Genome Evolution

- Finishing up chromosome rearrangements
- Polyploidy
- Genome size (picograms DNA)
- Genome size (genes)
- Repeated elements

Centromere identity

- Chromosome pairing reflects:
  - Identity of centromere
  - Identity of sequences in arms

- In chromosome segregation, the centromere tends to win out

- An arm of 8 stuck to the centromere of 10 will segregate with 10, not 8

Practice problem

- A healthy woman comes to the clinic for diagnosis
- Her karyotype shows that she has the following chromosomes (plus normal copies of all other chromosomes):
  - Almost all of 14, including centromere, attached to long arm of 21
  - Normal 21
  - Normal 14
- While the short arm of 21 is missing from the translocated chromosome, apparently this has no phenotype (it's mostly heterochromatin)
- Her spouse is chromosomally normal

Practice problem

What proportion of her children will:

- Have Down Syndrome due to 3 copies of 21
- Miscarry due to 3 copies of 14 or 1 copy of any chromosome (assume short arm of 21 is not needed)
- Be healthy carriers like their mother
- Be normal

Translocation

Polyploidy

- Having more than the normal number of chromosome sets
  - Many land plants: almost all cultivated plants (4N, 6N, 8N)
  - Salmon and trout (4N, long ago)
  - African Clawed frogs (Xenopus) 4N, 6N, 8N
  - Triploid toads (!?)

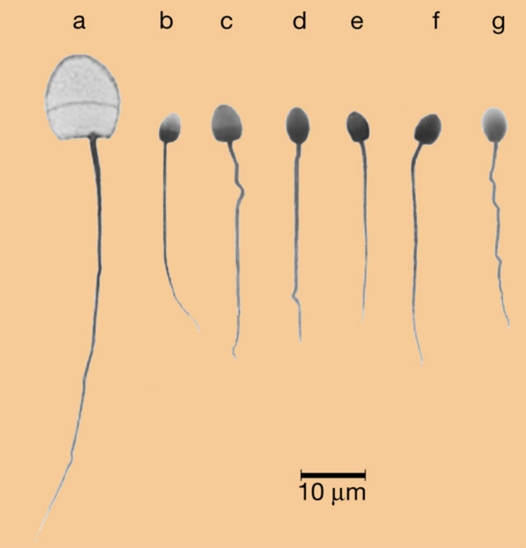
- Not well tolerated in mammals

- 3N and 4N human fetuses seldom survive to birth

Possible tetraploidy in a mammal?

- Red viscacha rat has 2N=112
- Only 1 X and one Y
- Chromosome painting suggests it's not tetraploid:
  - Does not seem to have 4 copies of any chromosome
  - Extra DNA mainly repeated sequences
- Alternative: runaway repeated DNA or transposons

