Stages of origin firing	Genes involved in egg shell function
7F	18–20	10B–11	*Cp7Fa*, *Cp7Fb*, *Cp7Fc*, *Cp36*, *Cp38*
22B	4	10B–13	None
30B	4	10B	*CG11381*, *CG13113*, *CG13114*[a]
34B	6	10B, 13	*Vm34Ca*
62D	4	10B, 13	*yellow-g*, *yellow-g2*
66D	60–80	10B–11	*Cp18*, *Cp15*, *Cp19*, *Cp16*

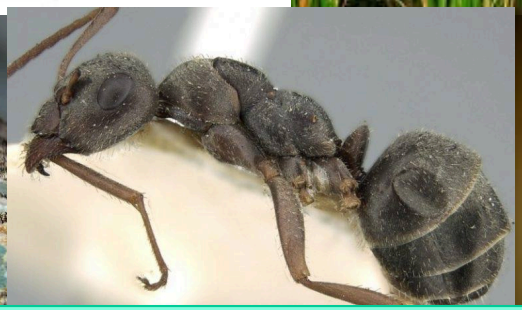
[a] Predicted chorion genes (Fakhouri *et al.* 2006; Tootle *et al.* 2011).

Transposable elements: Selfish genes









Toxin-Antidote Elements Across the Tree of Life.Burga A, Ben-David E, Kruglyak L
Annu Rev Genet. 2020 Nov 23;54:387-415. doi: 10.1146/annurev-genet-112618-043659.

ANNUAL REVIEWS

Annual Review of Genetics

Toxin-Antidote Elements Across the Tree of Life

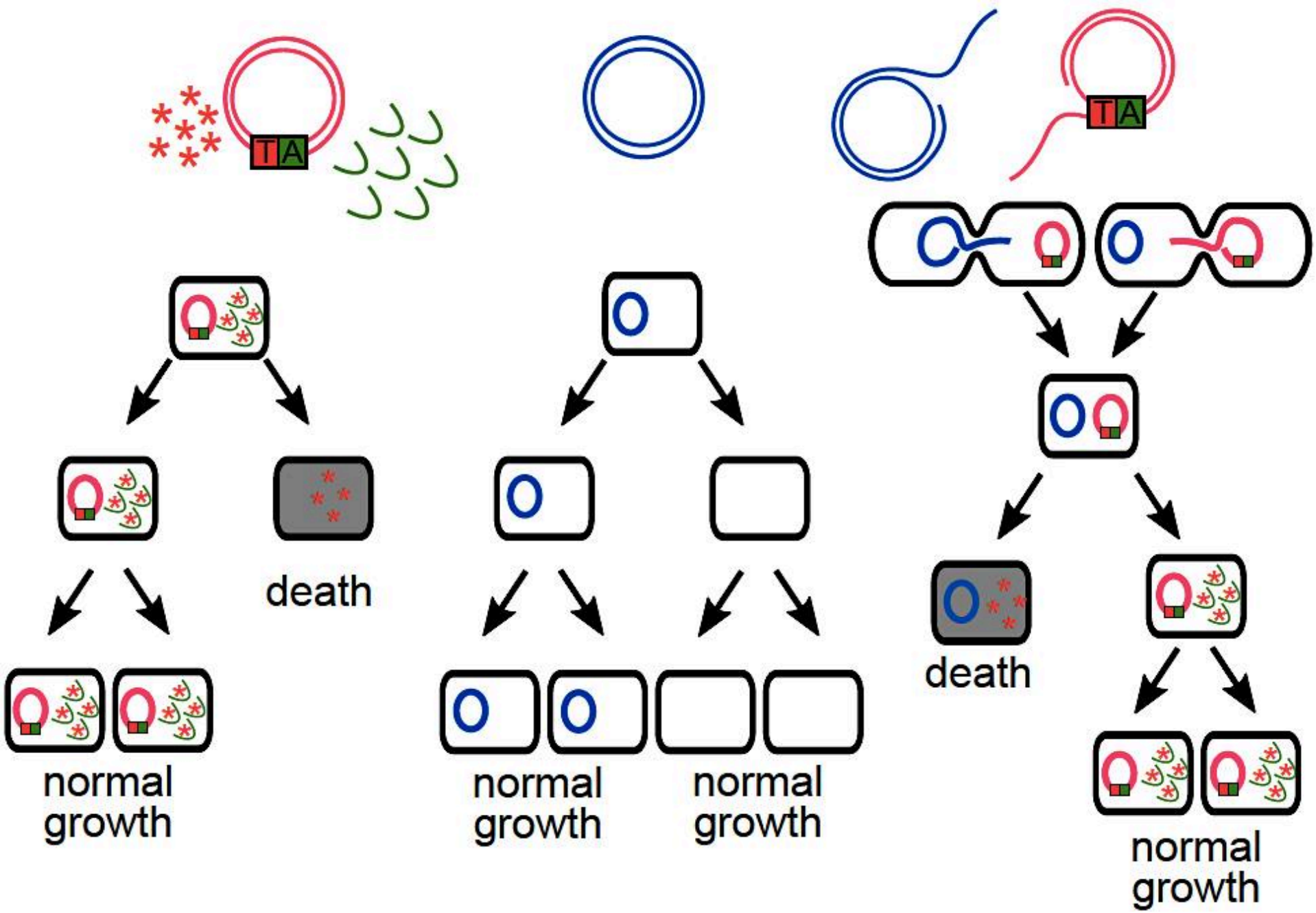
Alejandro Burga,[1,*] Eyal Ben-David,[2,3,*] and Leonid Kruglyak[2]

[1]Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna BioCenter (VBC), 1030 Vienna, Austria; email: alejandro.burga@imba.oeaw.ac.at

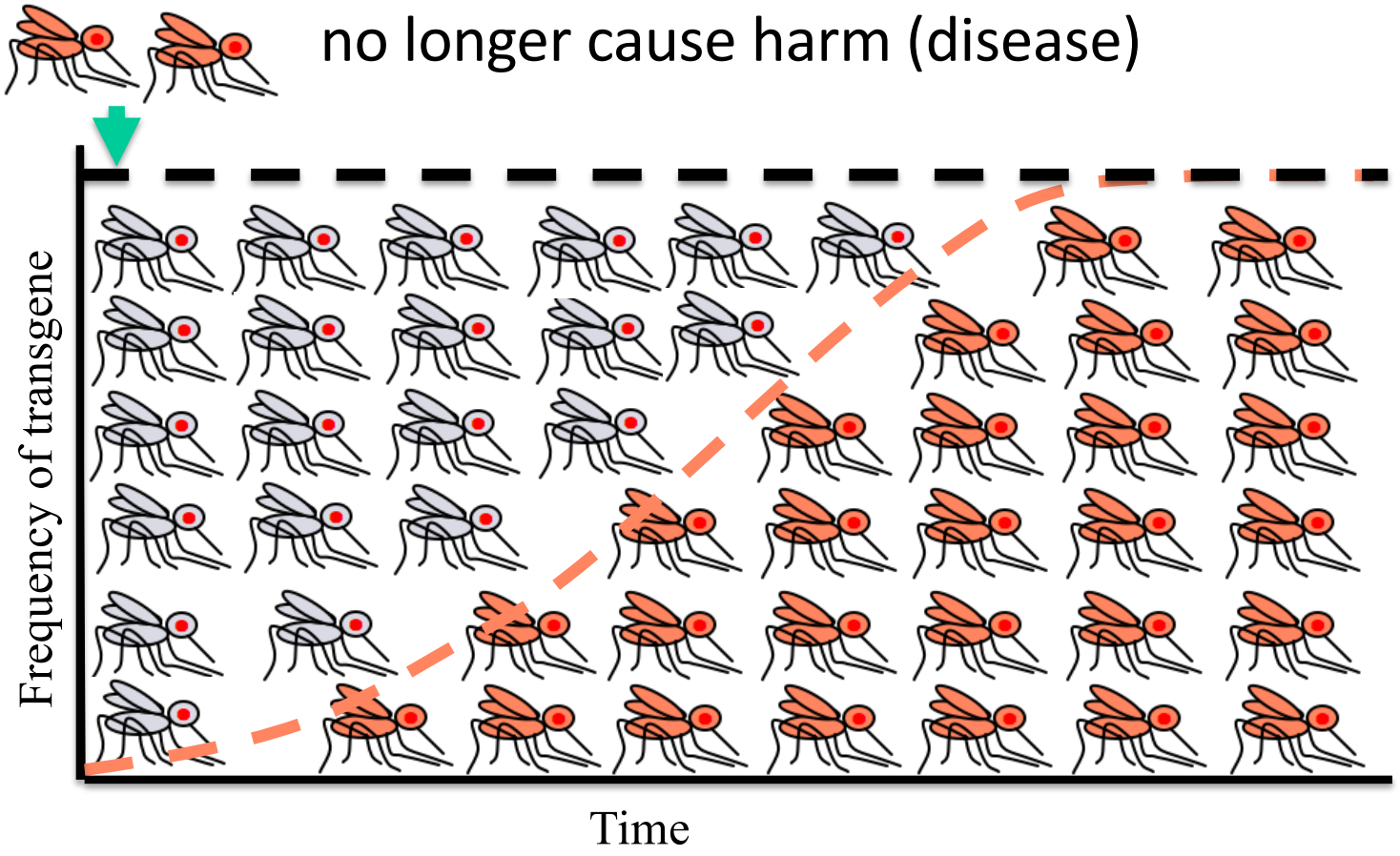
[2]Department of Human Genetics, Department of Biological Chemistry, and Howard Hughes Medical Institute, University of California, Los Angeles, California 90095, USA; email: lkruglyak@mednet.ucla.edu

[3]Department of Biochemistry and Molecular Biology, Institute for Medical Research Israel-Canada, The Hebrew University School of Medicine, Jerusalem 91120, Israel; email: eyal.bendavid@mail.huji.ac.il

TA cassettes “encourage” their own transmission and kick plasmids of the same incompatibiity group that lack them out of the population

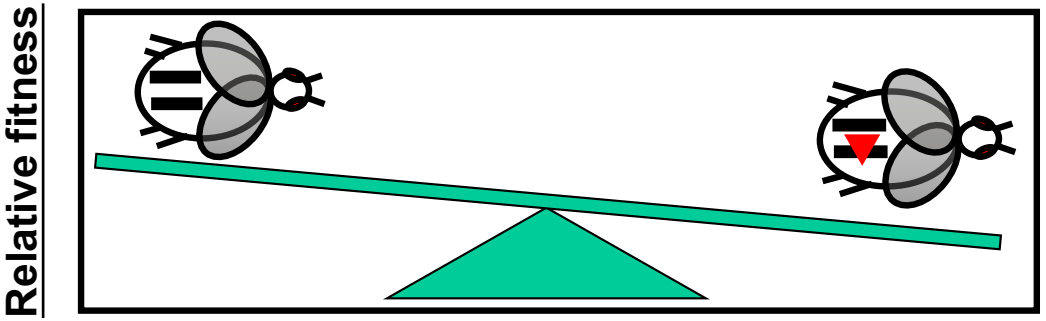


Modify the wild population so that its members no longer cause harm (disease)

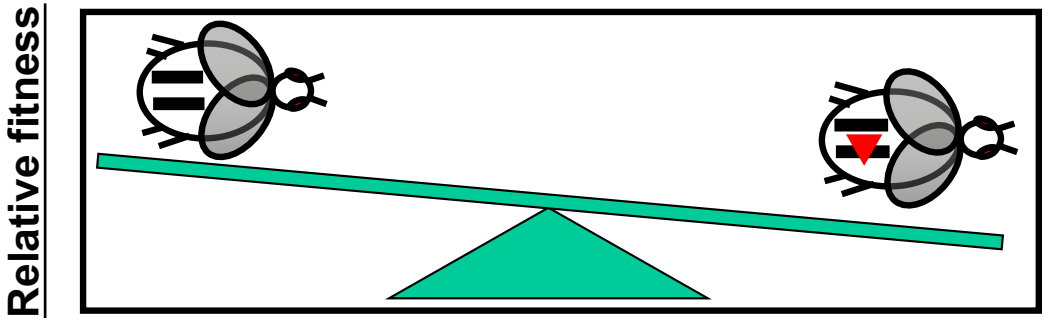


The challenge: How do we cause a trait to rapidly spread throughout a population?

Problem: traits of interest usually result In some fitness cost to carriers

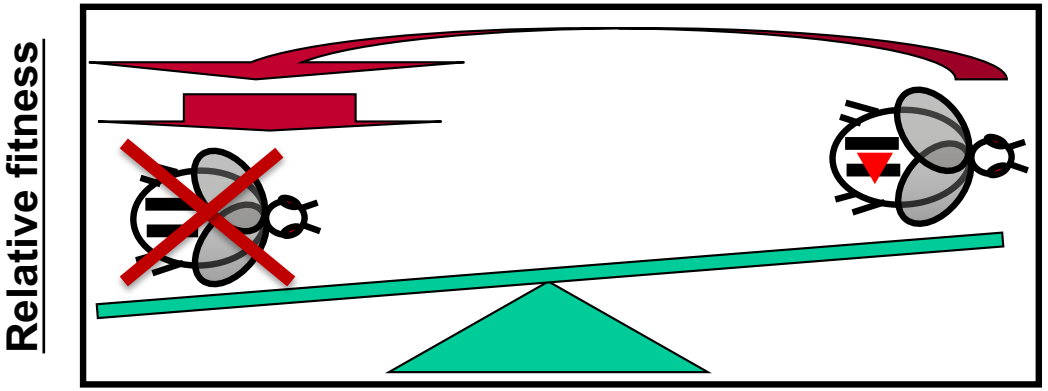


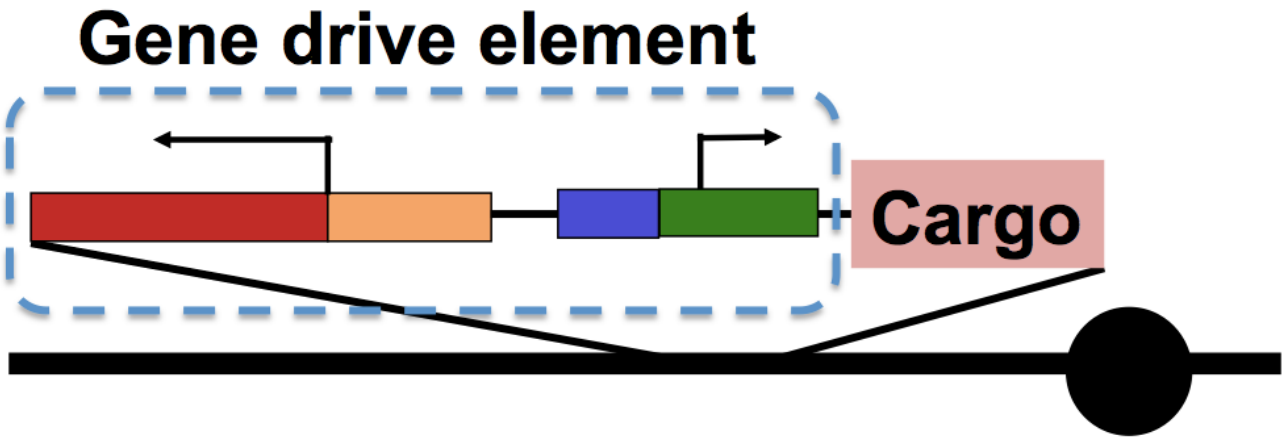
Problem: traits of interest usually result In some fitness cost to carriers



Solution: Increase the fitness cost associated with NOT carrying the gene of interest

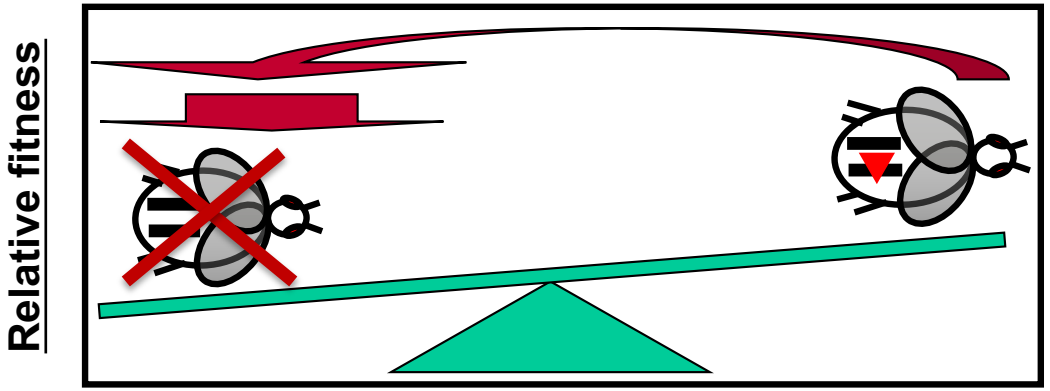
A form of Gene drive

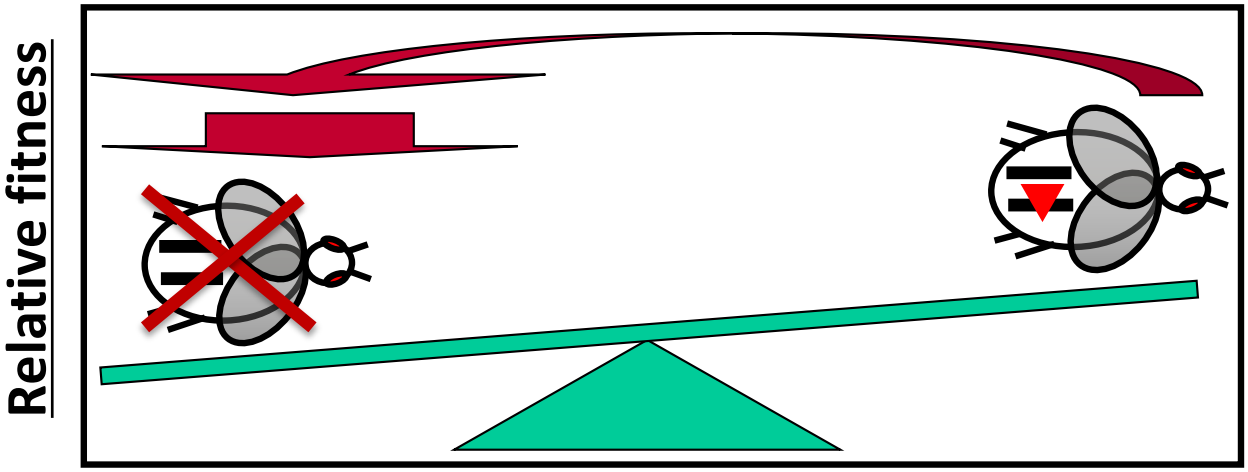




Solution: Increase the fitness cost associated with NOT carrying the gene of interest

A form of Gene drive

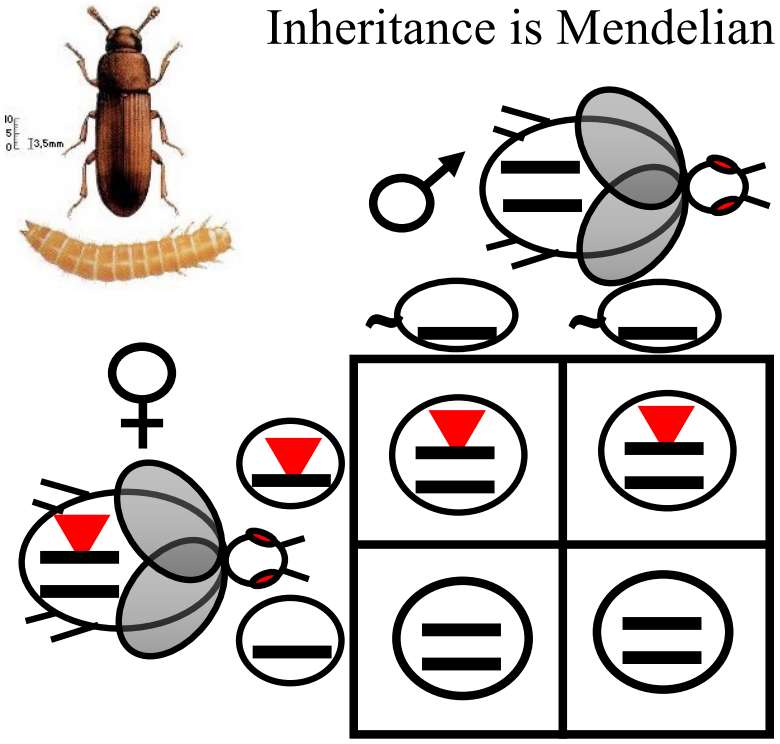




Medea

Cleave and Rescue (*ClvR*)

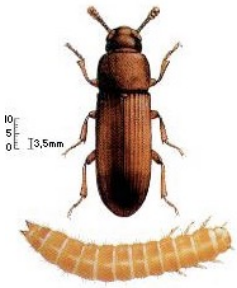
Medea (▼), a novel selfish genetic element



Beeman et al., (1992) Science 256, 89

Medea (▼), a novel selfish genetic element

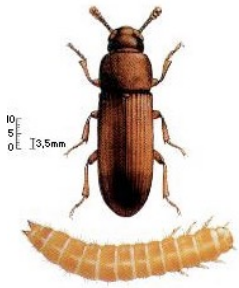
When *Medea* (▼) is present in females, only progeny that inherit *Medea* survive



Beeman et al., (1992) Science 256, 89

Medea (▼), a novel selfish genetic element

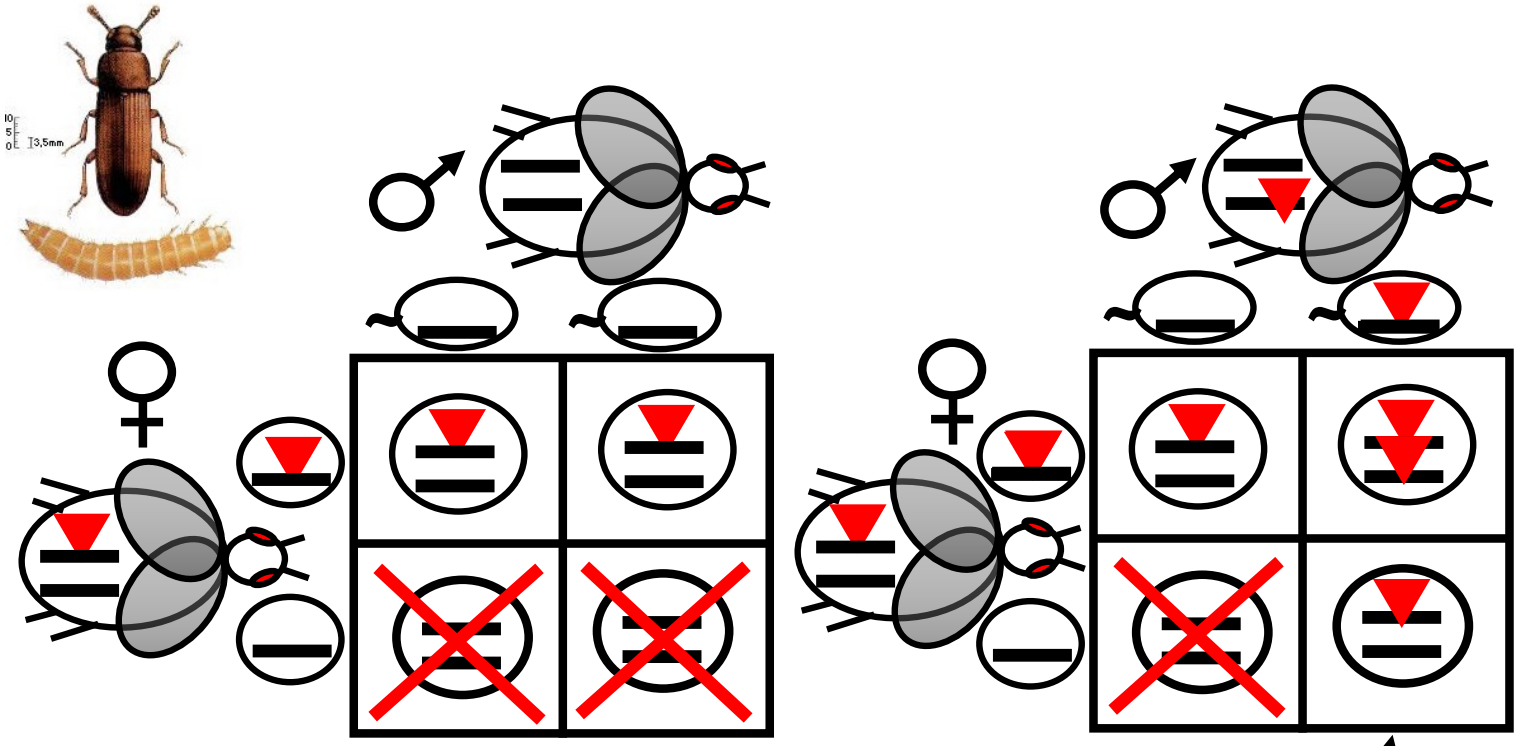
When *Medea* (▼) is present in females, only progeny that inherit *Medea* survive