Image by Michael A. Mares



Gallardo et al, Nature 1999

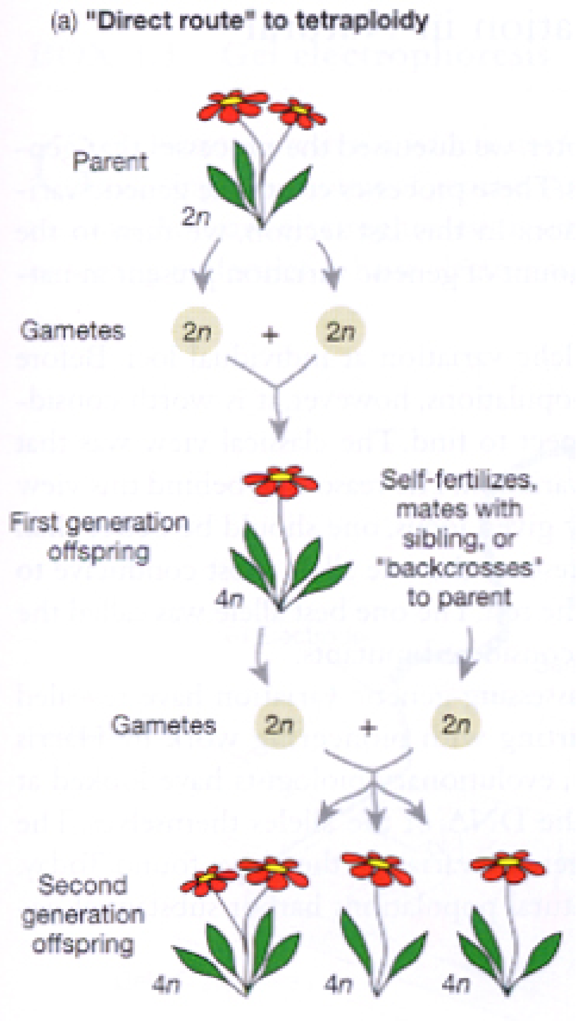
Discussion question

What are the reproductive options for a brand new tetraploid (4N)?

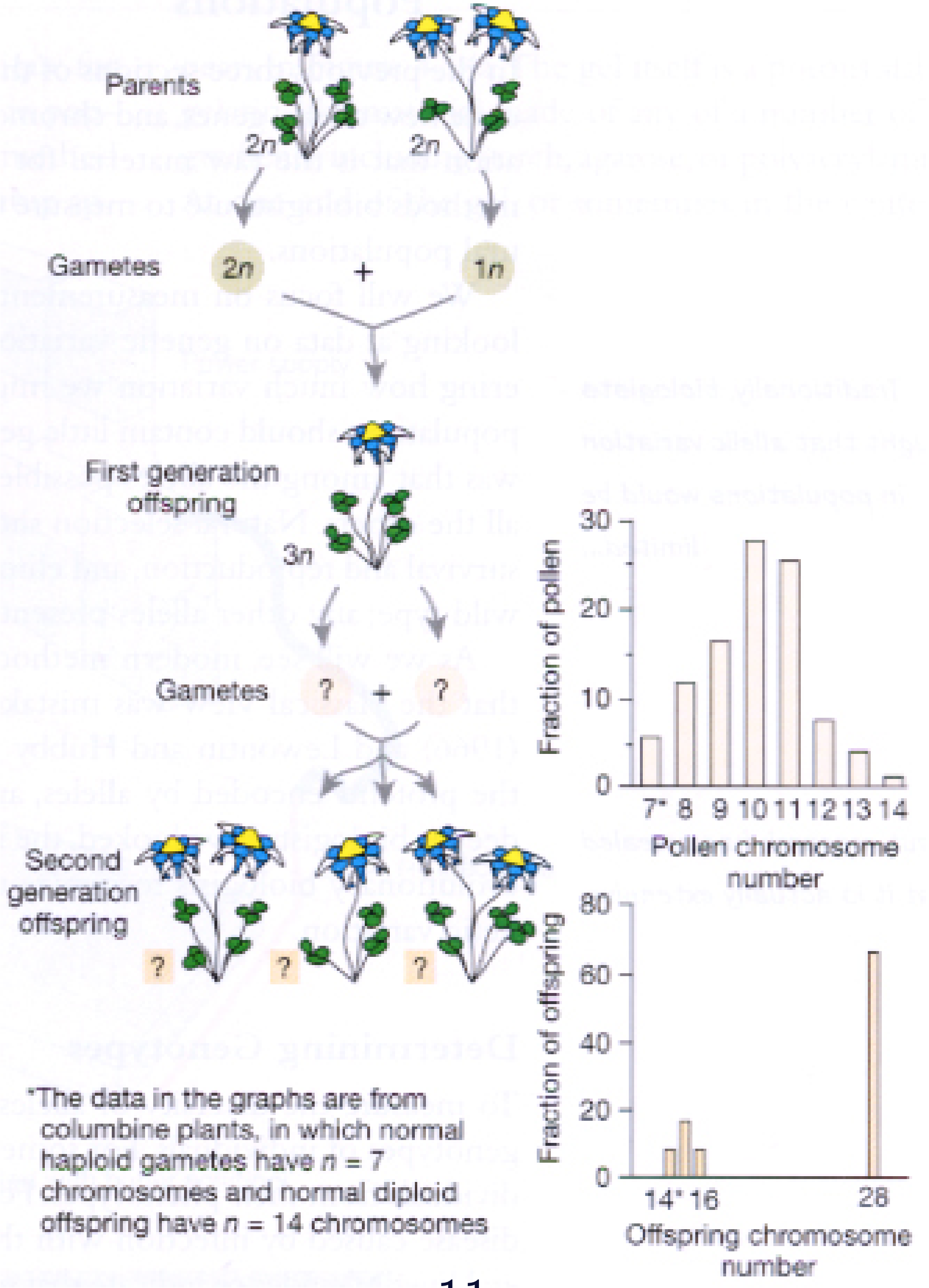
Discussion question

- What are the reproductive options for a brand new tetraploid (4N)?
  - Mate with another tetraploid, if it can find one
  - Self-fertilize
  - Clone itself
  - Reproduce vegetatively
  - Mate with a diploid (but offspring will be triploid, which is iffy)
  - Do something very weird in meiosis

Polyploidy


(a) "Direct route" to tetraploidy
Parent
2n
Gametes
2n
+
2n
First generation offspring
4n
Self-fertilizes, mates with sibling, or "backcrosses" to parent
Gametes
2n
+
2n
Second generation offspring
4n
4n
4n

(b) **Tetraploidy by passage through "triploid block"**



*The data in the graphs are from columbine plants, in which normal haploid gametes have n = 7 chromosomes and normal diploid offspring have n = 14 chromosomes

Types of polyploidization

- Within a species: chromosomes double or meiosis fails
- Between species:
  - F1 hybrid will fail at regular meiosis because its chromosomes cannot pair
  - One consequence of a failed meiosis is diploid eggs/sperm
  - These could generate a tetraploid offspring which would then be able to do meiosis (every chromosome has a partner)
- Such cross-species polyploids are fairly common in plants

Triploid toads

- Most triploids can't segregate their chromosomes in meiosis
- In *B. pseudoraddei baturae*:
  - Males eliminate "A" genome before meiosis and make haploid sperm
  - Females duplicate "A" genome and make diploid ova
- How did this evolve??
- (Probably an interspecies hybrid)