Beeman et al., (1992) Science 256, 89

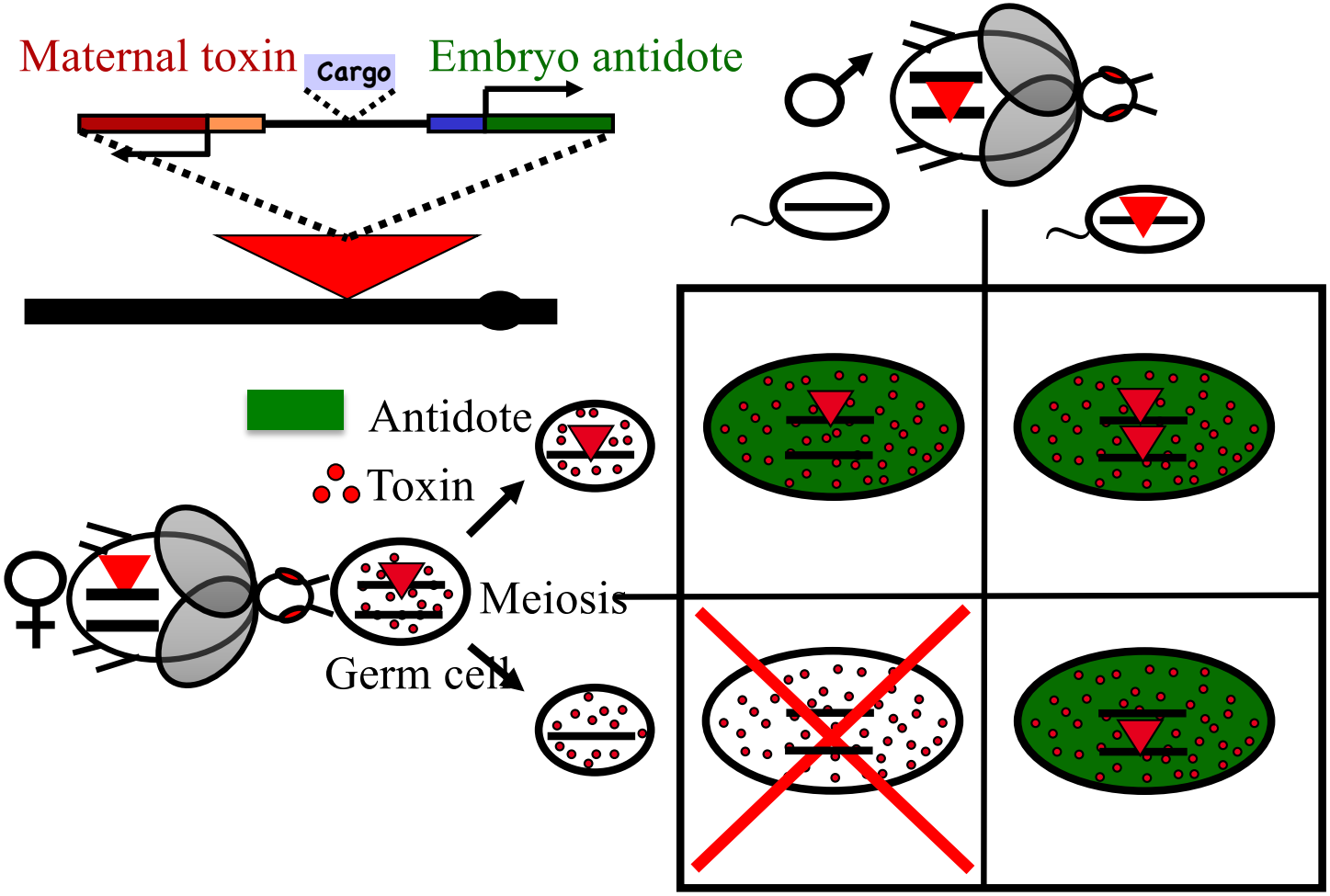
Paternal rescue

***Medea* gains a transmission advantage, not by increasing fitness of those that carry it, but by causing the death of those that lack it.**

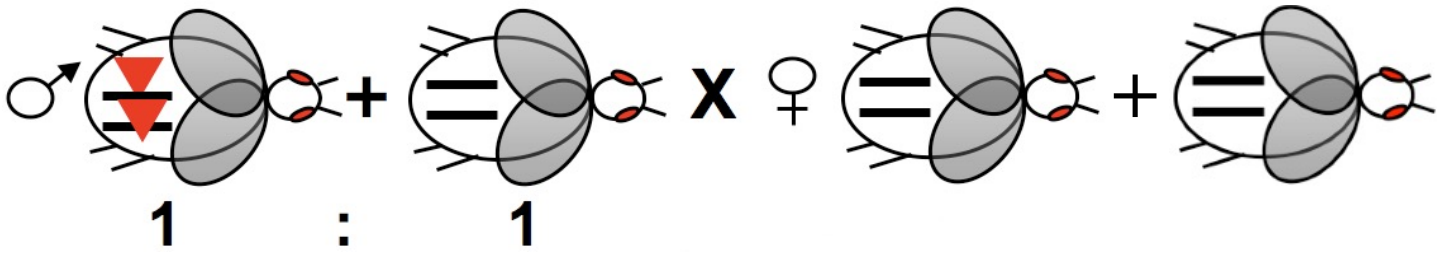


Beeman et al., (1992) Science 256, 89

Medea in insects: A maternal toxin deposited into all eggs.
Only those inheriting the antidote survive



Medea spreads, eliminating non-*Medea* individuals



Fraction *Medea* individuals in population

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

0
5
10
15
20

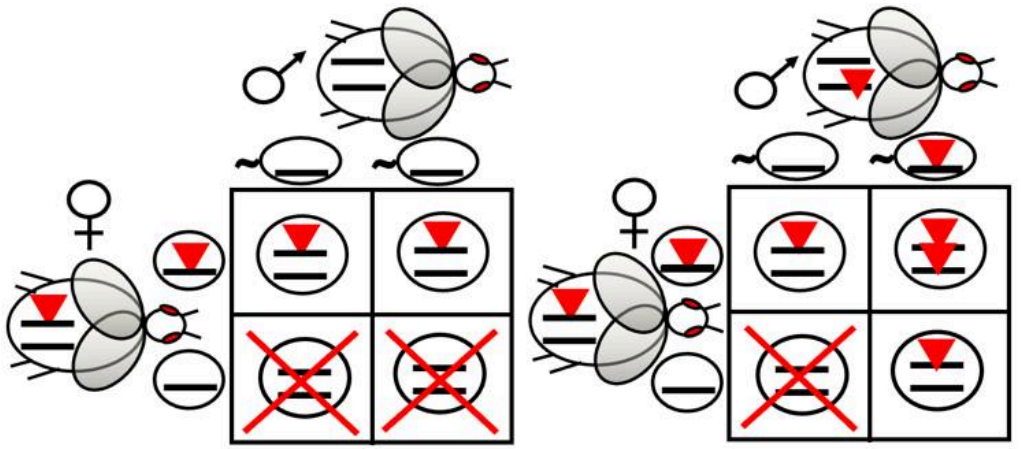
Generation

Fitness cost = 0

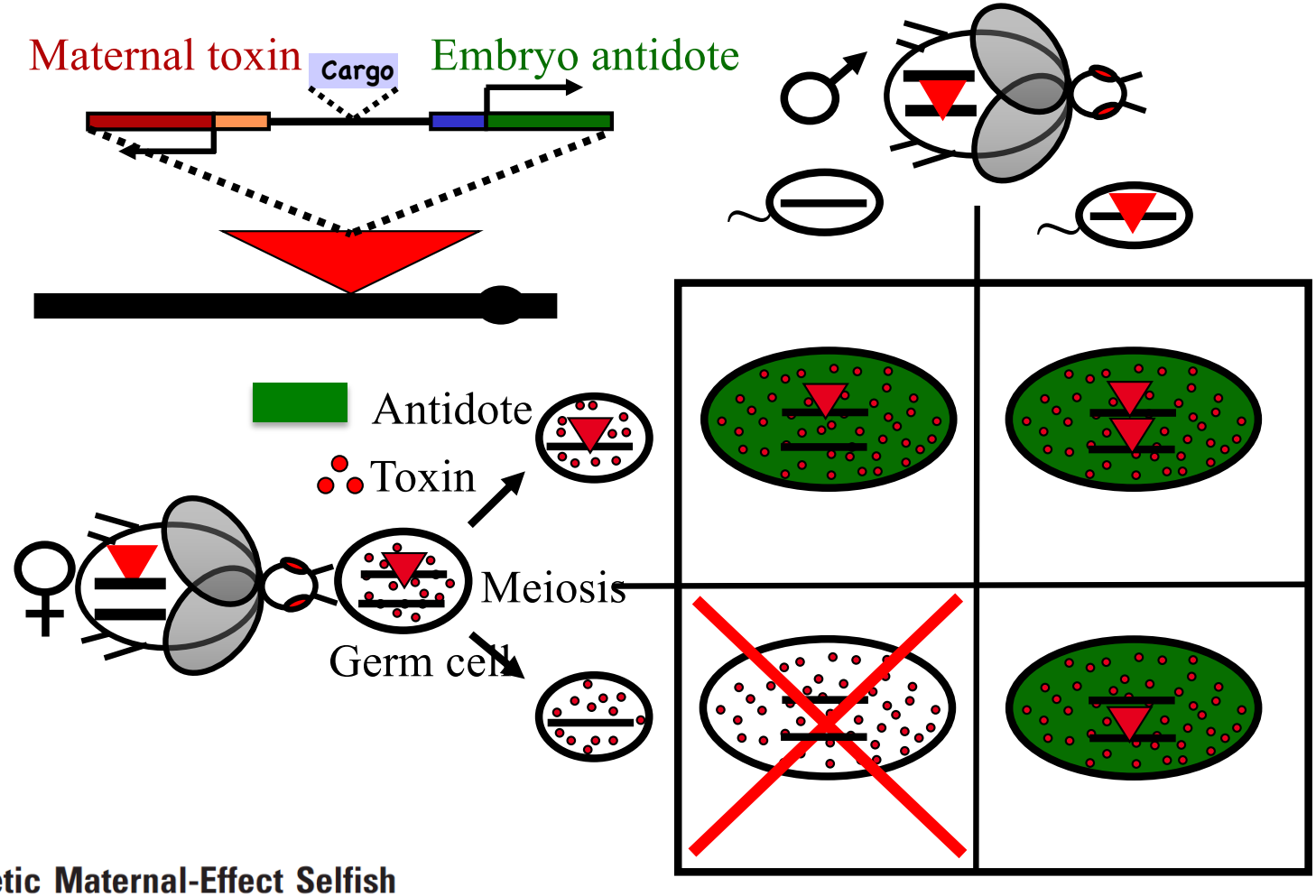
Fitness cost = 0.05

Fitness cost = 0.10

Fitness cost = 0.15



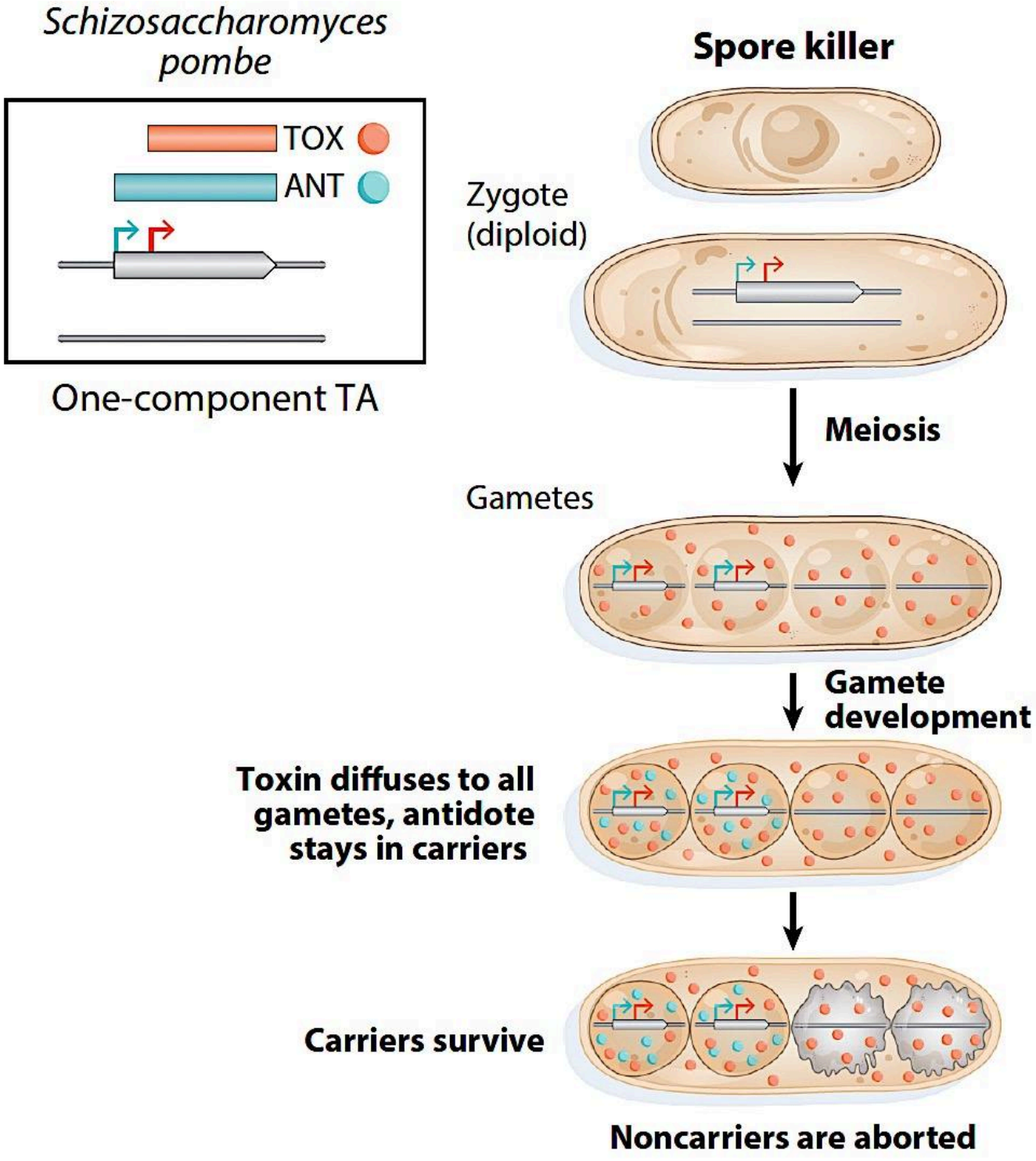
Medea in insects: A maternal toxin deposited into all eggs.
Only those inheriting the antidote survive



A Synthetic Maternal-Effect Selfish Genetic Element Drives Population Replacement in *Drosophila*

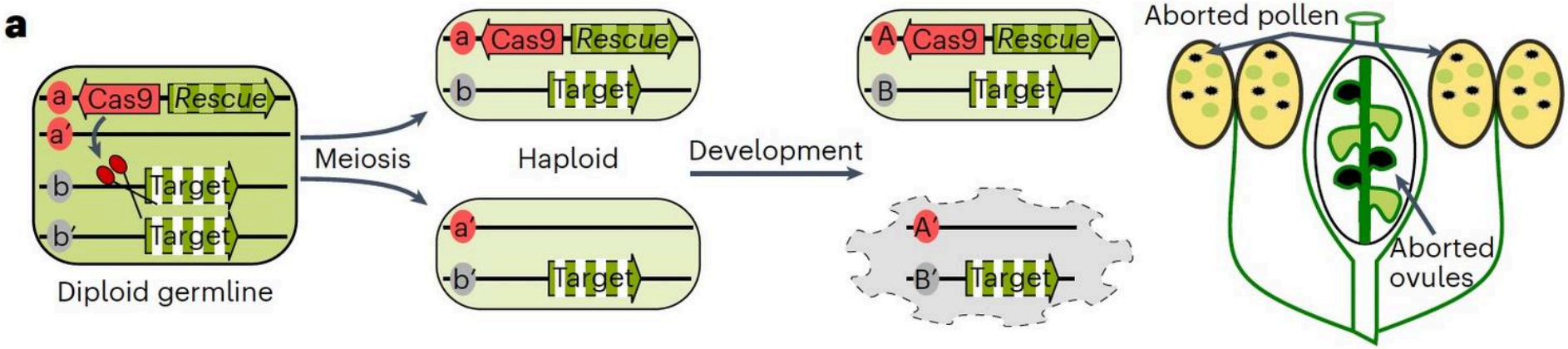
Chun-Hong Chen,[1] Haixia Huang,[1] Catherine M. Ward,[1] Jessica T. Su,[1] Lorian V. Schaeffer,[1] Ming Guo,[2] Bruce A. Hay[1]*

Science
AAAS


Schizosaccharomyces pombe
TOX
ANT
One-component TA
Spore killer
Zygote (diploid)
Meiosis
Gametes
Gamete development
Toxin diffuses to all gametes, antidote stays in carriers
Carriers survive
Noncarriers are aborted

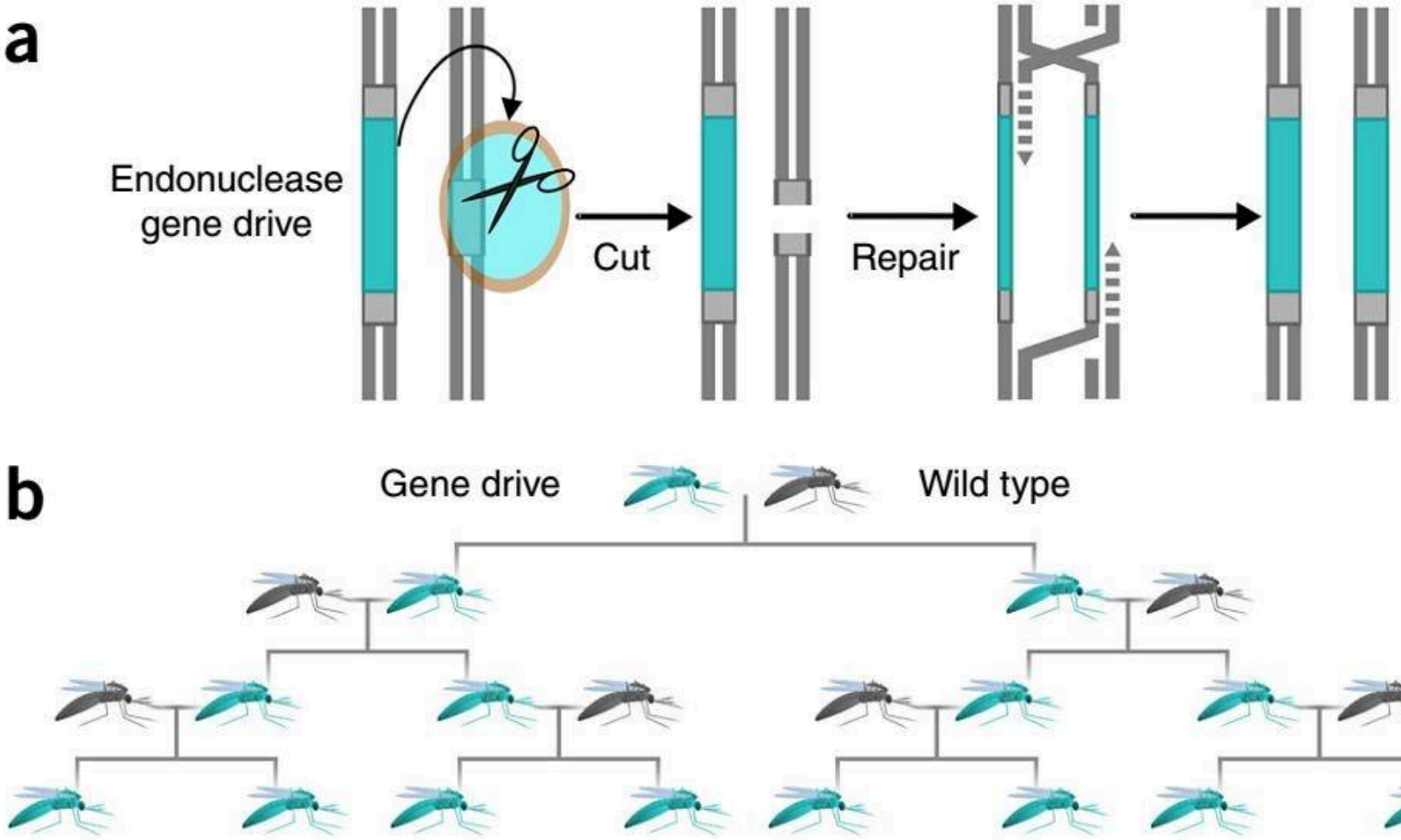
Cleavage and creation of loss-of-function (lof) alleles in the target (essential gene) occurs before meiosis. In this way both copies of the target gene are rendered non-functional before the first meiotic division segregates the vector-bearing chromosome into one gamete (which survives) and the non-vector-bearing gamete into the other one, which dies.

The promoters driving Cas9/gRNA function express during the diploid stage. This can either be in all cells (ubiquitous promoter), tissues that will give rise to the germline or during the diploid stages of meiosis, before the first meiotic division occurs, which separates the homologous chromosomes into haploid gametes

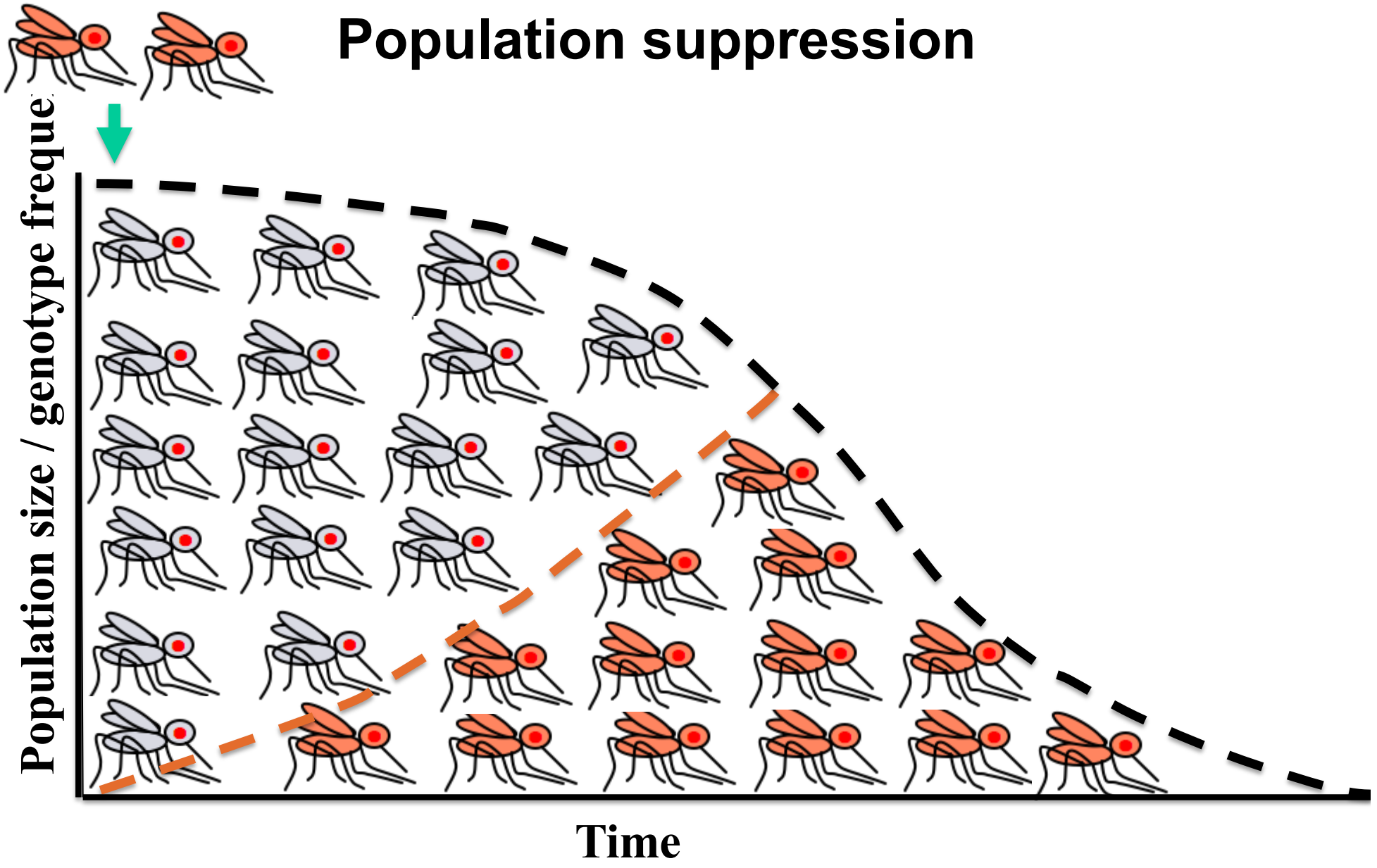


https://doi.org/10.1038/s41477-024-01701-3

Endonuclease cleaves the homologous chromosome at the site where the endonuclease gene is located. The gene drive element disrupts the target site. Thus, in heterozygotes only the sequence on the WT chromosome gets cleaved. If this break is repaired via HR in the germline then the drive element increases in frequency from 1 to 2.