Bufo pseudoraddei baturae

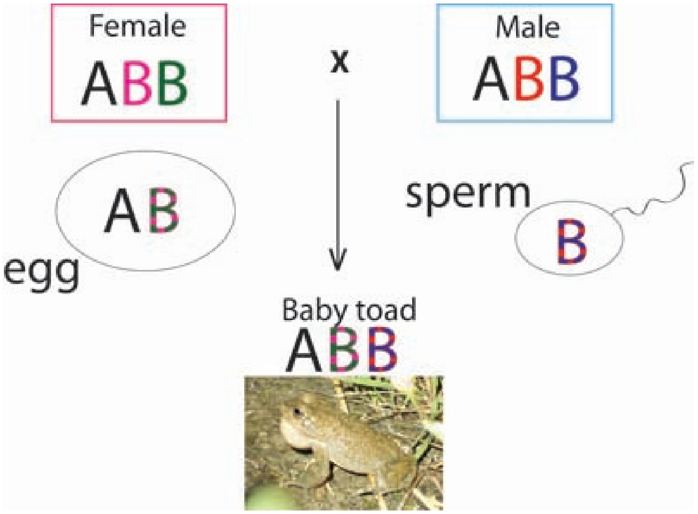
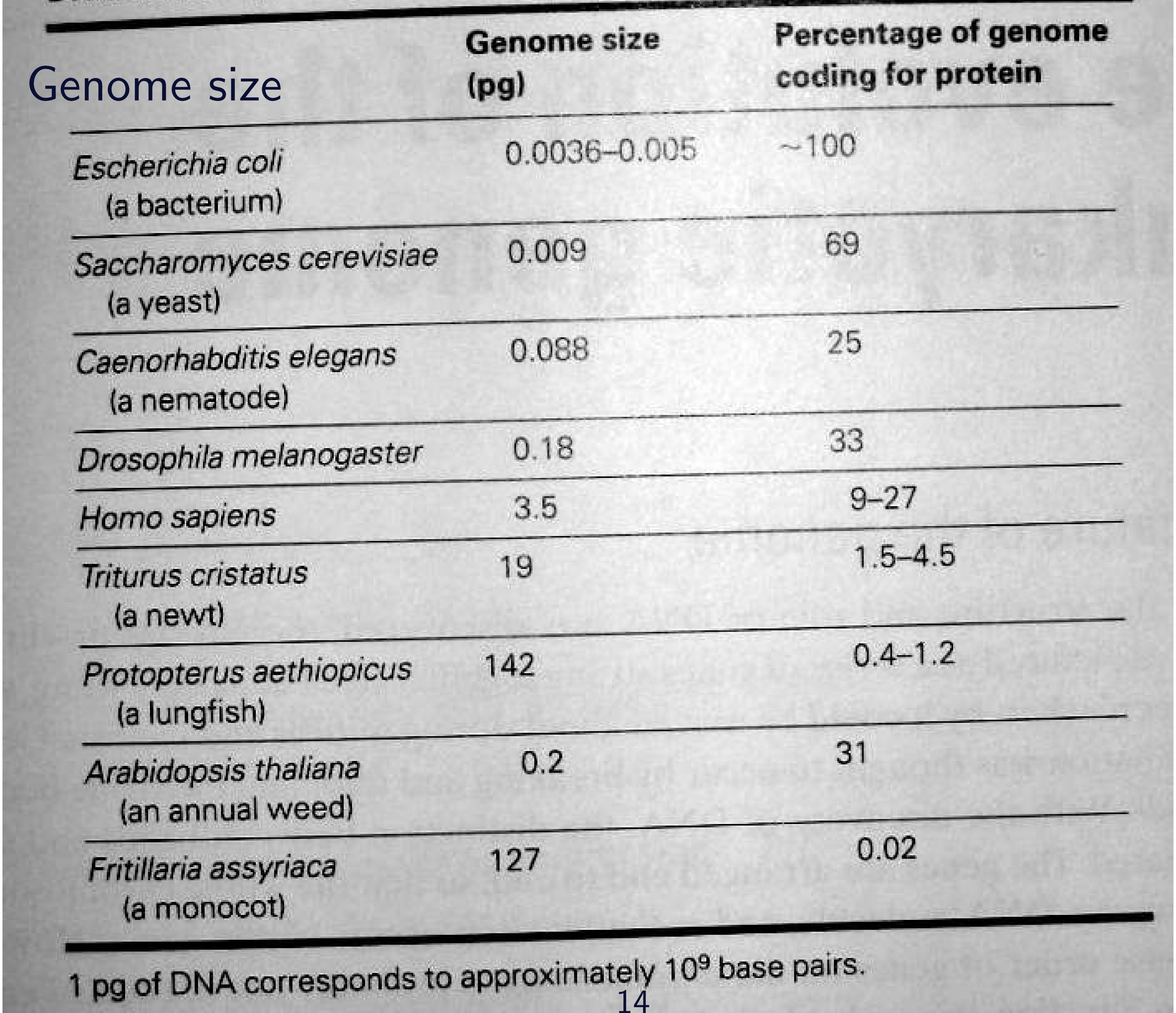


Image by Brooke

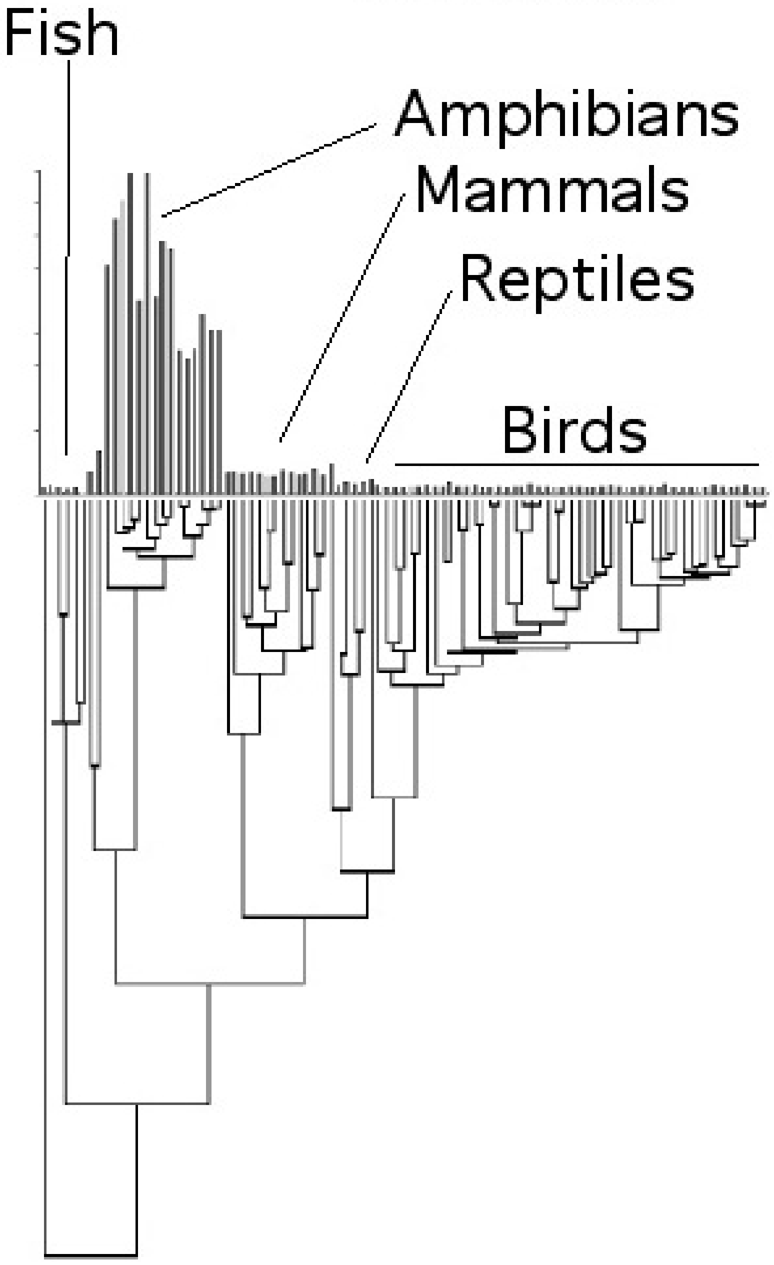
Genome size



	Genome size (pg)	Percentage of genome coding for protein
Escherichia coli (a bacterium)	0.0036–0.005	~100
Saccharomyces cerevisiae (a yeast)	0.009	69
Caenorhabditis elegans (a nematode)	0.088	25
Drosophila melanogaster	0.18	33
Homo sapiens	3.5	9–27
Triturus cristatus (a newt)	19	1.5–4.5
Protopterus aethiopicus (a lungfish)	142	0.4–1.2
Arabidopsis thaliana (an annual weed)	0.2	31
Fritillaria assyriaca (a monocot)	127	0.02

1 pg of DNA corresponds to approximately 10^9 base pairs.

Phylogenetic variation in vertebrate genome size



Typical number of genes?

Humans	25,000
Nematode	19,500
Drosophila	13,000
Arabadopsis	27,000

- Flies have a more dramatic metamorphosis than worms but do it with fewer genes

- Betting pool on the number of genes in the human genome:
  - In 2001 the mean guess was 66,000
  - Pool won by Lee Rowan of ISB with a guess of 25,947
  - This was the lowest guess received!