Eliminating a population, at least locally

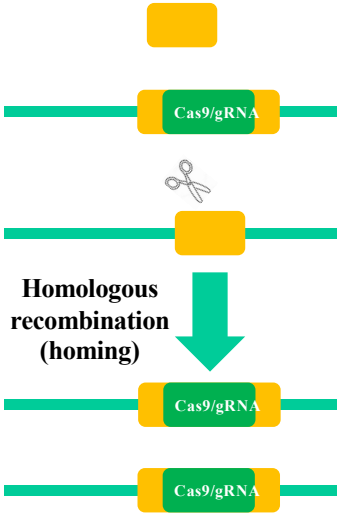


Over-replication: Homing/gene conversion...and suppression

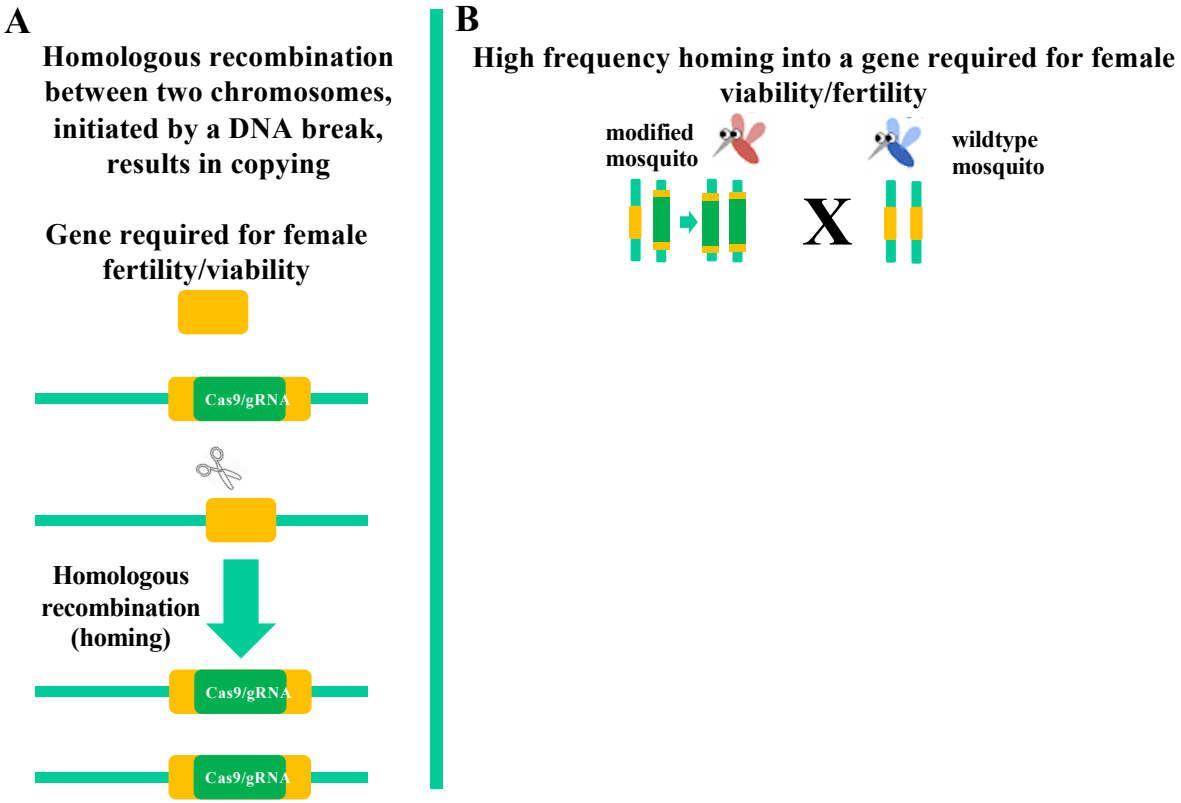
A

Homologous recombination between two chromosomes, initiated by a DNA break, results in copying

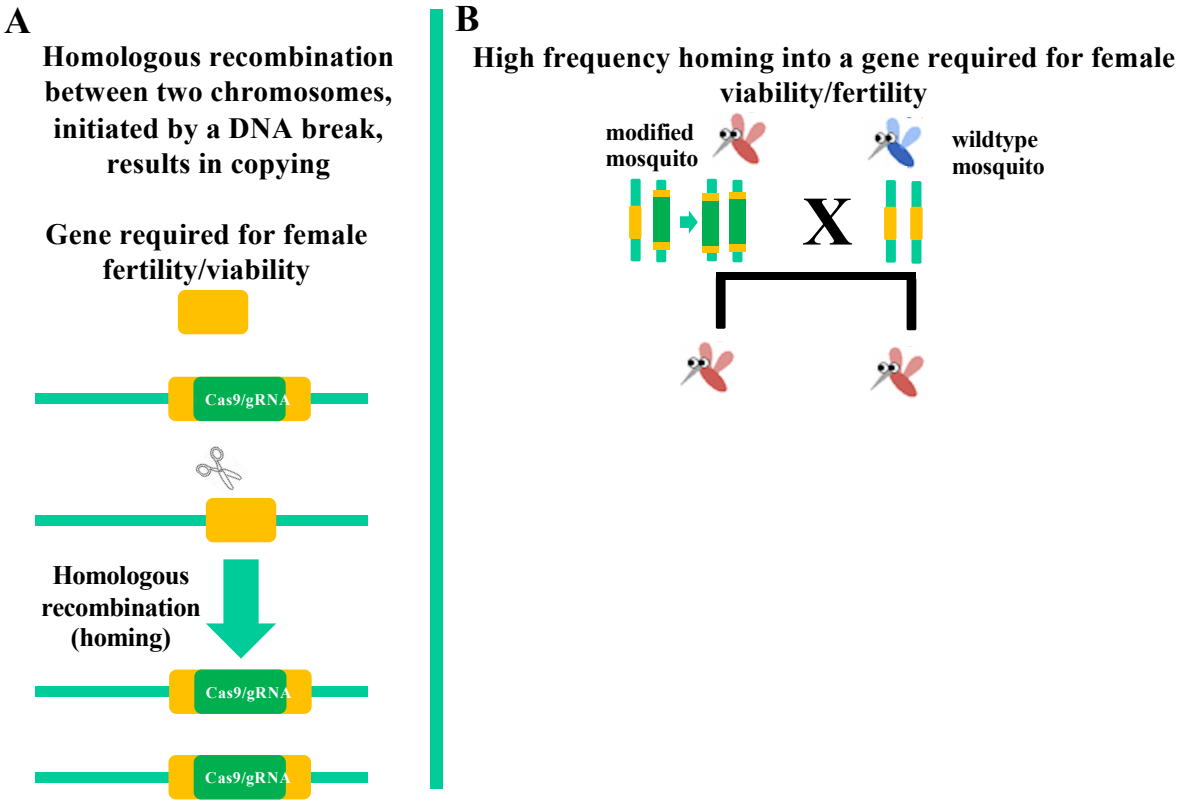
Gene required for female fertility/viability



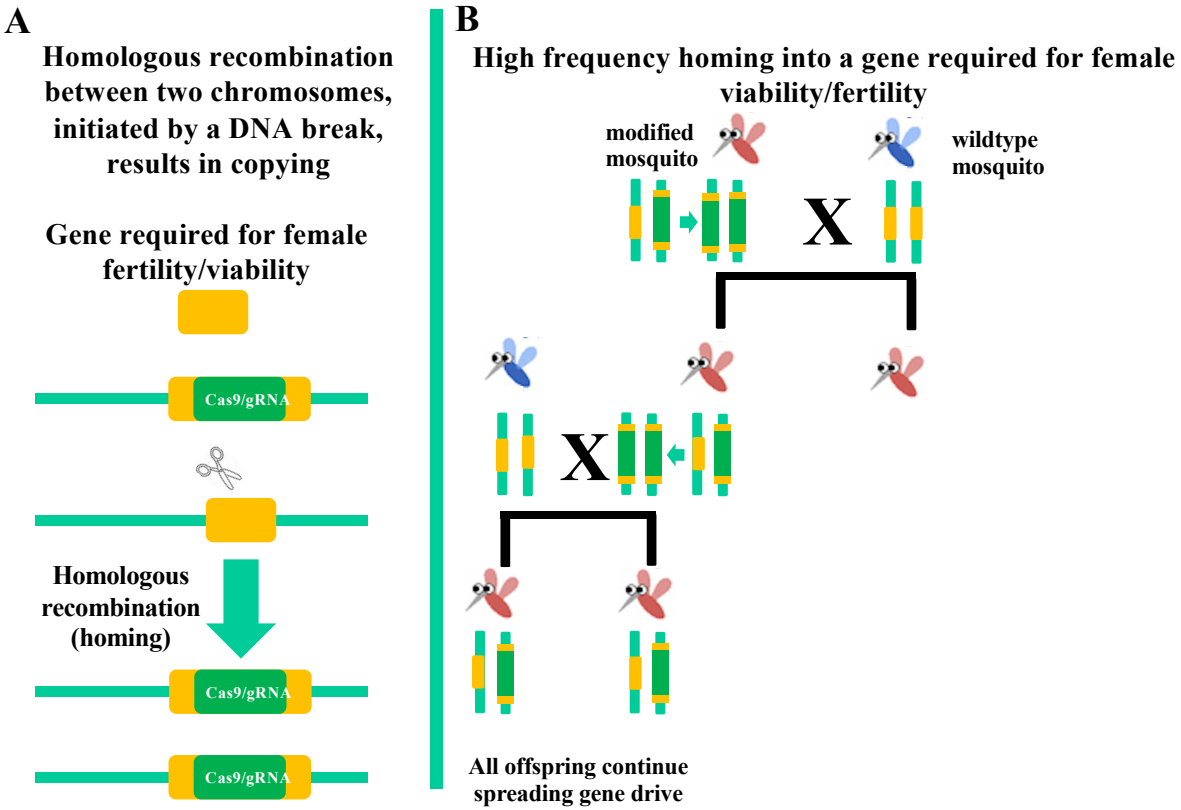
Over-replication: Homing/gene conversion...and suppression



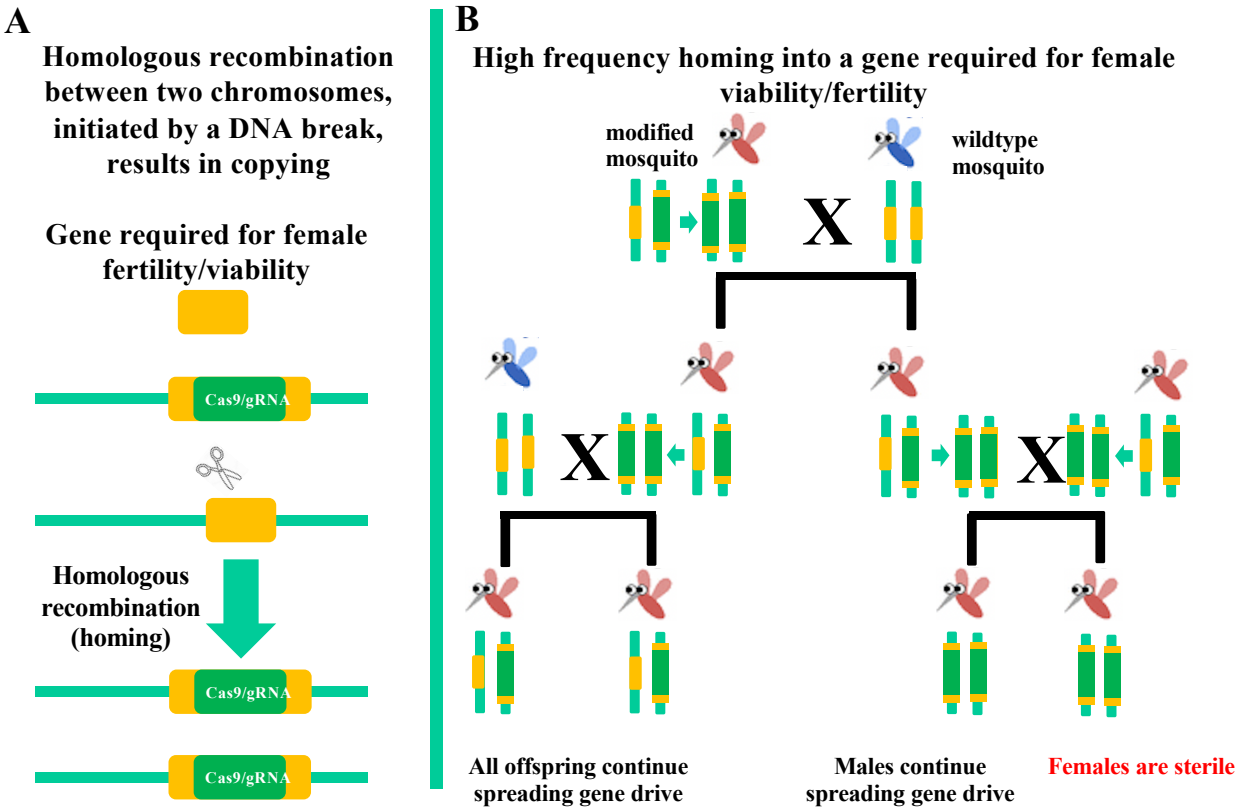
Over-replication: Homing/gene conversion...and suppression



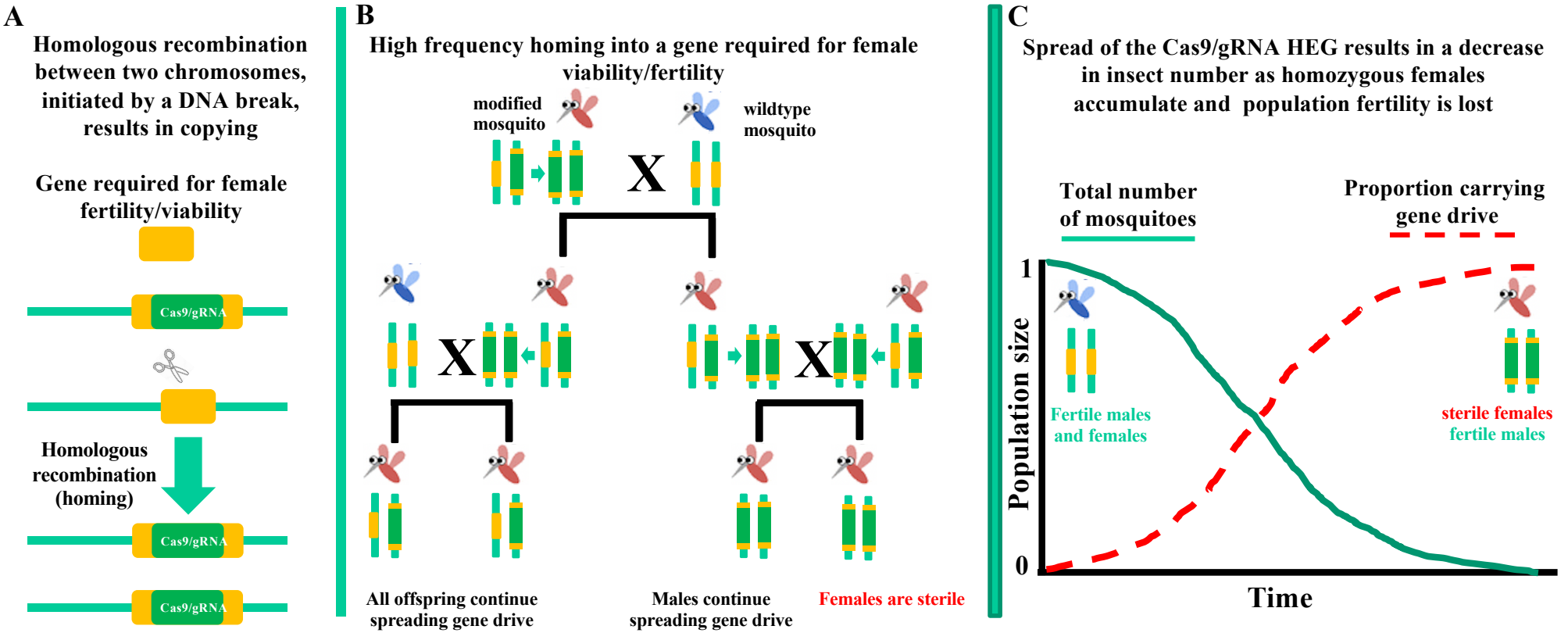
Over-replication: Homing/gene conversion...and suppression



Over-replication: Homing/gene conversion...and suppression

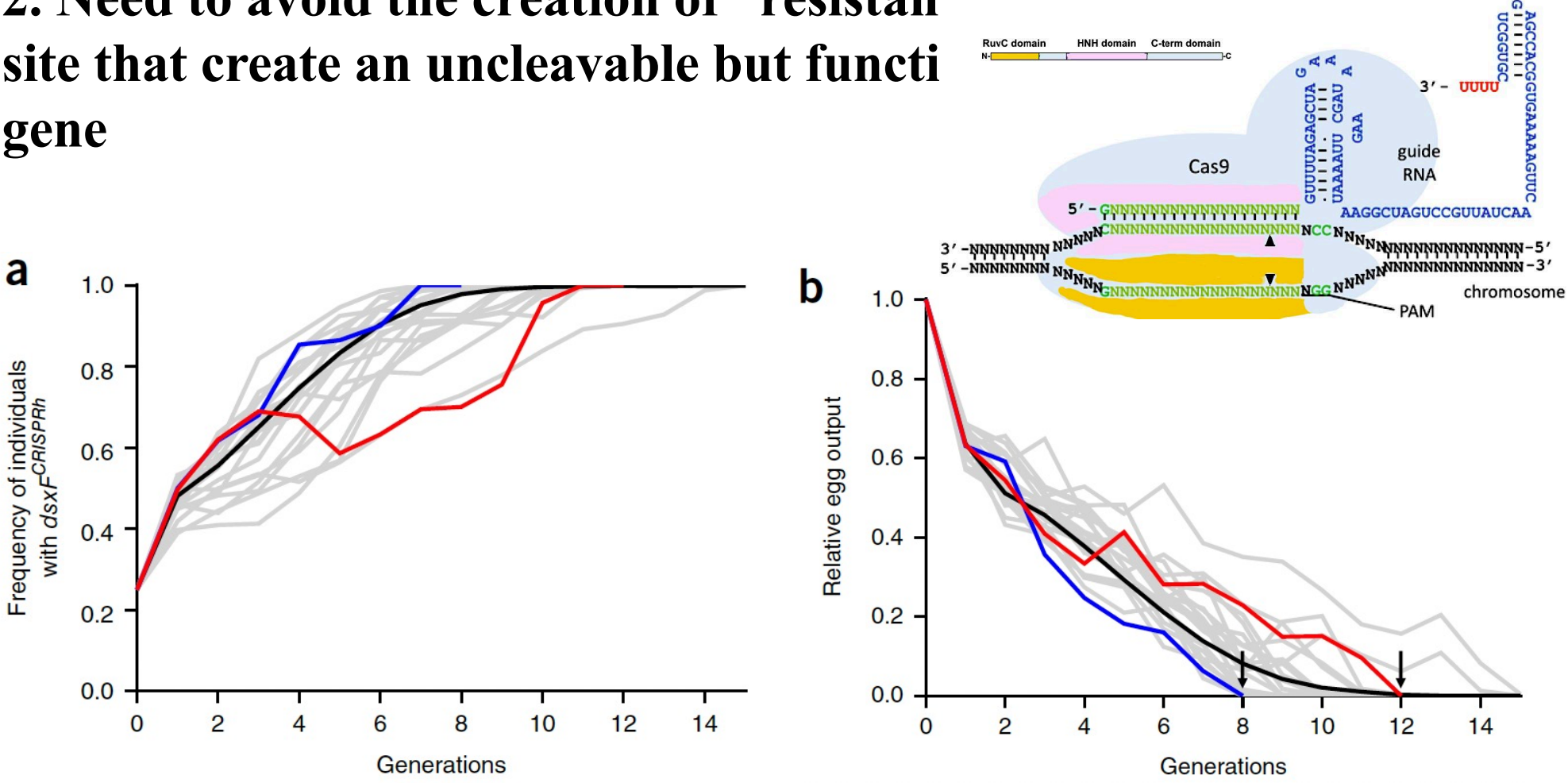


Over-replication: Homing/gene conversion...and suppression

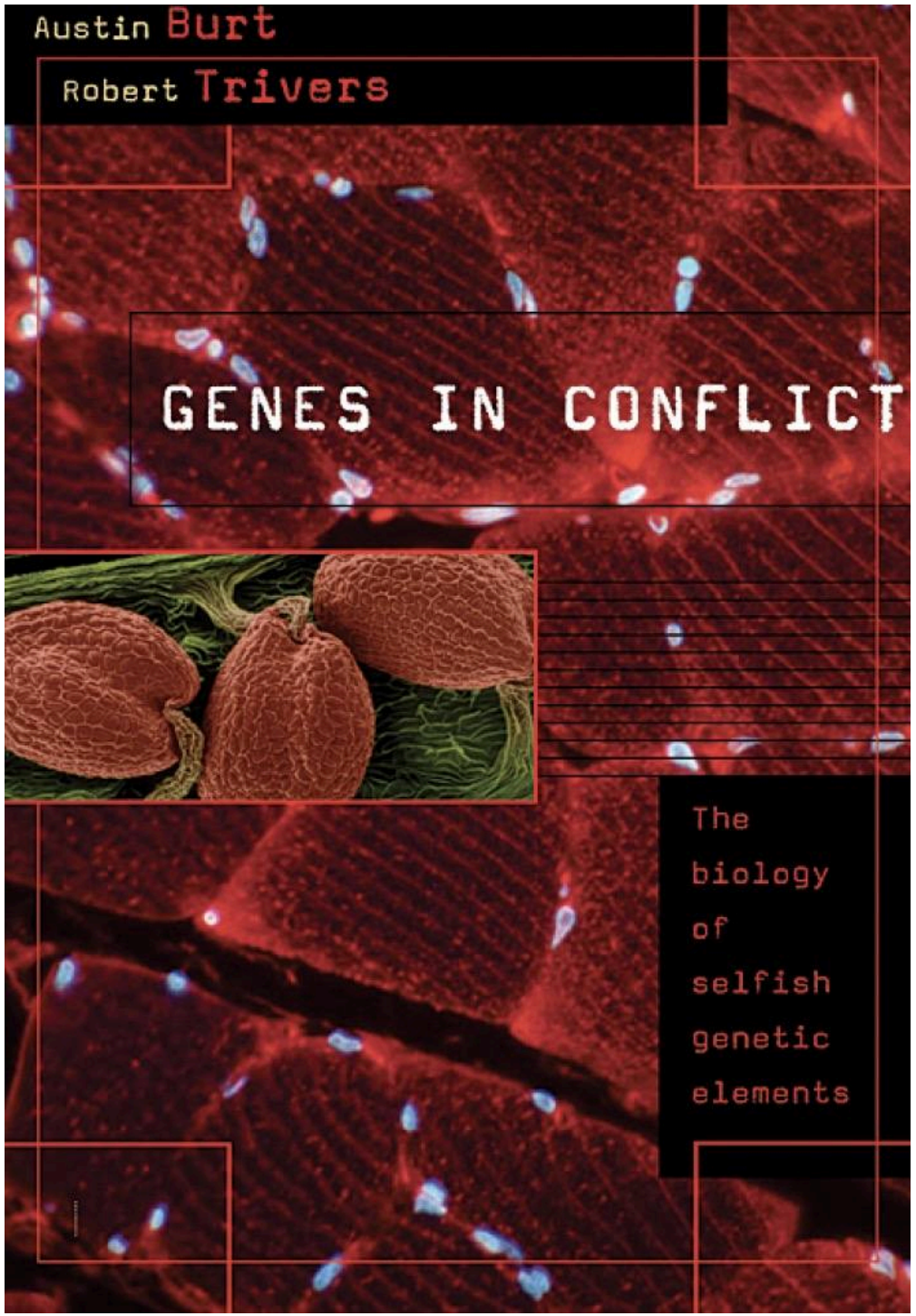


1. Requires that homing rates in the germline are very high

2. Need to avoid the creation of "resistance" alleles at the target site that create an uncleavable but functi gene



A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. Nat Biotechnol. 2018 Dec;36(11):1062-1066. doi: 10.1038/nbt.4245. Epub 2018 Sep 24.