Minimal number of genes

- *Mycoplasma genitalium* completely sequenced (580 kb)
- Transposons were used to interrupt genes
- Genome codes for 517 genes (480 proteins, 37 RNA)
  - 1354 of the 2209 insertions were viable
  - An estimated 256 to 350 genes are essential
- “Essential” here means it’s possible to keep the organism alive without them; more are probably essential in the wild
- Craig Venter of Celera has proposed trying to create a minimal genome organism as a proof of concept

Unanswered questions

- Why does genome size vary so much more than gene count?
  - Are some species under selection to “fatten” their genome?
  - Are others under selection to streamline it?
- Smallest genomes are in intracellular parasites
- Largest genomes are in plants and amphibians
- Genome size appears to correlate with:
  - Cell size (big genome = big cell)
  - Cell division rate (big genome = slow division)
  - In some species, developmental rate (big genome = slow development)

Dinosaur genome size

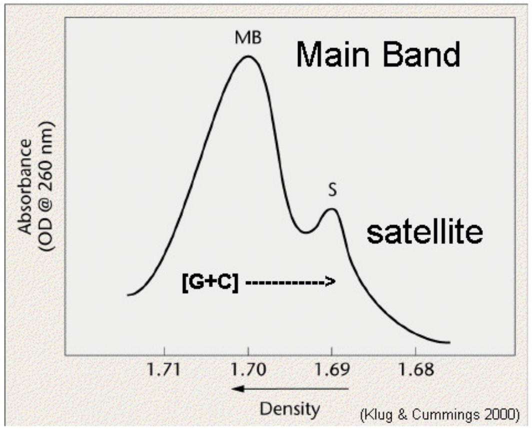
- Organ and Edwards tried to estimate dinosaur genome size:
  - Measure correlation between bone-cell size and genome size in modern animals
  - Measure bone-cell size in fossils
  - Infer genome size
- The dinosaur lineages thought to lead to birds had bird-sized (tiny) genomes
- Other lineages had larger (reptile-sized) lineages

Satellite DNA: localized repeats

- Large areas of the genome often composed of short repeats
  - In kangaroo rats the repeats AAG, TTAGGG, CAACAGCGCGGG make up about half of the genome

- First detected as minor peaks ("satellites") in a distribution of DNA densities

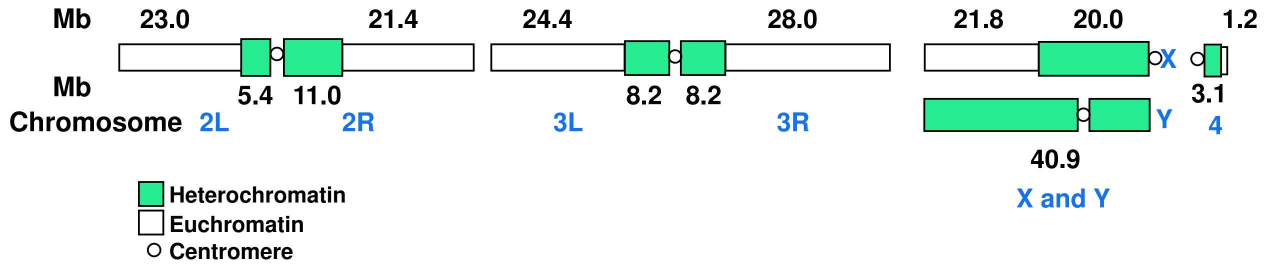
- Satellite DNA is often located near the centromere or at the telomeres



(Klug & Cummings 2000)

Satellite DNA is generally heterochromatin

- Heterochromatin is densely packaged DNA
  - Most genes are inactivated if they are moved into heterochromatin
  - A few genes can remain active there
- Fly heterochromatin patterns:



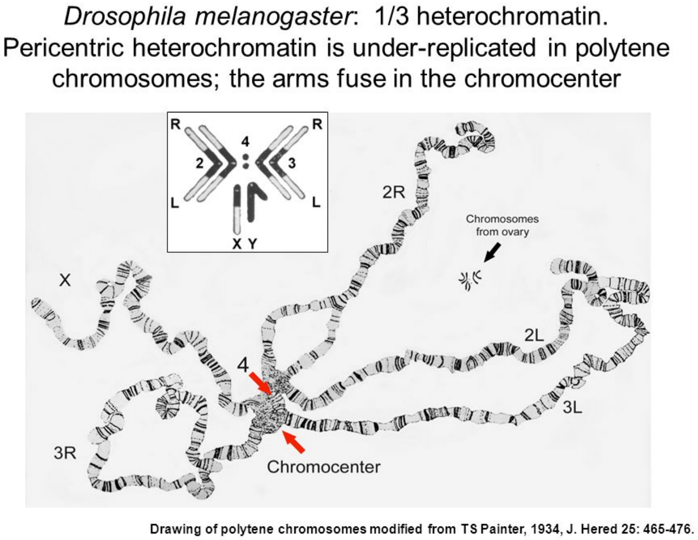
Function vs selfishness

- Does highly repeated DNA have a function? It might be involved in:
  - Chromosome physical structure–spindle fiber attachment, etc
  - Recognition of homologous centromeres
  - Chromosome pairing

- Possible evidence for selfishness:
  - fitness in Drosophila is not reduced when heterochromatin is reduced
  - Heterochromatin content varies widely between species

Satellite DNA varies among species

- *D. melanogaster* has about 30% satellite DNA, mostly at centromeres
- *D. nasutoides* has about 60% satellite DNA, but all on one chromosome.
- *D. littoralis, D. ezoana* have no satellite DNA



Drosophila melanogaster: 1/3 heterochromatin.
Pericentric heterochromatin is under-replicated in polytene chromosomes; the arms fuse in the chromocenter

Drawing of polytene chromosomes modified from TS Painter, 1934, J. Hered 25: 465-476.

D. melanogaster heterochromatin

Evolution of satellite DNA

- Strong concerted evolution:
  - Sequence is homogenized
  - Number of repeats varies widely among individuals

- If same satellite DNA on multiple chromosomes:
  - Illegitimate crossing-over can occur
  - Creates translocations and chromosome fusions at a high rate

- Could this help explain the Indian muntjak?

Tranposons as genome components

- Transposons are genetic elements that can move or copy themselves into new locations

- DNA transposons:
  - Cut themselves out and insert in a new site
  - 2% of human genome

- Retrotransposons:
  - Make RNA copy, reverse transcribe back to DNA, insert in a new site
  - 42% of human genome

Transposons as genome components

- Many transposon families have both active and inactive members
  - Initial transposon codes for genes needed in transposition
  - A copy can lose these genes but use proteins from other copies to transpose
  - Eventually copies deteriorate and stop transposing
- Within a genome, transposons evolve like organisms on their own
- However, if they evolve in ways contrary to host survival they die

Full-size transposons in humans

- LINE elements
  - “Long interspersed elements”
  - Retrotransposons which code for a reverse transcriptase
  - Three families in humans (LINE1, LINE2, LINE3)
  - LINE1 is still active and can move
  - LINE2 and LINE3 have stopped transposing
  - LINEs have about 850,000 copies in human genome or 21% of genome

Truncated transposons in humans

- SINE elements
  - “Short interspersed elements”
  - Have lost the reverse transcriptase gene
  - Can transpose only in presence of related LINE
  - Alu family of SINEs is the biggest gene family in humans
  - SINEs have about 1,500,000 copies or 13% of genome