Austin Burt
Robert Trivers
GENES IN CONFLICT
The biology of selfish genetic elements

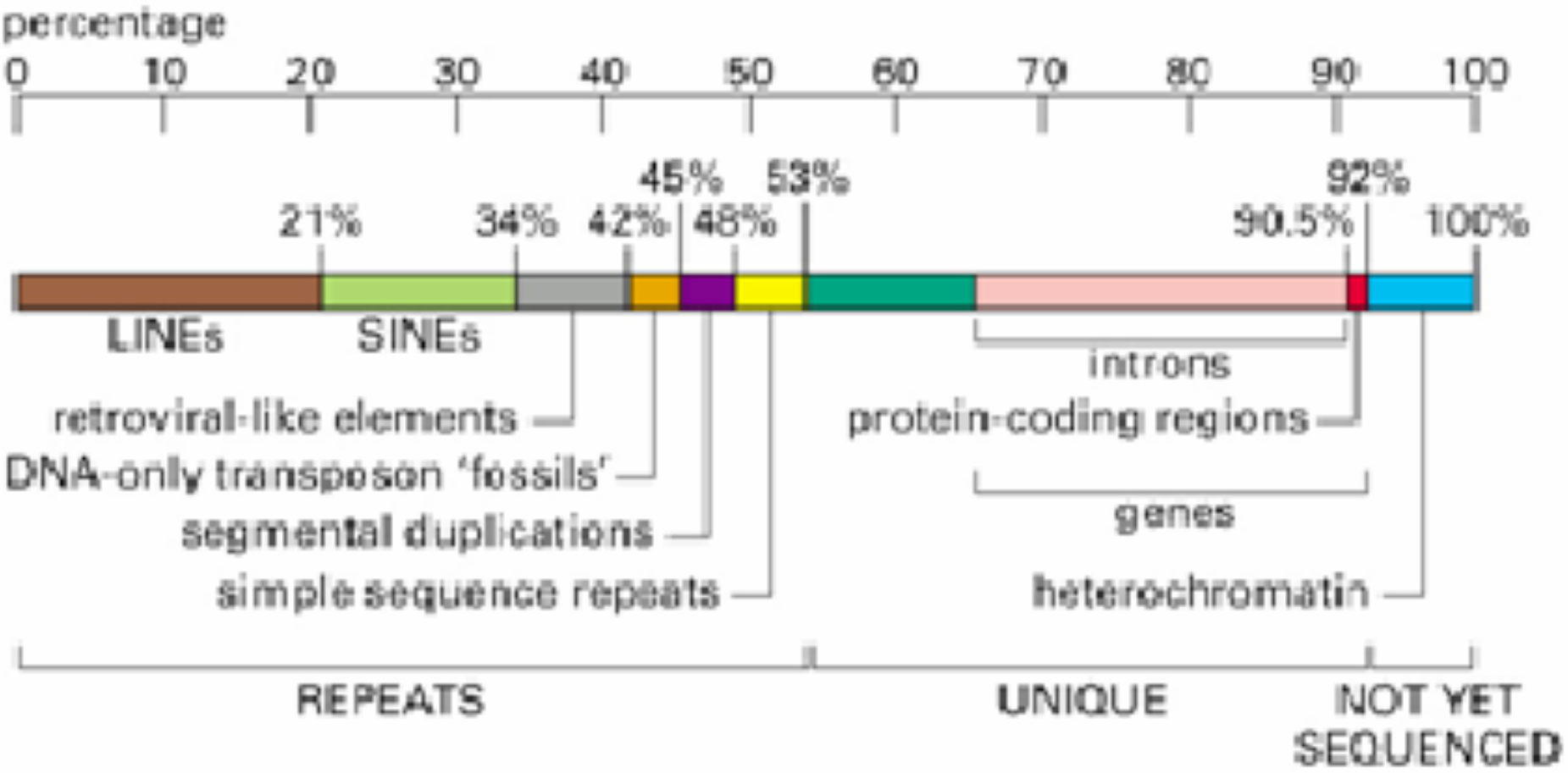


Figure 4–17. Molecular Biology of the Cell, 4th Edition.

Repetitive DNA sequences in the human genome are the consequences of transposition.

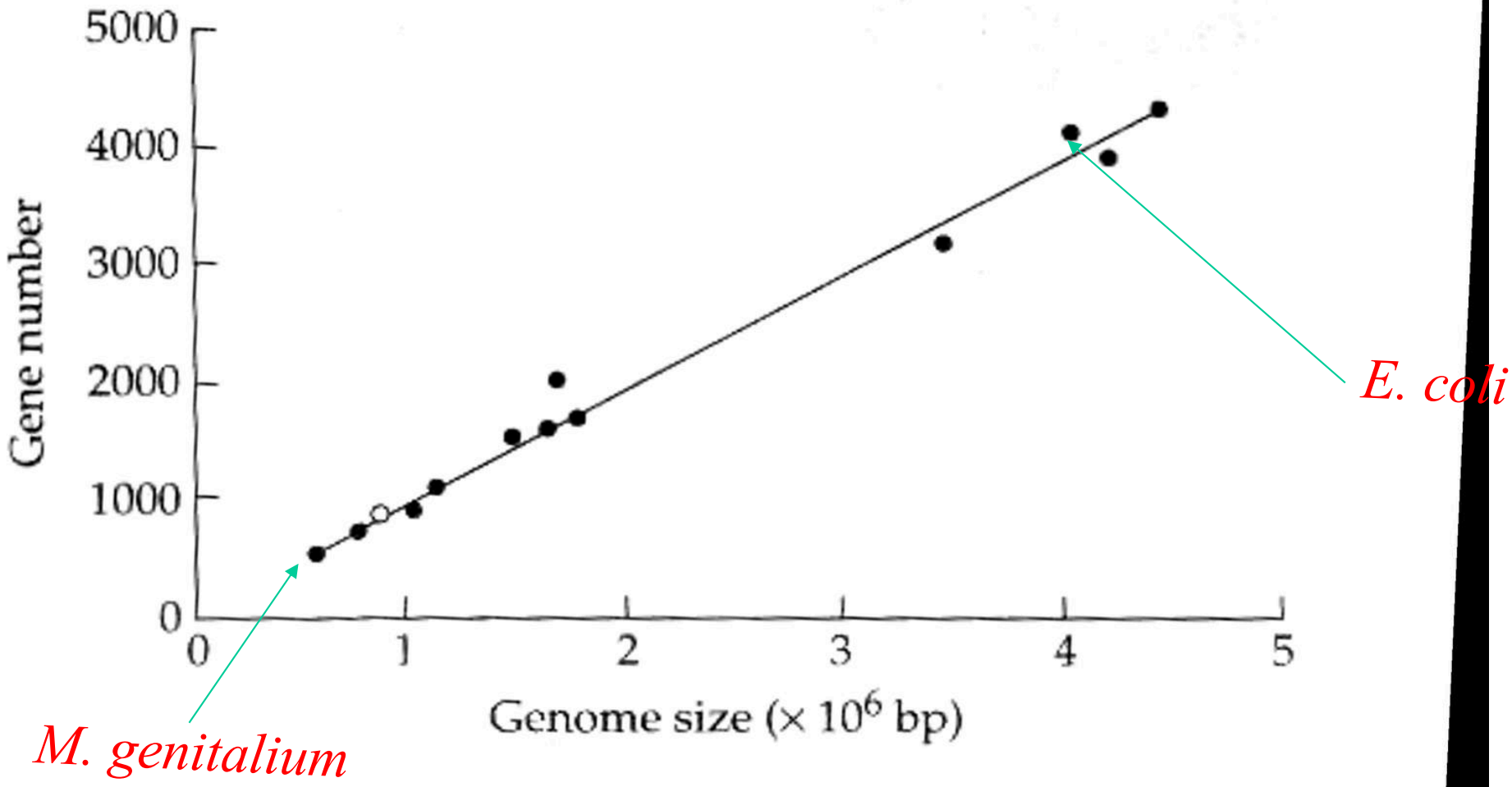
Range of prokaryotic genome sizes

	Range in genome size (kb)	Ratio (highest/lowest)
Eubacteria	580–13,200	20
Mycoplasmas	580–1,800	3
Gram negative	650–7,800	12
Gram positive	1,600–11,600	7
Cyanobacteria	3,100–13,200	4
Archaebacteria	1,600–4,100	3

From Cavalier-Smith (1985), updated by knowledge of the *Mycoplasma genitalium* (Fraser et al. 1995).

Graur & Li. Fundamentals of Molecular Evolution (1999)

Genome size in prokaryotes is correlated with gene number

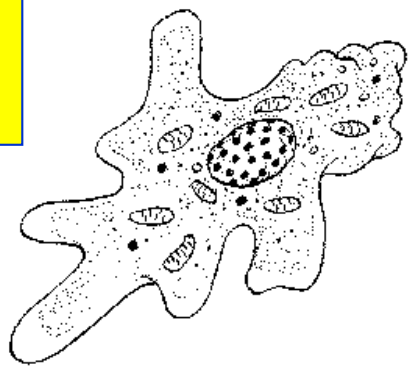


Graur & Li. Fundamentals of Molecular Evolution (1999)

The genome size, or 'C-value', of an organism is defined as the total amount of DNA contained within its haploid chromosome set.

	(Gb)
H. sapiens	2.9
M. musculus	2.5
D. melanogaster	0.18
C. elegans	0.097
S. cerevisiae	0.012

the single-celled amoeba, one of the simplest of eukaryotic creatures, has a genome size of ~200,000 Mb!



The range of genome sizes in various eukaryotic groups

Taxon	*Genome size range (Kb)*	*Ratio*
All eukaryotes	8,800–686,000,000	77,955
Alveolata	23,500–201,000,000	8,553
Apicomplexians	9,400–201,000,000	21,383
Ciliates	23,500–8,620,000	367
Dinoflagellates	1,370,000–98,000,000	72
Diatoms	35,300–24,500,000	694
Amoebae	35,300–686,000,000	19,433
Euglenozoa	98,000–2,350,000	24
Fungi	8,800–1,470,000	167
Animals	49,000–139,000,000	2,837
Sponges	49,000–53,900	1
Cnidarians	323,000–715,000	2
Aschelminthes	80,000–2,450,000	31
Annelids	882,000–5,190,000	6
Mollusks	421,000–5,290,000	13
Crustaceans	686,000–22,100,000	32
Insects	98,000–7,350,000	75
Echinoderms	529,000–3,230,000	6
Non-vertebrate chordates	157,000–1,470,000	9
Agnathes	637,000–2,790,000	4
Elasmobranchs	1,470,000–15,800,000	11
Bony fishes	340,000–139,000,000	409
Amphibians	931,000–84,300,000	91
Reptiles	1,230,000–5,340,000	4
Birds	1,670,000–2,250,000	1
Mammals	1,700,000–6,700,000	4
Monotremes	3,470,000–3,700,000	1
Marsupials	3,470,000–4,560,000	1
Placentals	1,700,000–6,700,000	4
Plants	50,000–307,000,000	6,140
Algae	80,000–30,000,000	375
Pteridophytes	98,000–307,000,000	3,133
Gymnosperms	4,120,000–76,900,000	17
Angiosperms	50,000–125,000,000	2,500

Eukaryotes vary over a range of 80,000!

Graur & Li. Fundamentals of Molecular Evolution (1999)

The C-value paradox:

What accounts for the often massive, counter-intuitive and seemingly arbitrary differences in genome size observed among eukaryotic organisms?

The fruit fly
Drosophila melanogaster

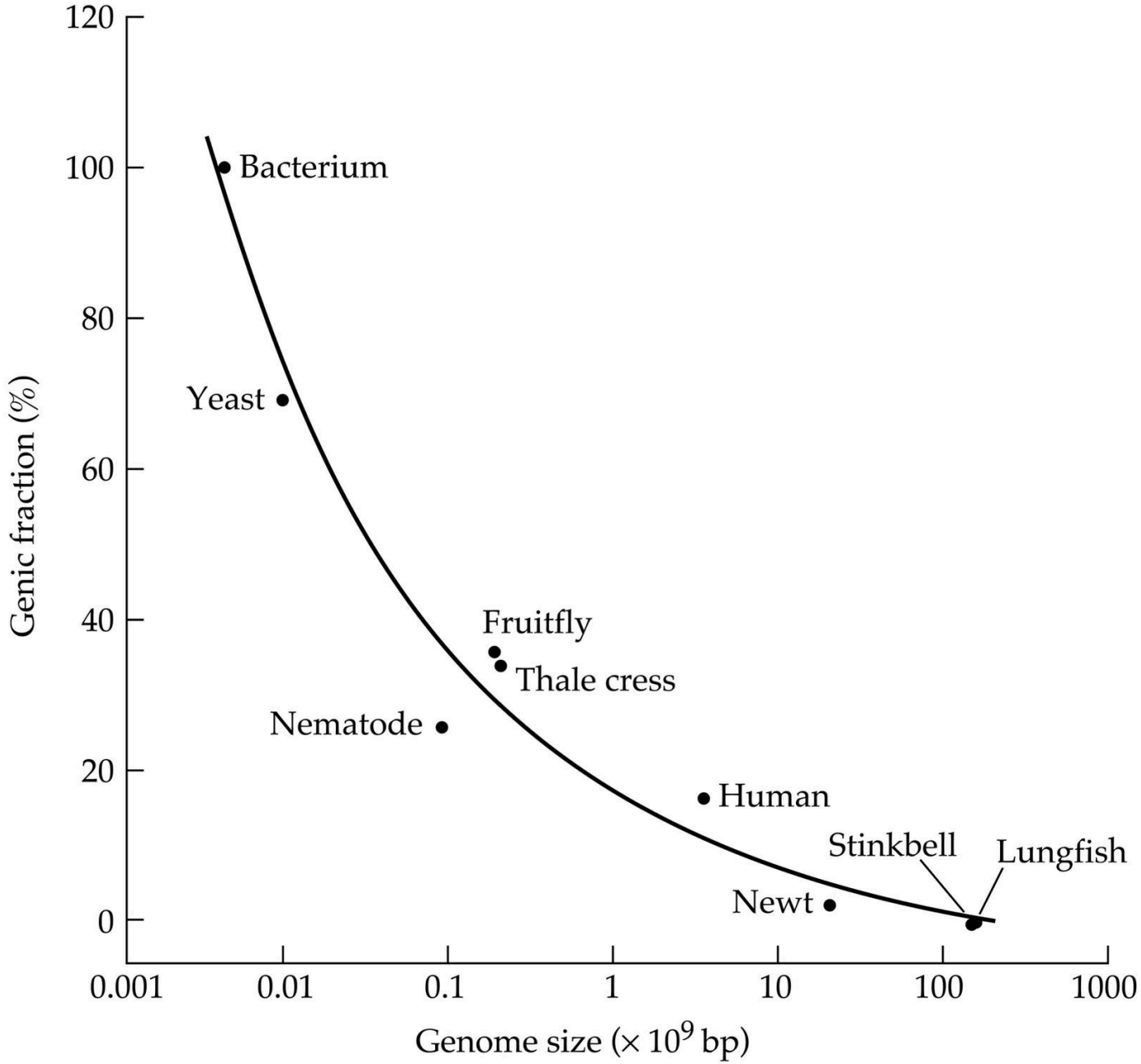
180 Mb

The mountain grasshopper
Podisma pedestris

18,000 Mb

The difference in genome size of a factor of 100 is difficult to explain in view of the apparently similar levels of evolutionary, developmental and behavioral complexity of these organisms.

The key to the C-value paradox lies in the nongenic regions



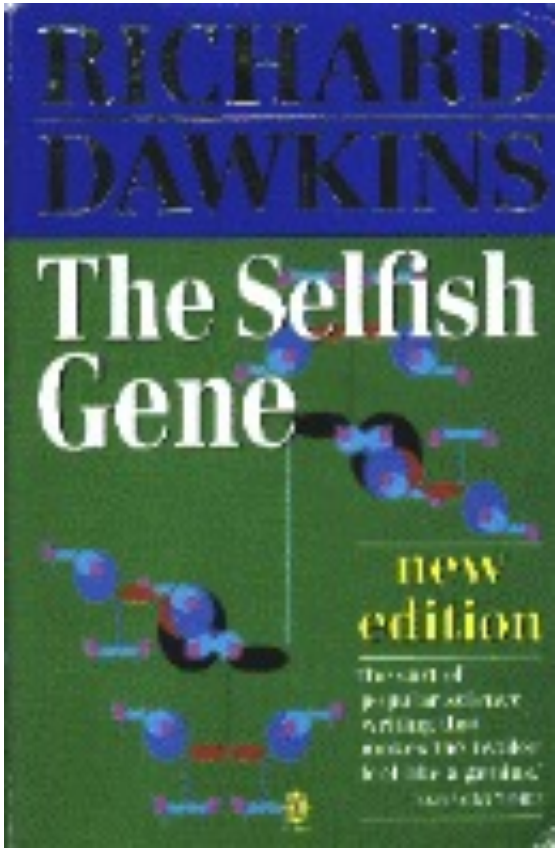
A chicken is just an egg's way of making more eggs

The gene view of evolution:
genes are the basic units of evolution

"Individuals are not stable things, they are fleeting. Chromosomes too are shuffled to oblivion, like hands of cards soon after they are dealt. But the cards themselves survive the shuffling. The cards are the genes. The genes are not destroyed by crossing-over, they merely change partners and march on. Of course they march on. That is their business. They are the replicators and we are their survival machines. When we have served our purpose we are cast aside. But genes are denizens of geological time: genes are forever."

The Selfish Gene
Richard Dawkins

The Selfish Gene
Richard Dawkins



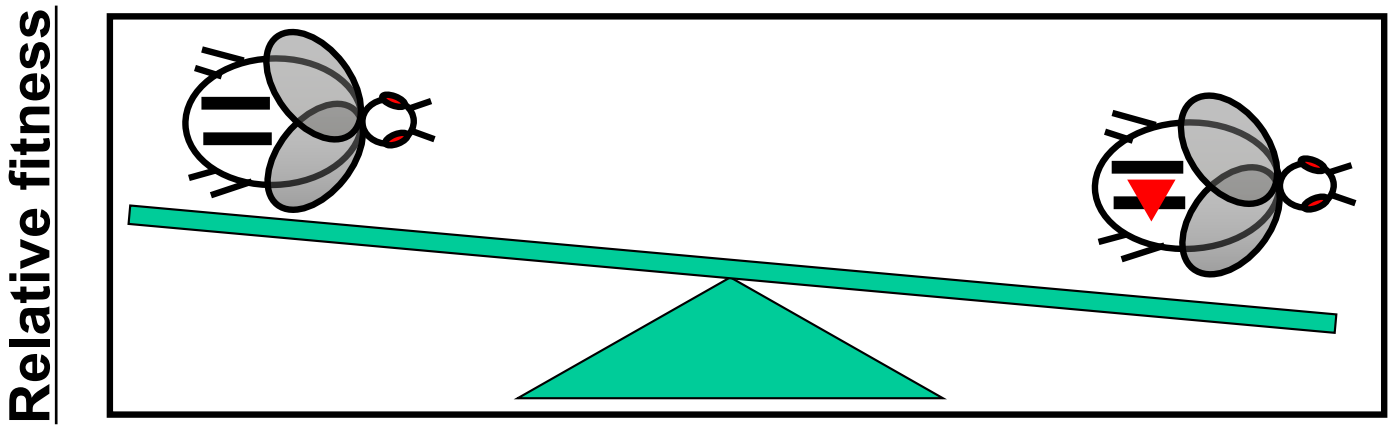
genes are dynamic entities directly subject to evolutionary forces

"Genes are competing directly with their alleles for survival, since their alleles in the gene pool are rivals for their slot on the chromosomes of future generations. Any gene that behaves in such a way as to increase its own survival chances in the gene pool at the expense of its alleles will, by definition, tautologously, tend to survive. The gene is the basic unit of selfishness."

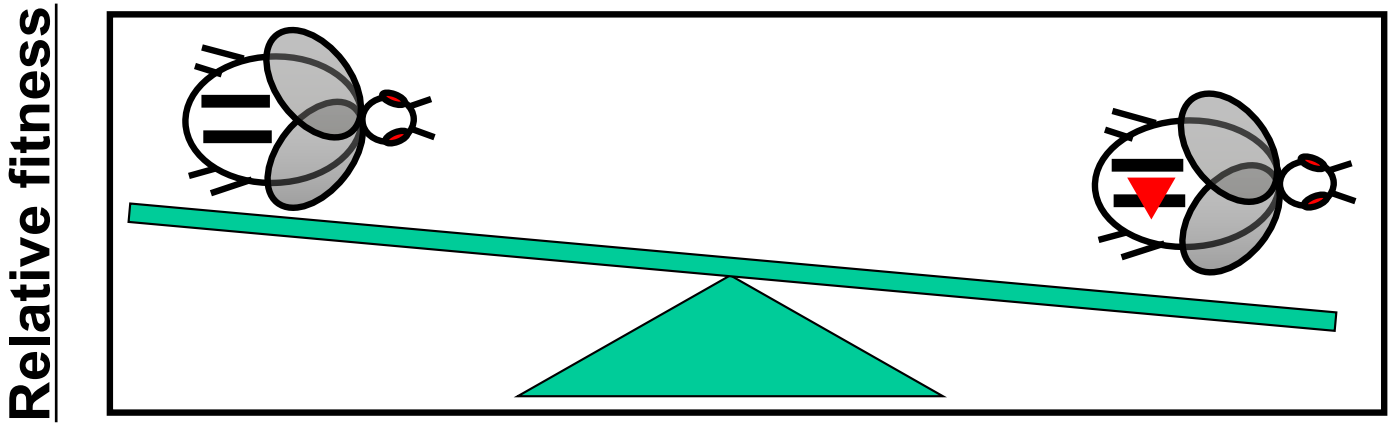
The Selfish Gene
Richard Dawkins

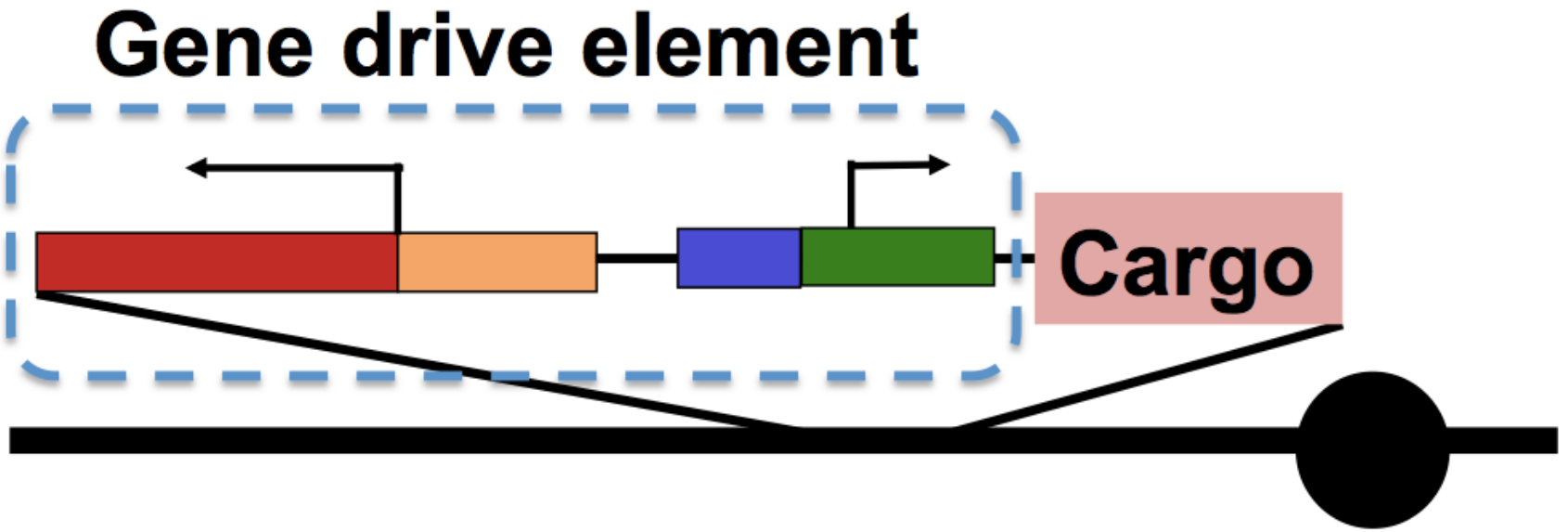
"…it appears that the amount of DNA in organisms is more than is strictly necessary for building them: a large fraction of the DNA is never translated into protein. From this point of view of the individual organism this seems paradoxical. As the 'purpose' of DNA is to supervise the building of bodies, it is surprising to find a large quantity of DNA which does no such thing. Biologists are racking their brains to think what this surplus DNA is doing. But from the point of view of the selfish genes themselves, there is no paradox. The true 'purpose' of DNA is to survive, no more and no less. The simplest way to explain the surplus DNA is to suppose that it is a parasite, or at best a harmless but useless passenger, hitching a ride in the survival machines created by the other DNA." pg 156