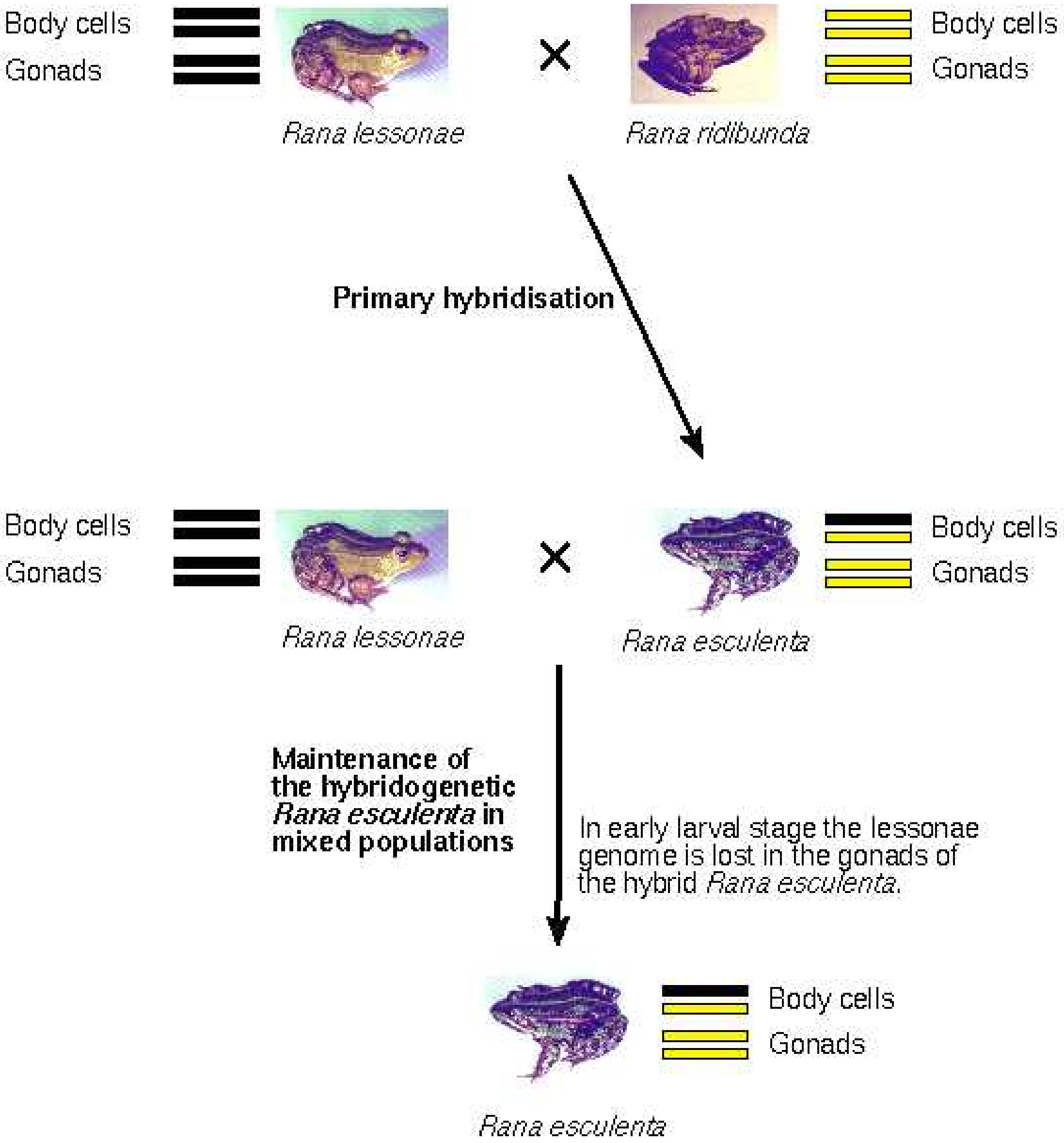
Are retrotransposons viruses?

Two hypotheses:

- Retrotransposon evolved from an RNA virus:
  - Reverse transcriptase function evolved by virus
  - Virus lost ability to leave the cell
- Virus evolved from a retrotransposon:
  - Cell evolved reverse transcriptase for its own use
  - Transposon picked up a protein that could form a virus capsule and became a virus
- Hard to see how an RNA virus could evolve reverse transcriptase; easier for me to imagine a cell doing it

Pure weirdness

- *Pelophylax esculenta* appears to be a species of water frog
- Originally it seems to have been an inter-species hybrid
- It prefers to mate with one of its parent species
- Why doesn't it get diluted out?



Primary hybridisation

Maintenance of the hybridogenetic *Rana esculenta* in mixed populations

In early larval stage the lessonae genome is lost in the gonads of the hybrid *Rana esculenta*.

One-minute responses

- Tear off a half-sheet of paper
- Write one line about the lecture:
  - Was anything unclear?
  - Did anything work particularly well?
  - What could be better?
- Leave at the back on your way out