Problem: traits of interest usually result In some fitness cost to carriers

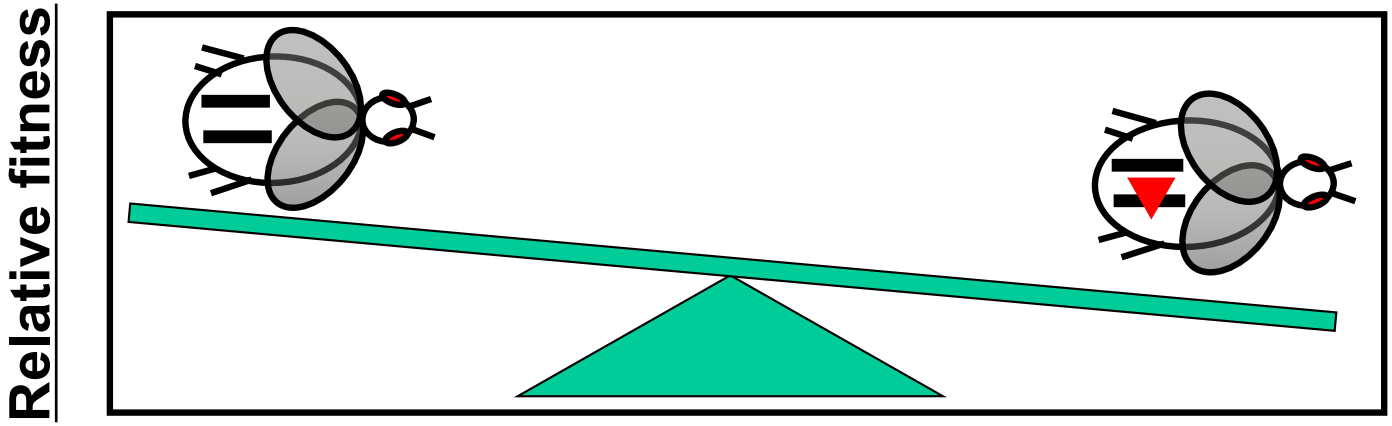


Problem: traits of interest usually result In some fitness cost to carriers



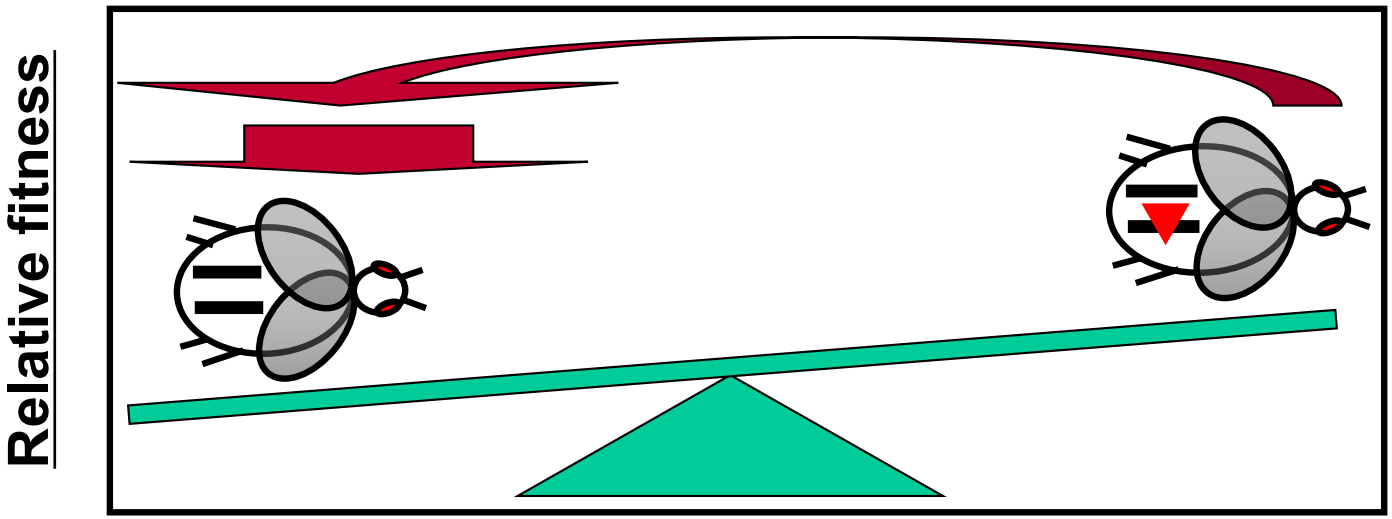


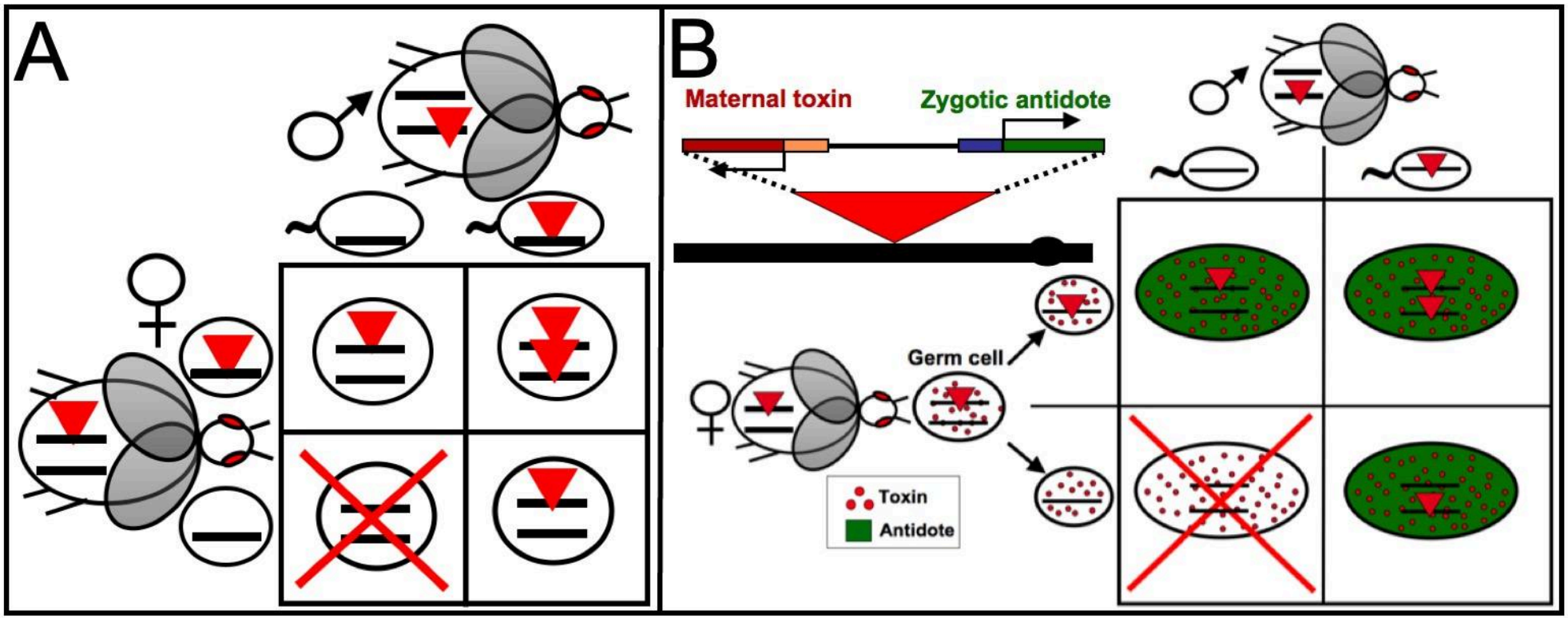
Problem: traits of interest usually result In some fitness cost to carriers



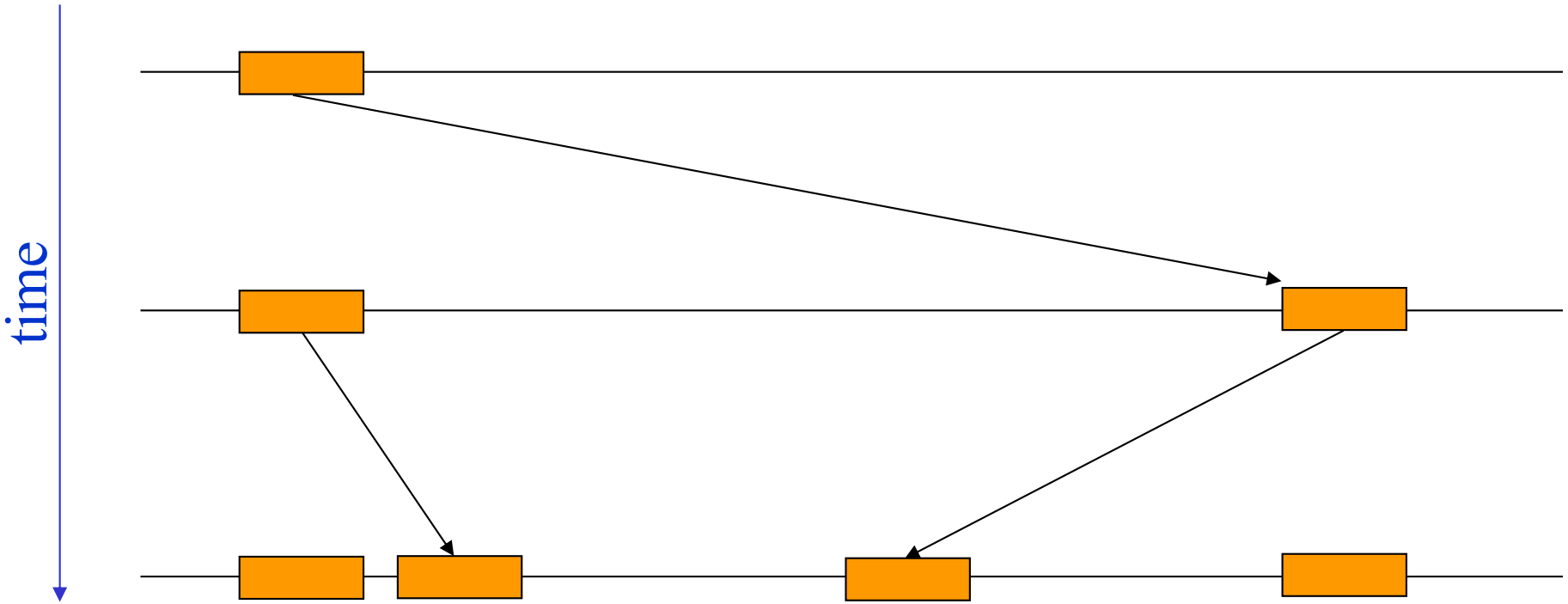
Solution: Increase the fitness cost associated with NOT carrying the gene of interest

Gene drive




A
B
Maternal toxin
Zygotic antidote
Germ cell
Toxin
Antidote

Imagine a gene that could duplicate copies of itself within the genome



A piece of selfish DNA has two distinct properties:

- Arises when DNA sequence spreads by forming additional copies of itself within the genome
- Makes no specific contribution to the phenotype

"Selfish sequences spread and make no contribution to the phenotype of the organism, except in insofar as it is a slight burden to the cell that contains it. The spread of DNA sequences within the genome can be compared to the spread of a not-too-harmful parasite within its host."

Intra-genomic selection

"The idea of selfish DNA is firmly based on this general theory of natural selection, but it deals with selection in an unfamiliar context"

"In short, we may expect a kind of molecular struggle for existence within the DNA of the chromosomes, using the process of natural selection.

Orgel, L. E. & Crick, F. H. C. 1980. Nature 284: 604-607

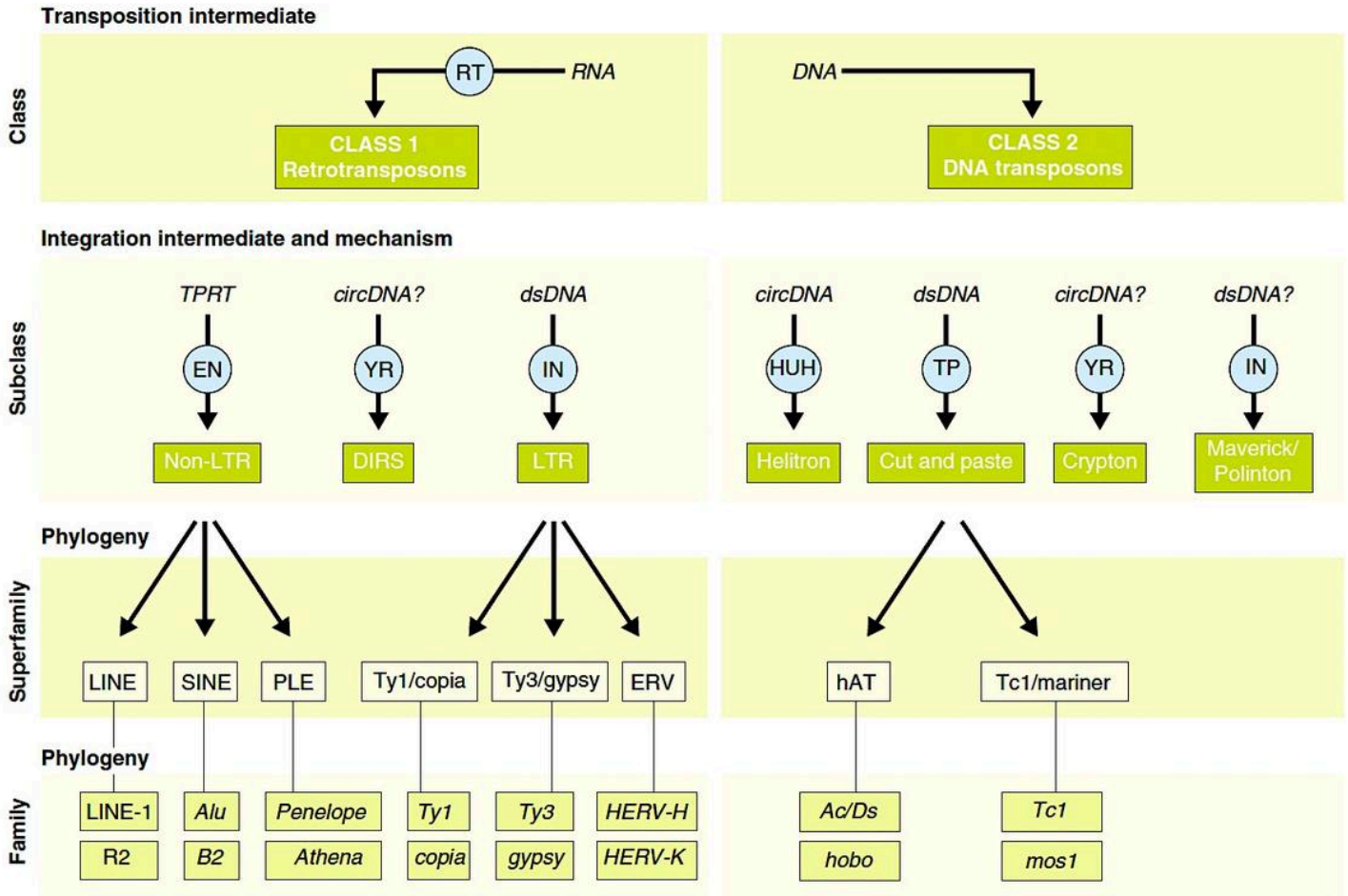


Fig. 1 Classification of eukaryotic transposable elements. Schematic and examples showing the key features and relationships between TE classes, subclasses, superfamilies, and families. Blue circles represent TE-encoded enzymes. circDNA circular DNA intermediate, DIRS Dictyostelium repetitive sequence, dsDNA linear double-stranded DNA intermediate, EN endonuclease, IN integrase, PLEs Penelope-like elements, HUH, Rep/Helicase protein with HUH endonuclease activity, RT reverse transcriptase, TP transposase, TPRT target primed reverse transcription, YR tyrosine recombinase (for other abbreviations, see text)

Structure	Key Genes	Mode of Movement
 Short inverted repeats at each end	Transposase	Moves as DNA by excision or replication
 Long direct repeats at both ends	Reverse Transcriptase Integrase	Moves by a RNA intermediate
 Poly-A at 3' end 5' end often truncated	Reverse Transcriptase/ Endonuclease	Moves by a RNA intermediate

Bacterial transposons Insertion (IS) sequences

What defines most transposons

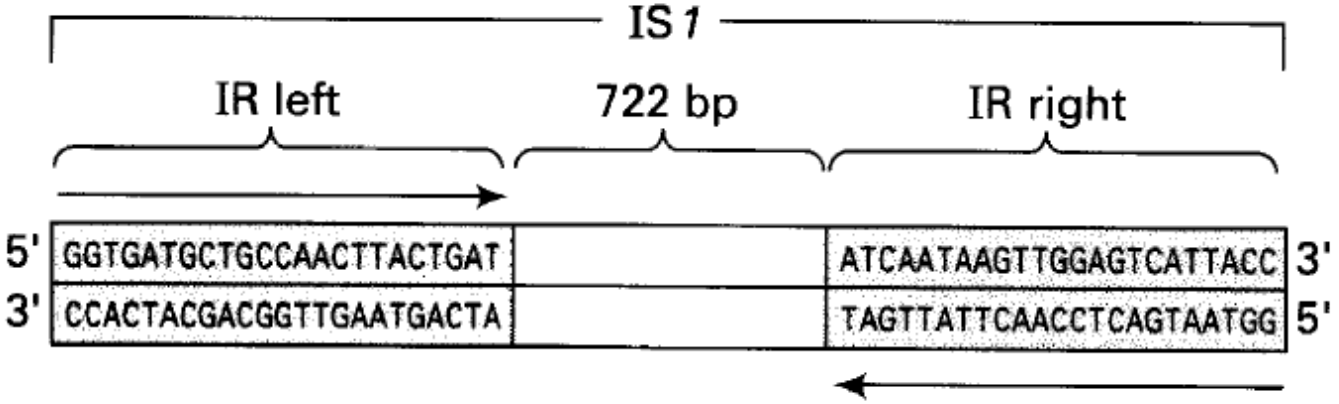
1. **Inverted repeats**
2. **Sequences between the inverted repeats that may be transported**
3. **A source of transposase, and in the case of replicative transposition, a source of resolvase. Transposase provides a mechanism to move these elements from one site to another.**

The basic prokaryotic unit of transposon unit is an IS and a source of transposase.

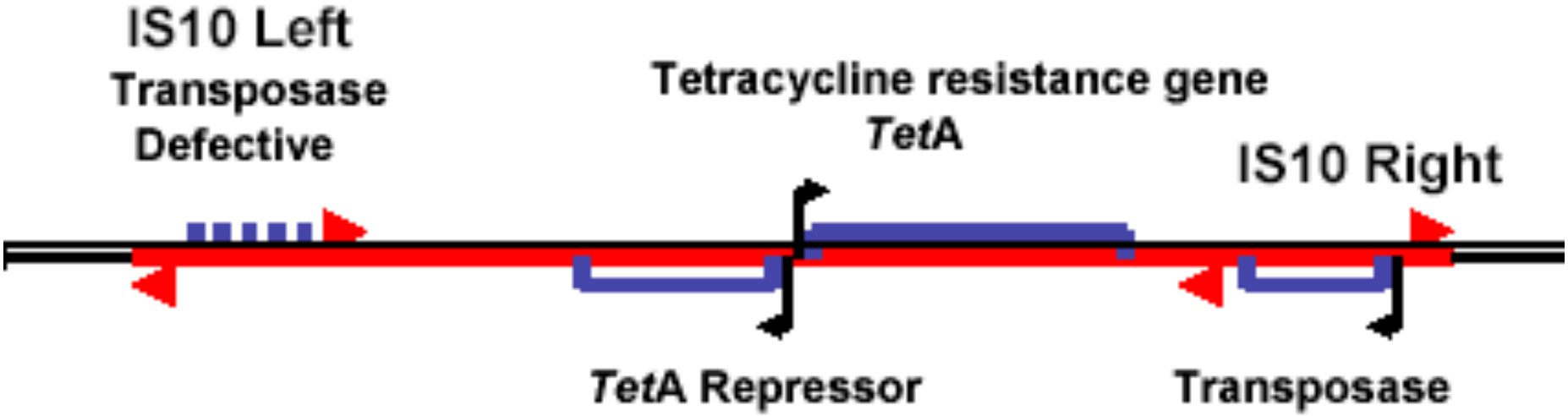
Transposase may be provided either in cis (between the inverted repeats) or in trans (somewhere else on the bacterial chromosome).

Insertion sequence

inverted repeats



Structure of a Complex Transposon Tn10



IS10 Right may transpose by itself
or Tn10 can be moved as a unit

A composite transposon has a central region carrying markers (genes) unconnected with transposition, flanked by IS modules.

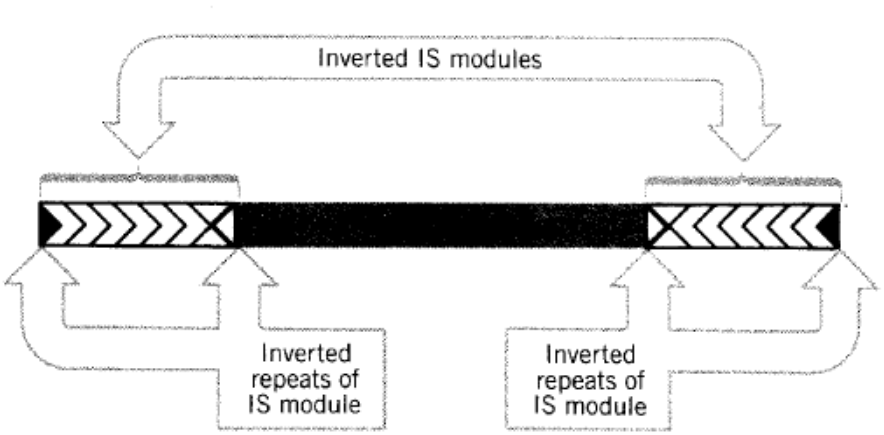
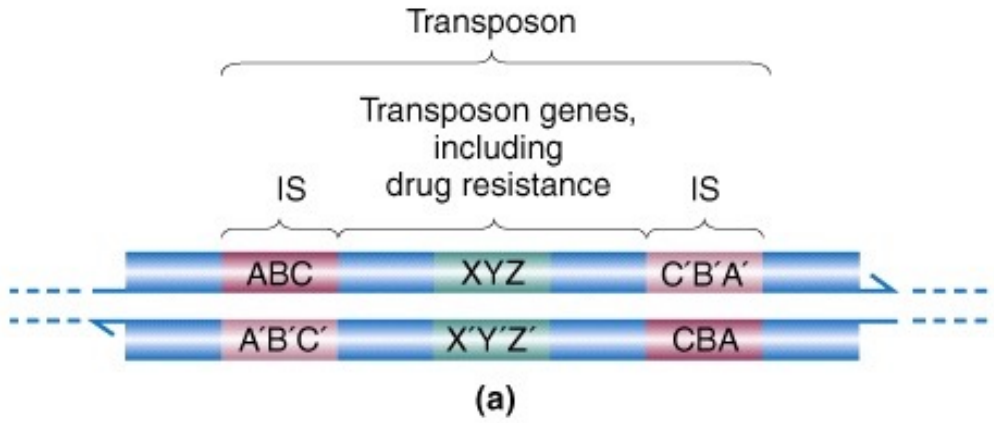
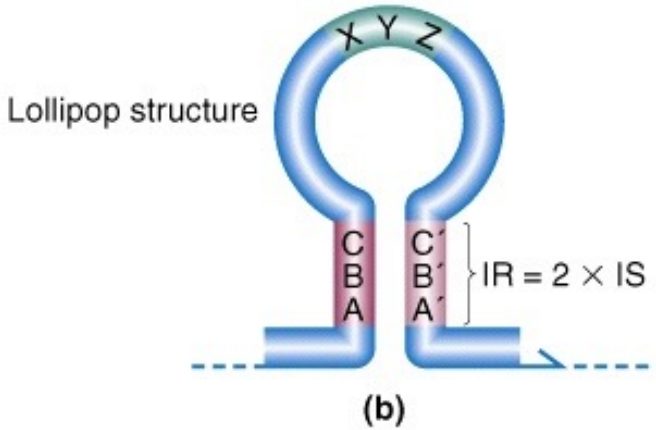
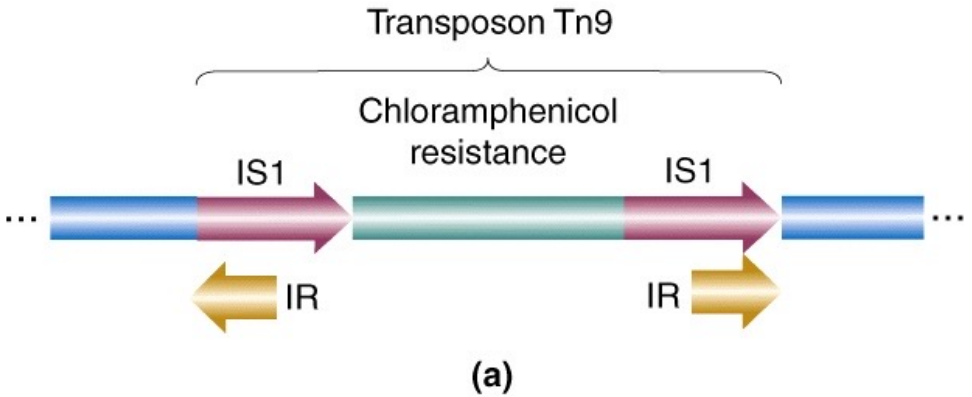
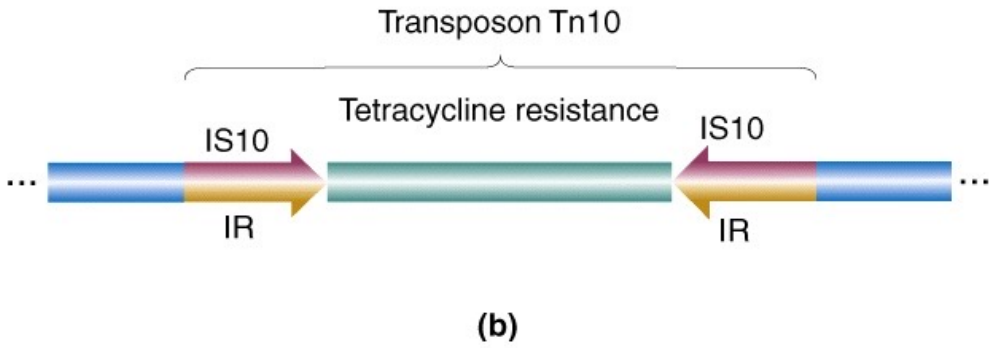


Figure 35.3
A composite transposon has a central region carrying markers unconnected with transposition (such as drug resistance) flanked by IS modules. The modules have short inverted terminal repeats. If the modules themselves are in inverted orientation (as drawn), the short inverted terminal repeats at the ends of the transposon are identical.

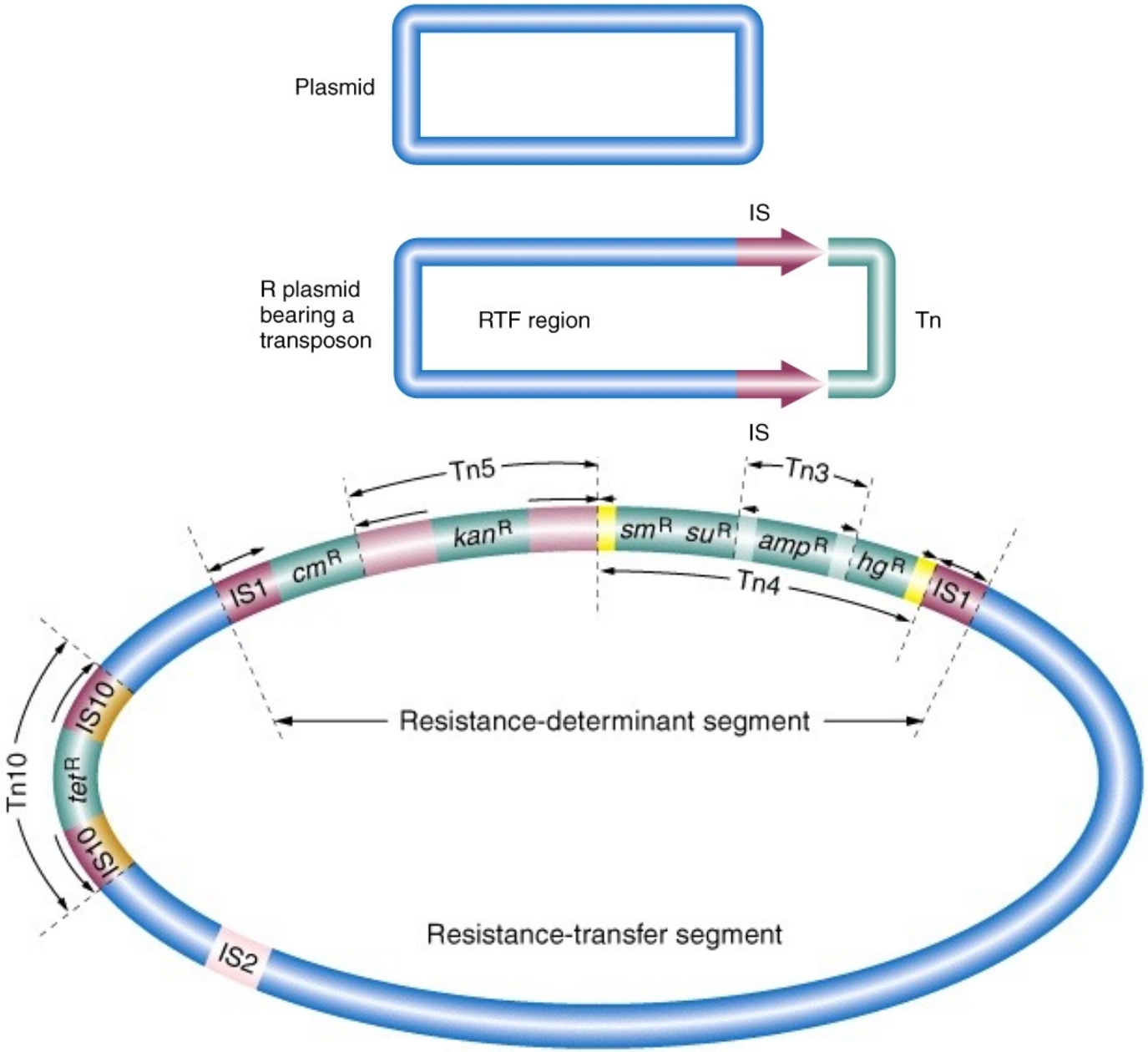




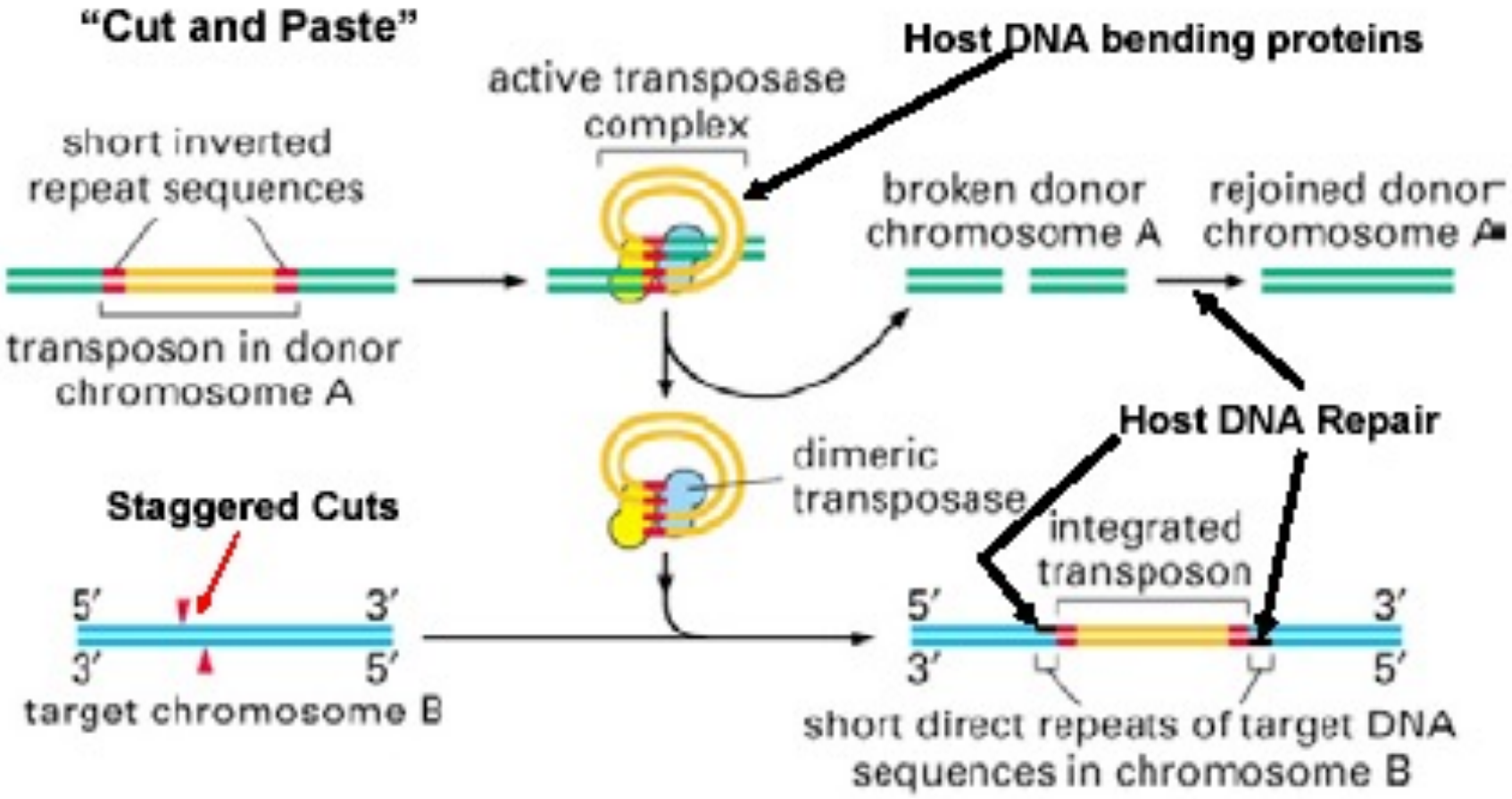




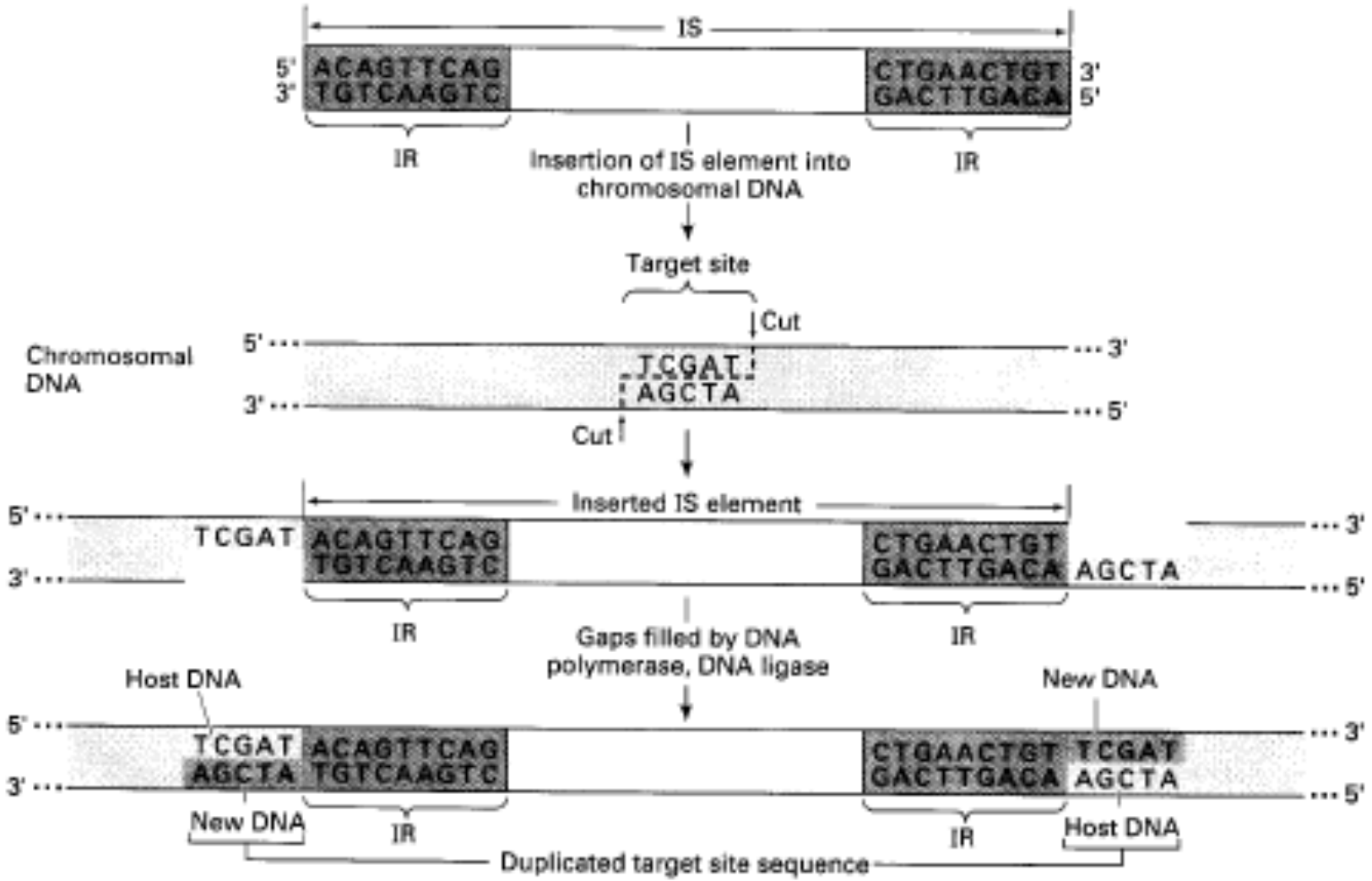
Plasmids are autonomously replicating elements located in the bacterial cytoplasm. They can move from cell to cell. Insertion into a plasmid of a transposon carrying an antibiotic resistance gene provides a mechanism by which antibiotic resistance can be transferred from cell to cell.



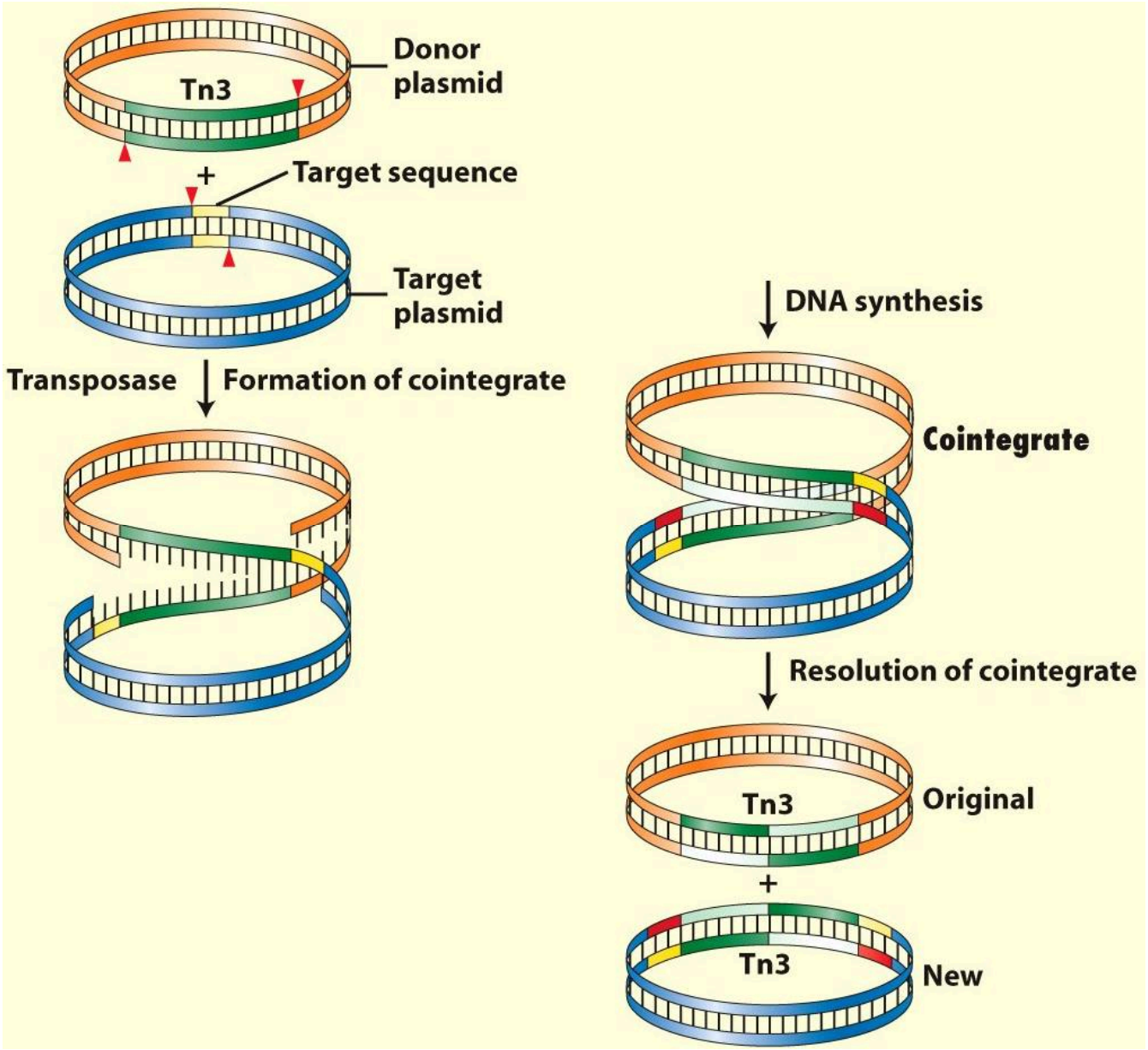
Mechanism of Simple Transposition by IS10



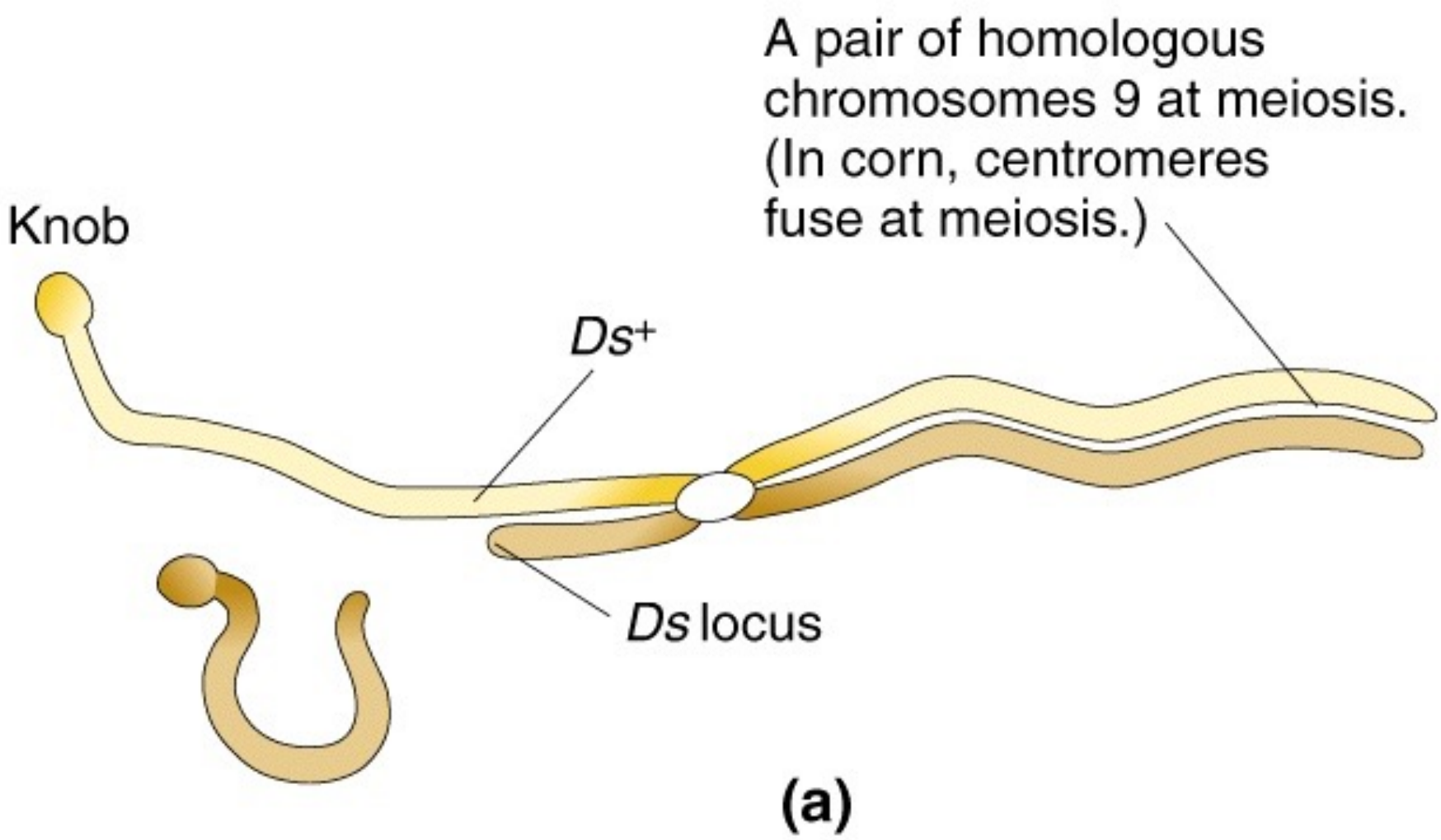
Schematic of the integration of an IS element into chromosomal DNA. As a result of the integration event, the target site becomes duplicated to produce direct target repeats. Thus, the integrated IS element is characterized by its inverted repeat (IR) sequences flanked by direct target site duplications. Integration involves making staggered cuts in the host target site. After insertion of the IS, the gaps that result are filled in with DNA polymerase and DNA ligase. (Note: The base sequences given for the IR are for illustration only and are not the actual sequences found, either in length or sequences.)



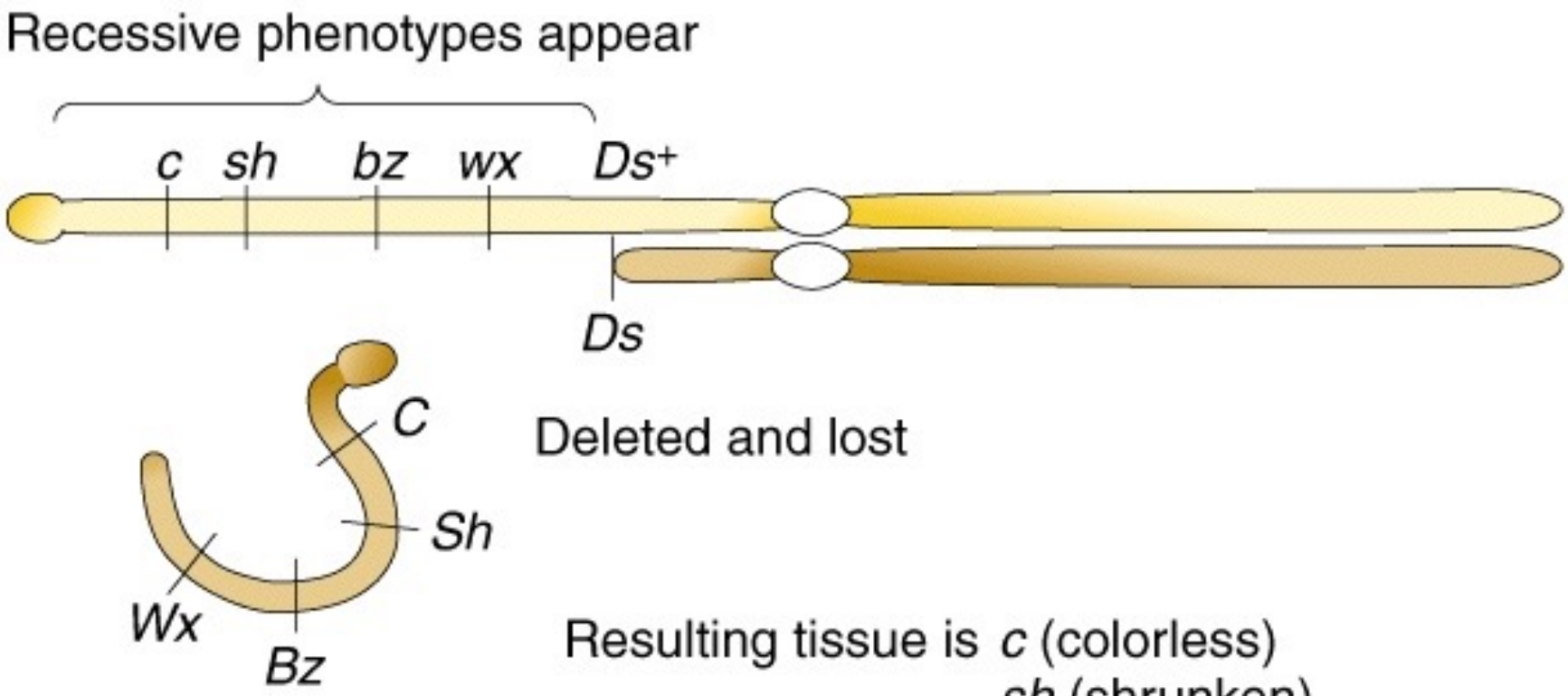
Replicative transposition involves the formation of a cointegrate. The initial event involves cleavage of donor and recipient strands at only two sites, followed by ligation and DNA synthesis.homologous recombination followed by resolution of the Holliday junction results in the formation of two molecules, both containing transposons. The new chromosomal insertion is flanked by direct repeats of the target sequence.



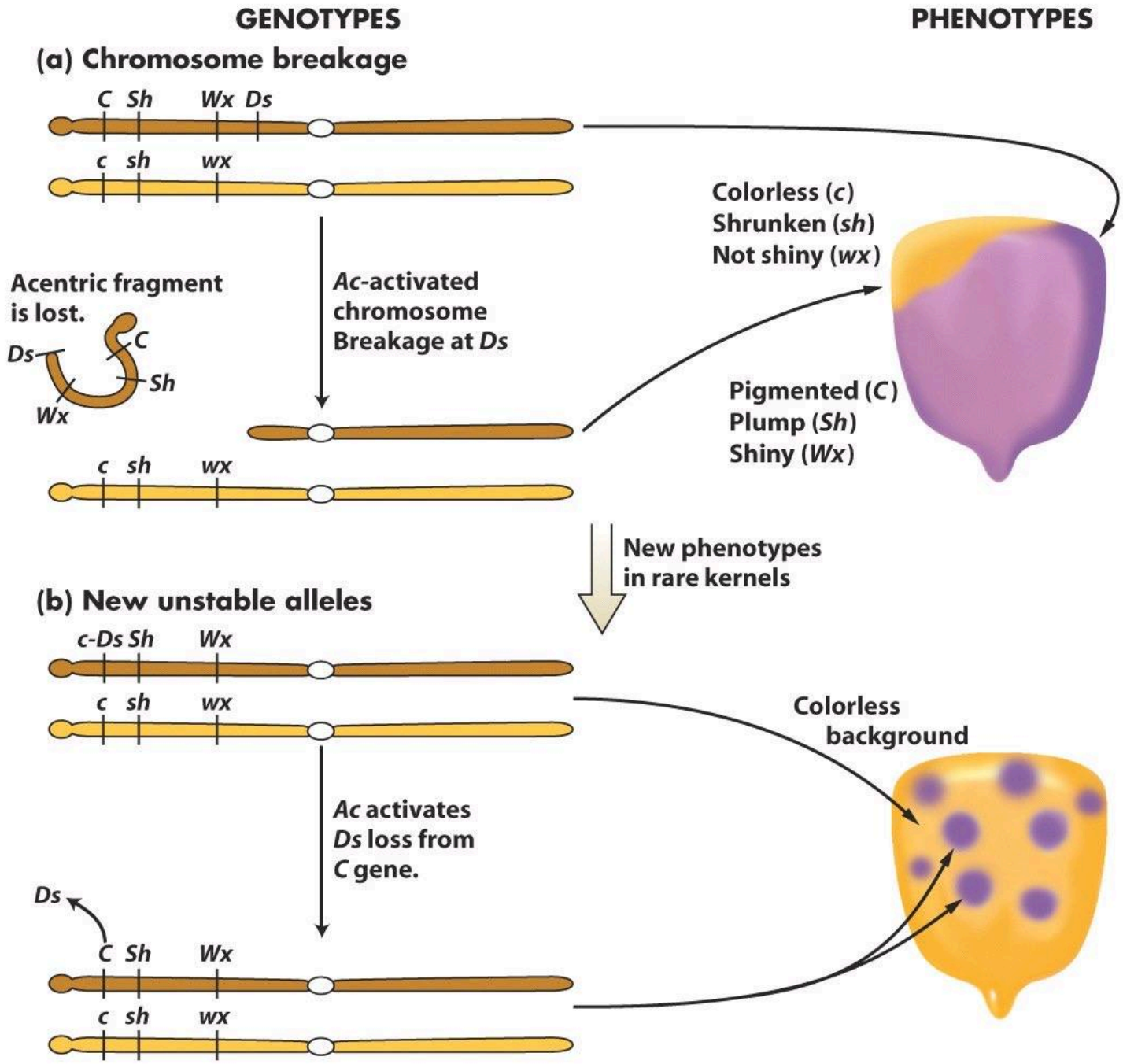
Autonomous and non-autonomous transposons in corn

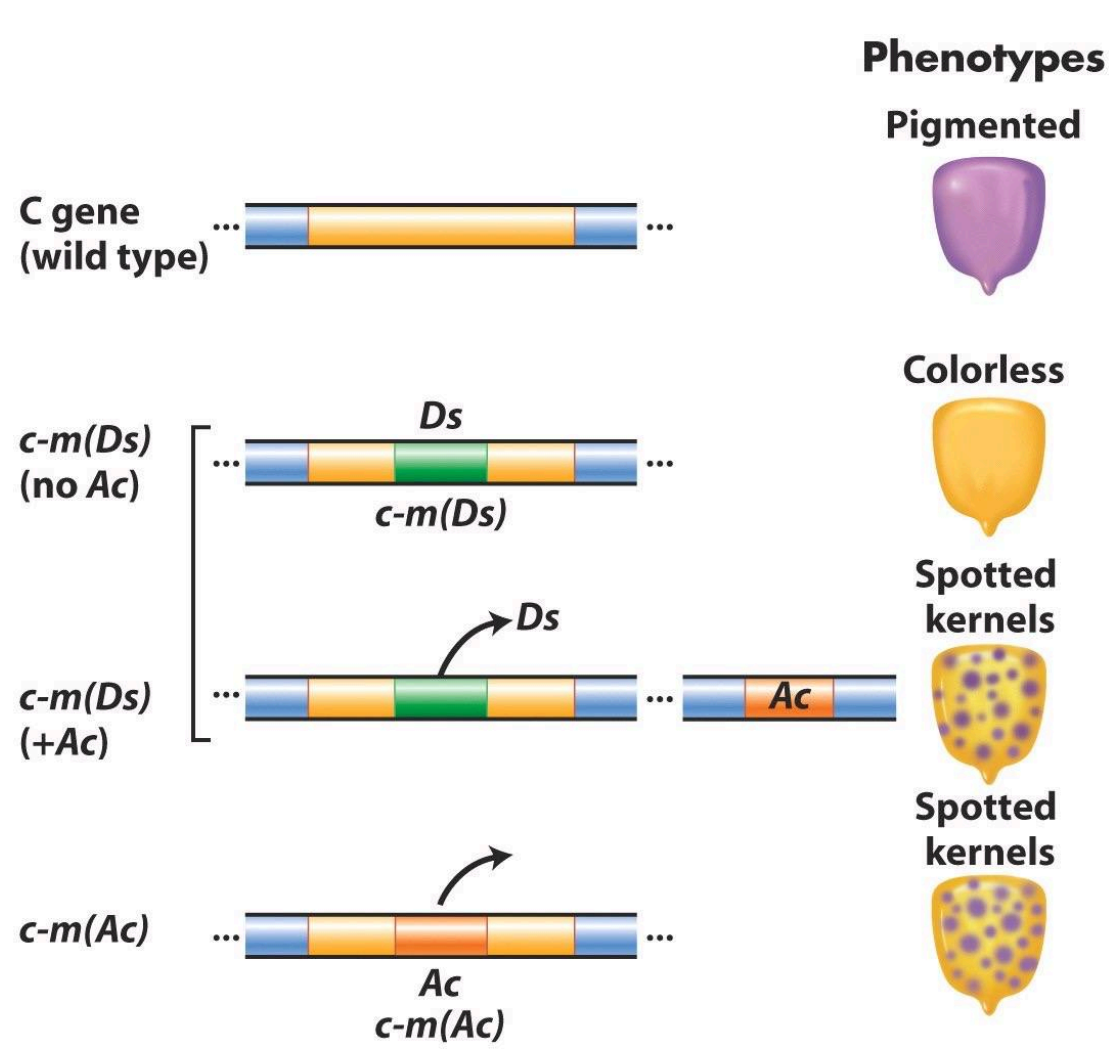


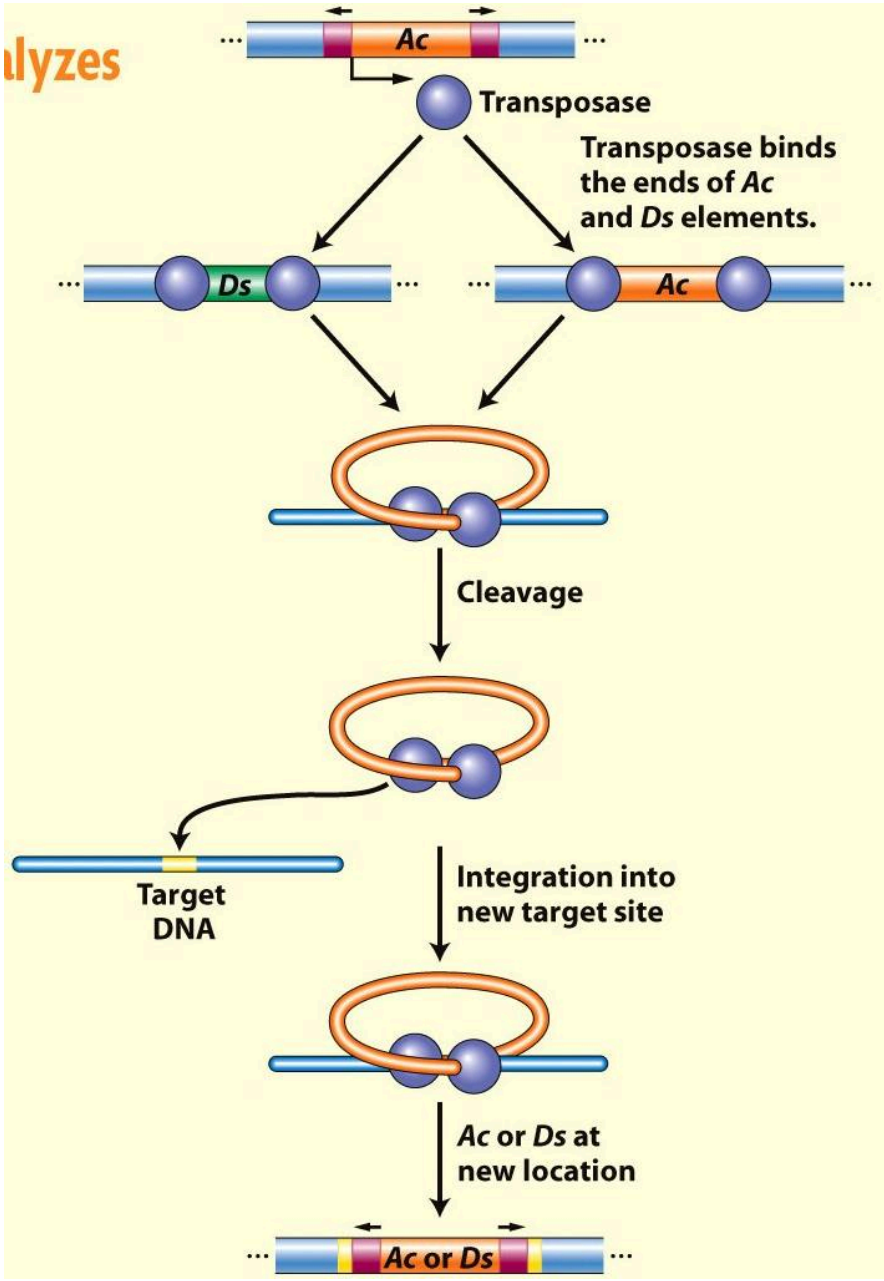
A large chromosome fragment is sometimes lost in the presence of the Ds allele.



(b)


GENOTYPES
PHENOTYPES
(a) Chromosome breakage
C Sh Wx Ds
c sh wx
Colorless (c)
Shrunken (sh)
Not shiny (wx)
Acentric fragment
is lost.
Ds
C
Sh
Wx
Ac-activated
chromosome
Breakage at Ds
Pigmented (C)
Plump (Sh)
Shiny (Wx)
c sh wx
New phenotypes
in rare kernels
(b) New unstable alleles
c-Ds Sh Wx
c sh wx
Colorless
background
Ac activates
Ds loss from
C gene.
Ds
C Sh Wx
c sh wx





Transposon art. Mobile genetic elements caused these prized color patterns in morning glories.

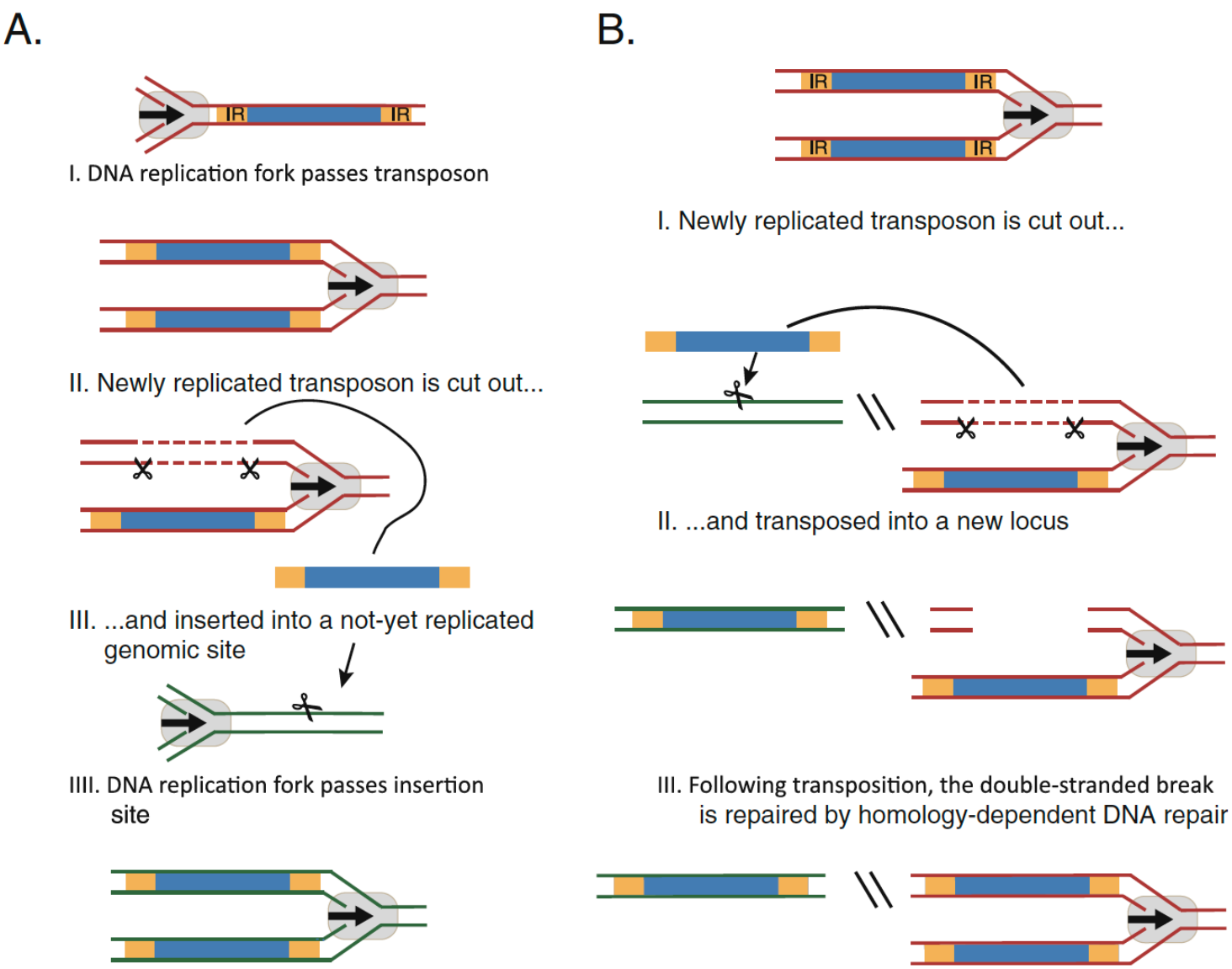


Figure 3 Models of replicative transposition. (A) After replication, the transposon is excised and integrated into a yet unreplicated genomic site thus duplicating the newly inserted transposon. **(B)** The double-stranded break created by transposition at newly replicated DNA is repaired using the sister chromatid as a template for homology-directed DNA repair, leading to reconstitution of the excised transposon. IR, terminal inverted repeat.

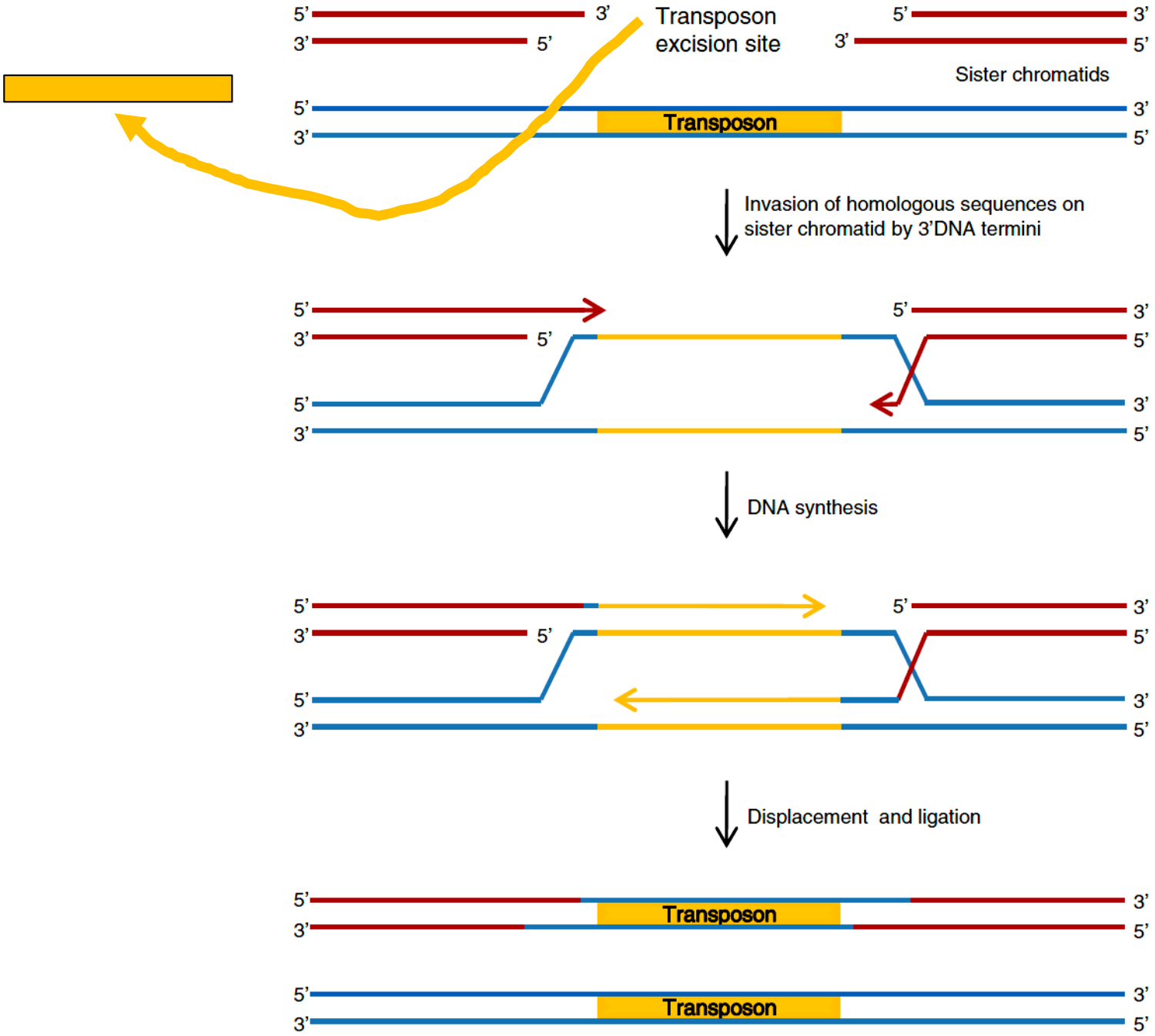
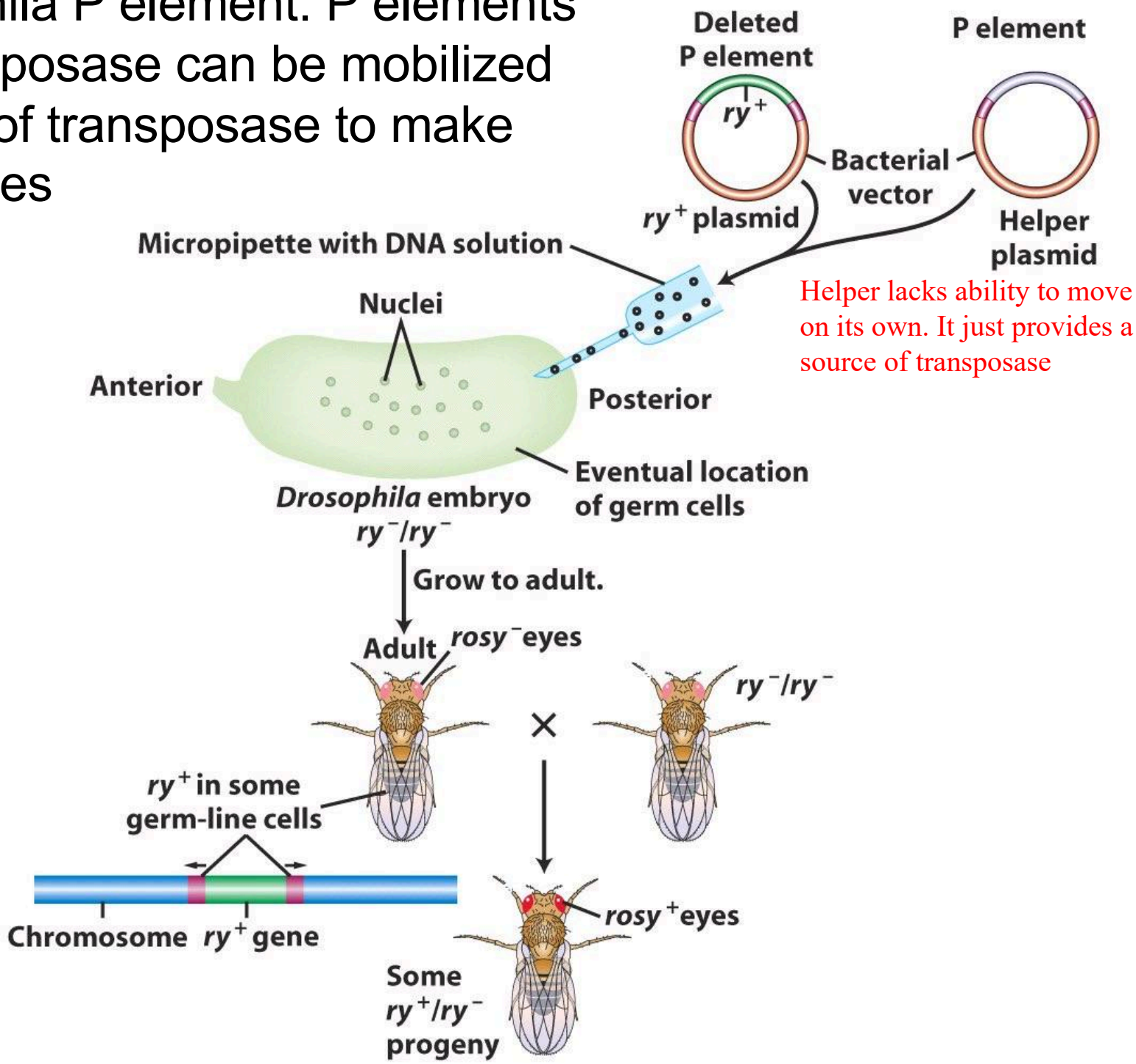


Figure 4 Homology-dependent DNA repair by synthesis-dependent strand annealing (SDSA). Following excision, the 3′-DNA termini from the double-stranded breaks invade the homologous sequence on the sister chromatid. The homologous sequence is then used as template for DNA synthesis and the final, elongated 3′-DNA termini anneal to each other and are joined with the 5′ ends by ligation.

The Drosophila P element. P elements lacking transposase can be mobilized by a source of transposase to make transgenic flies

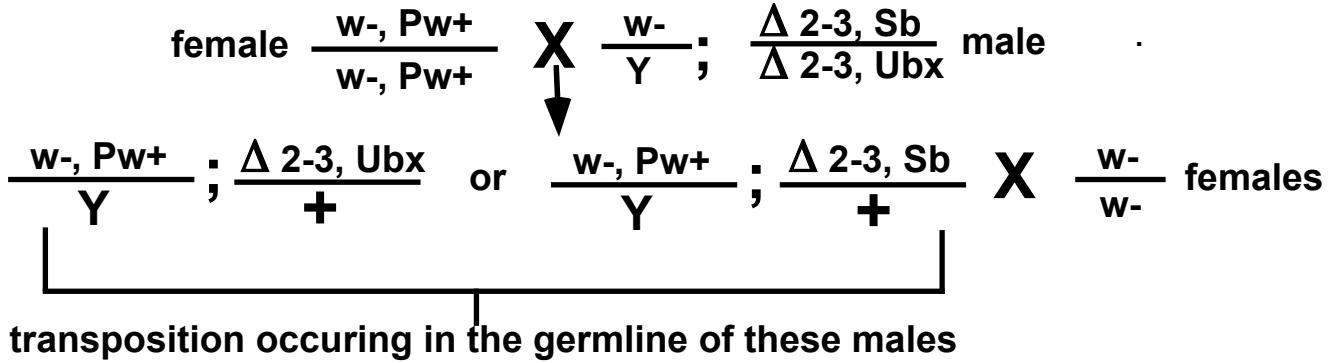


Helper lacks ability to move on its own. It just provides a source of transposase

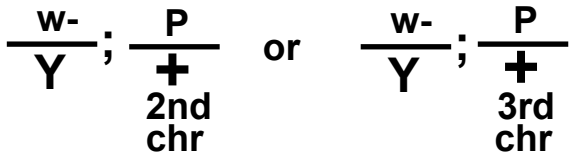
Hopping a transposon in flies off the X chromosome, and onto an autosome.

Pw+ = a transposon that carries a wildtype copy of the w gene on the X chromosome.

Δ 2-3, Ubx and Δ 2-3, Sb = these are third chromosomes that carry an immobilized source of transposase (Δ 2-3), as well as a dominant visible marker (Ubx or Sb). Transposase is normally only active in the germline. However, Δ 2-3 is active in somatic cells as well. This has the consequence that transposons are continually being mobilized in somatic tissues. If the P element contains a cell autonomous marker gene such as w+ (red pigment) this mobilization can be observed as eye color mosiacism.



The P element can hop into the + version of the second or third chromosome. If this occurs, and the meiotic product (the sperm) also carries a Y chromosome, then the progeny will be a red-eyed male. If no transposition occurs or the P hops into outer space then male progeny will have white eyes.



There are a number of other male progeny types that will also be genrated. Some of these will still have the Δ 2-3 source of transposase. Others will simply have no P element transposition onto one of the chromosomes present in the gamete. These will all lead to progeny that are w-

1. Any red eyed male resulting from the above crossmust have had a transposition event off the X chromosome and onto an autosome or onto the Y chromosome (happens rarely), since males get their X (which in this case is w-) from the female.

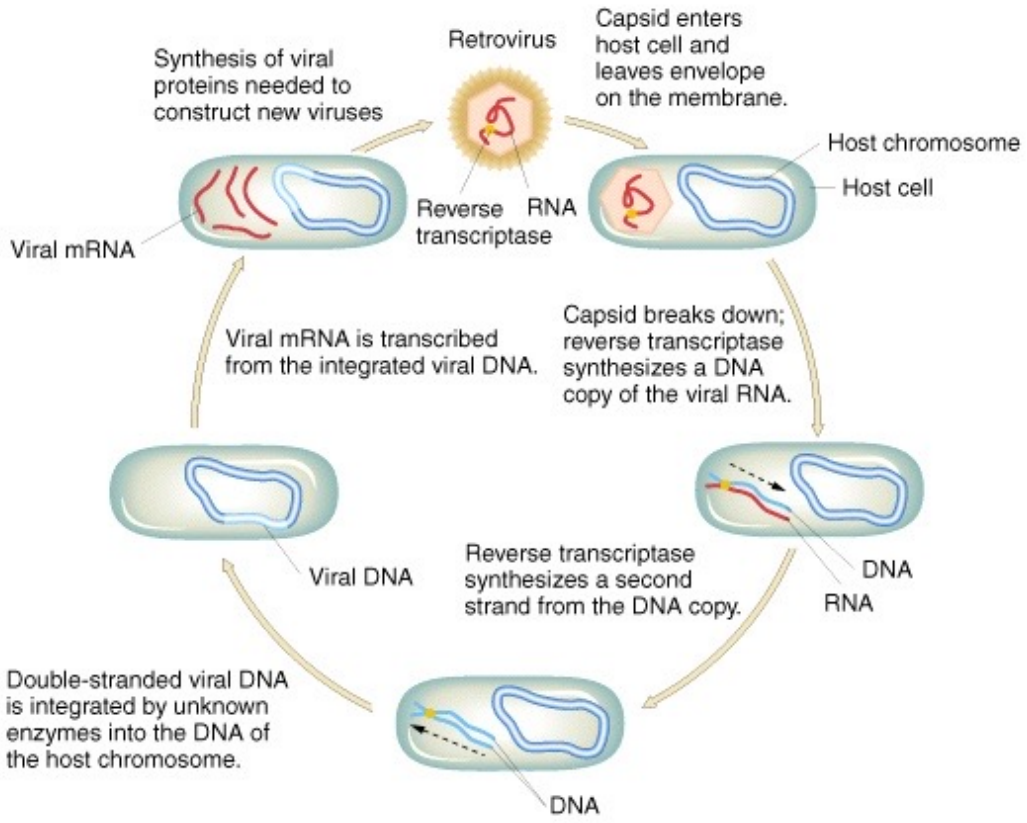
2. Males that have both a P element and Δ 2-3 will have red and white variegated eyes as a result of somatic transposition of the Pw+ in somatic eye tissue.

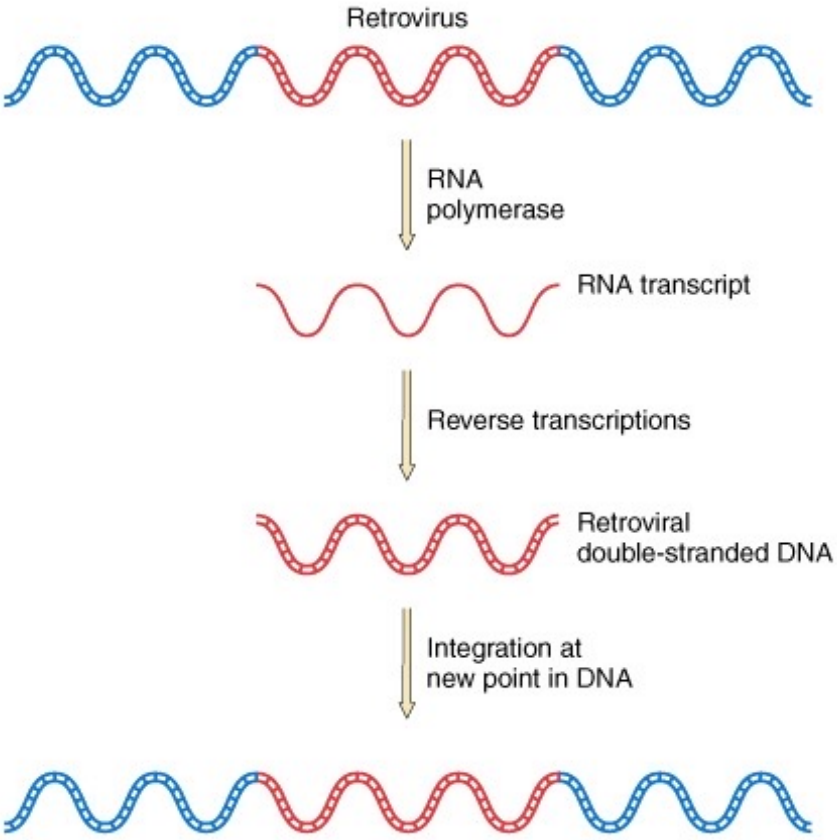
The power of the P element

In the early 20th century wild Drosophila melanogaster populations lacked the P element. We know this because people collected animals from throughout the world and kept them as stocks up through the present. They lack any P elements.

The P element entered D. melanogaster sometime in the 20th century, from another species. It is now ubiquitous (30+ copies) in ALL wild populations.

Retroviruses are RNA viruses that generate a double-stranded DNA intermediate using the enzyme reverse transcriptase. The double-stranded DNA intermediate can integrate into the genome, generating a provirus.





Retrovirus-like Retrotransposon

Yeast Ty1

LTR: Long Terminal Repeat contains both promoter and poly-A sites

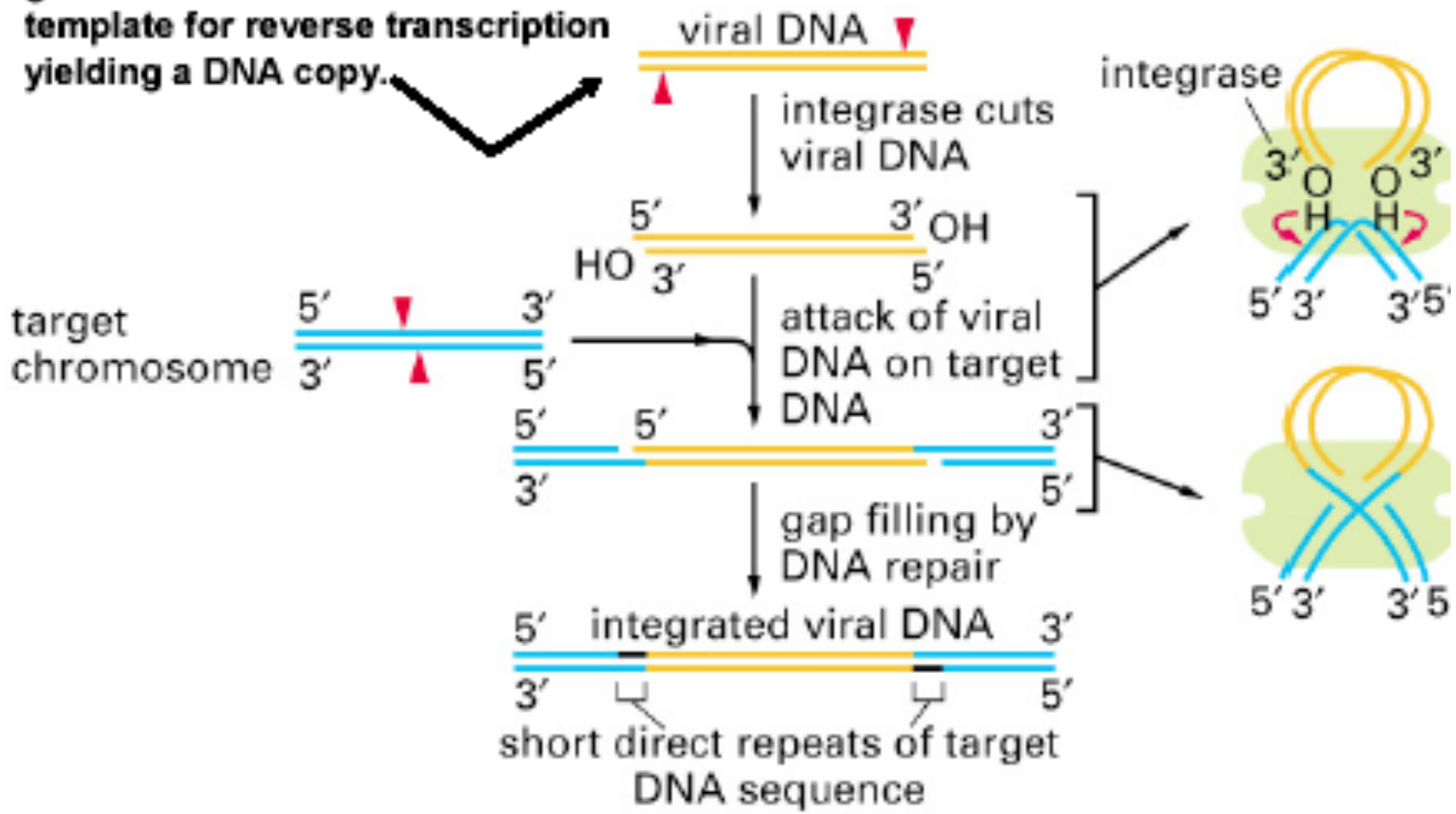
TYA: encoding a polyprotein for the capsid (*gag*-like proteins)

TYB: encoding a polyprotein for enzymes used in transposition

- ***pro*** **proteinase**
- ***int*** **integrase**
- ***rt*** **reverse transcriptase**
- ***rnh*** **RNaseH**

Integrase Catalyzes Insertion of Retrotransposons

Retrotranposon transcription generates RNA that serves as a template for reverse transcription yielding a DNA copy.



Mechanisms by which transposons and retrotransposons increase in copy number

Replicative transposition

Repair of a DNA Double-stranded break

RNA synthesis followed by reverse transcription

Movement in front of a replication fork

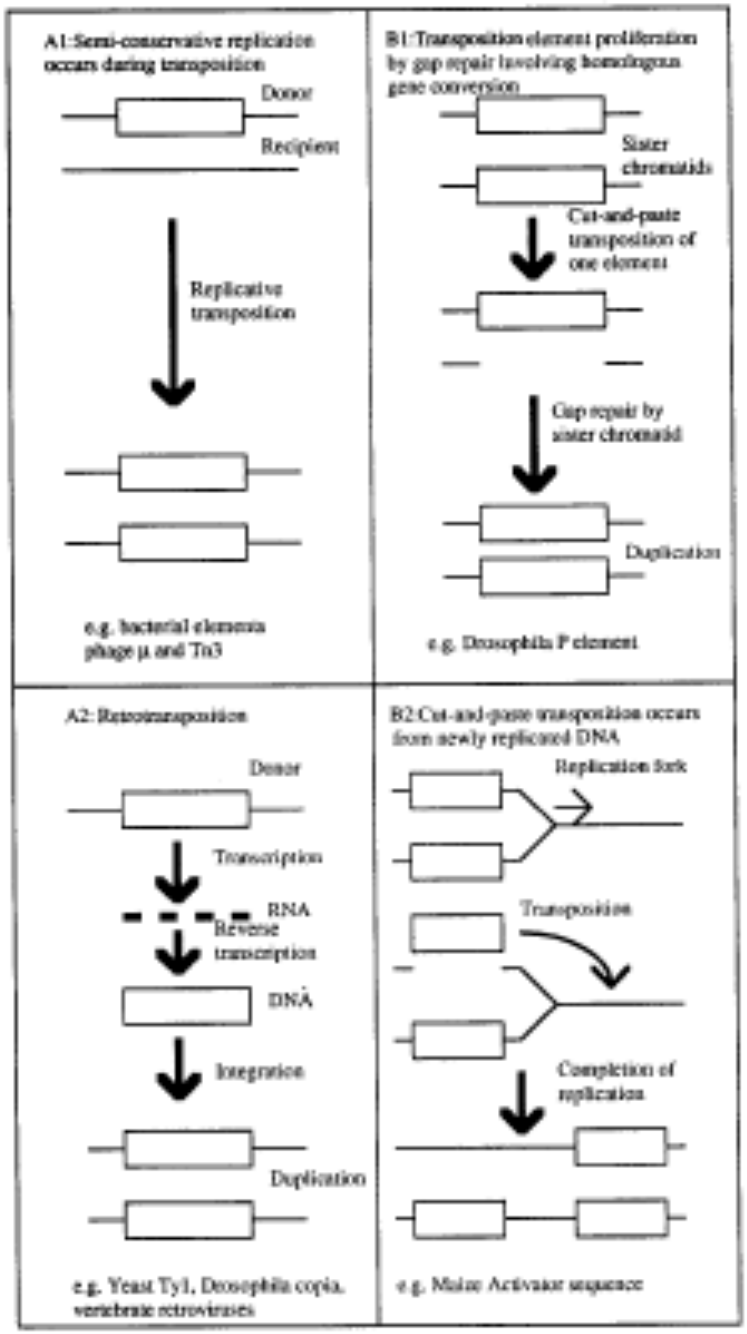
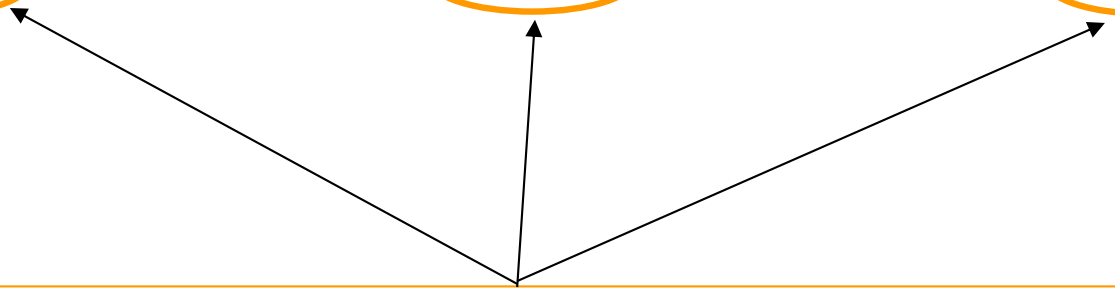


Fig. 1 Four ways in which transposition can lead to an increase in transposable element copy number. For

Number and nature of interspersed repeats in human, fly and worm

	Human		Fly		Worm	
	Percentage of bases	Approximate number of families	Percentage of bases	Approximate number of families	Percentage of bases	Approximate number of families
LINE/SINE	33.40%	6	0.70%	20	0.40%	10
LTR	8.10%	100	1.50%	50	0.00%	4
DNA	2.80%	60	0.70%	20	5.30%	80
Total	44.40%	170	3.10%	90	6.50%	90



The euchromatic portion of the human genome has a much higher density of transposable element copies than fly and worm

Transposons and genome structure

Drosophila melanogaster: 90 different elements = 3% of genome
Mammals: LINES and SINES = 34% of genome

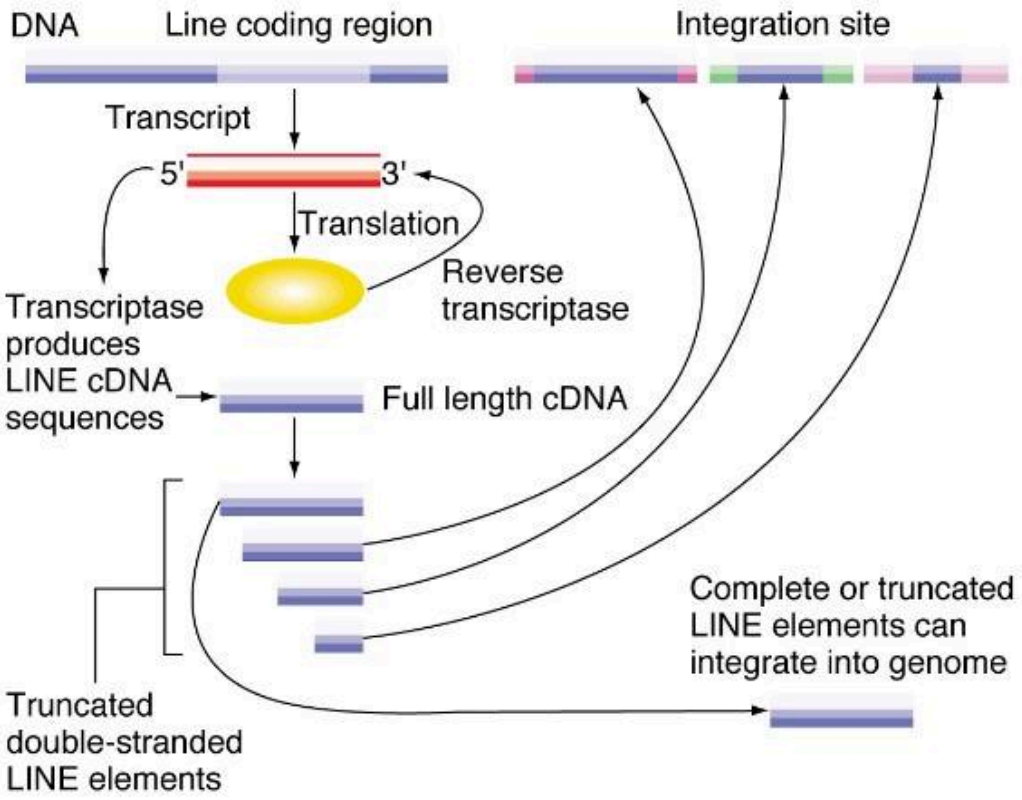
LINES are retroposons that derive from a "selfish" DNA sequence that encodes reverse transcriptase. A complete LINE sequence can be copied into RNA. Reverse transcriptase makes cDNA copies from the RNA. These copies may be complete or truncated. Both may then integrate into other sites in the Genome.

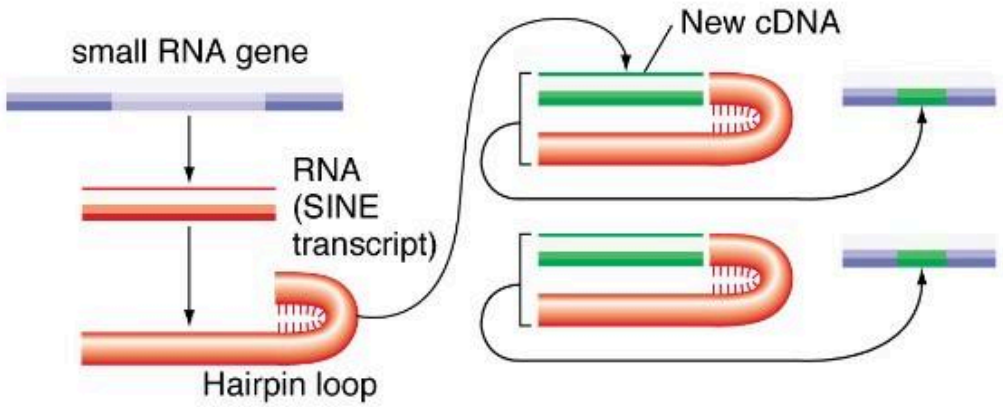
SINES replicate through an RNA intermediate, but do not encode reverse transcriptase themselves. Instead They utilize reverse transcriptase present in the cell.

The Alu family of repetitive elements in the human Genome is an example of a SINE family. There are Than 500,000 Alu elements/ genome at 300bp/element.

The SINE element has evolved from normal cellular RNAs, most often tRNAs, but also an RNA that forms A component of the signal recognition particle, which Is required for protein translocation across the Endoplasmic reticulum membrane.

The important modification of these cellular RNAs that Has occurred are modifications in the 3' end of the Transcript that lead to self complementarity. This Provides a binding structure for reverse transcriptase.





How a Quarter of Cow DNA Came From Reptiles

By hopping between species, jumping genes have radically altered the course of animal evolution.

By Ed Yong

Transposon mobilization: Consequences for genome evolution over time

1. Inversions as a result of recombination between inverted repeats.

2. Deletions as a result of recombination between direct repeats.

3. Unequal crossing over between repeats (regions of micro-homology) leads to translocations, deletions, and gene duplications.

4. Gene inactivation as a result of transposon insertion.

5. Gene activation as a result of a transposon (or retrovirus) inserting near a gene, and driving the transcription of that gene from a transposon promoter.

6. Once transposons enter the germline of a species they can multiply and become ubiquituous very rapidly.

Drosophila P elements. World-wide spread within 30 years

Alu sequences in primates and humans.

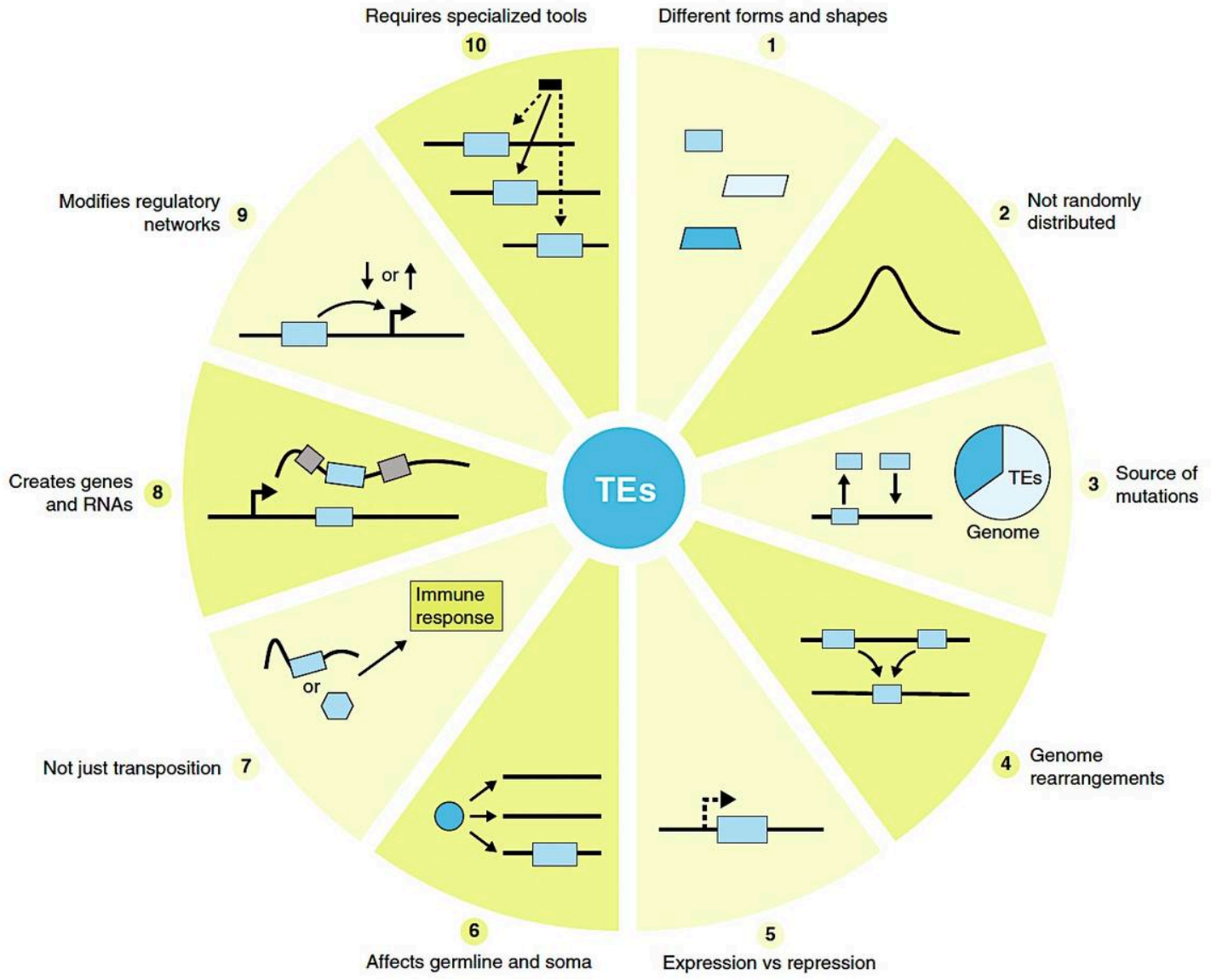


Fig. 2 Ten things you should know about transposable elements (TEs). Examples of how TEs can impact genomes in direct and indirect ways. Blue boxes represent TEs, gray boxes represent canonical exons, and the black box represents a sequencing read. Right-angled arrows represent gene or TE promoters