

# Genome Sequencing: What are We Dealing With?

part of “High-Throughput Analyzes of Genome Sequences”

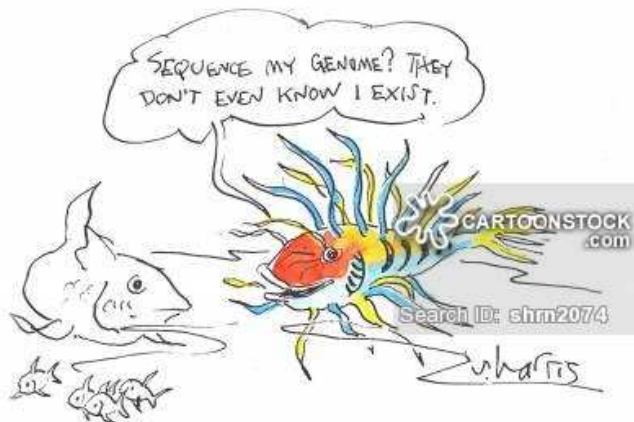
Sonja J. Prohaska

Computational EvoDevo  
University of Leipzig

Leipzig, WS 2014/15

# Sequenced Genome?

What would you want to know?

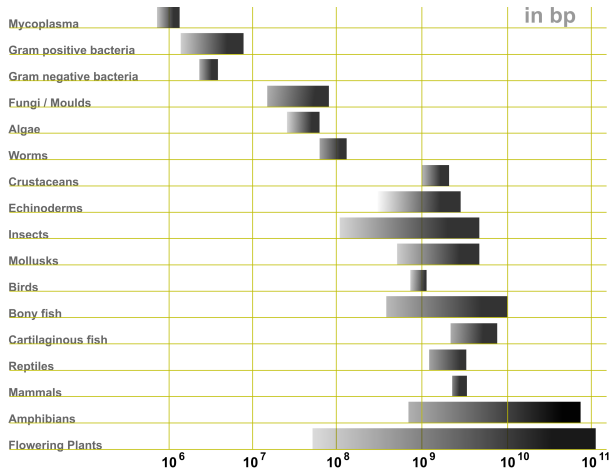


# What would you want to know?

- ▶ what species is the genome from?
- ▶ what is the genome size?
- ▶ what is the ploidy?
- ▶ what is it composed of?
- ▶ what is its structure?
  - ▶  $k$ -mer distributions
  - ▶ self-similarity
  - ▶ junk? (simple repeats, transposable elements)
  - ▶ functional protein coding genes?
  - ▶ functional non-coding RNAs?
  - ▶ functional regulatory elements?

# Genome Size per Nucleus

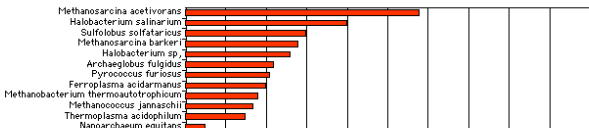
- ▶ **mass:**  $10^{-9}$  grams = 1pg
- ▶ **base pairs (bp):** 1pg 978Mb =  $0.978^9$  bp



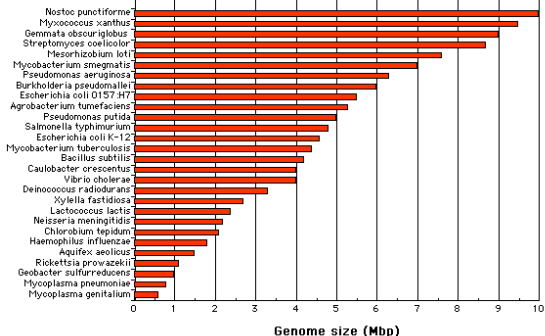
Mycoplasma – smallest genome size among bacteria. Sizes are given for the haploid genomes.

# Genome Size

## Archaea:



## Bacteria:

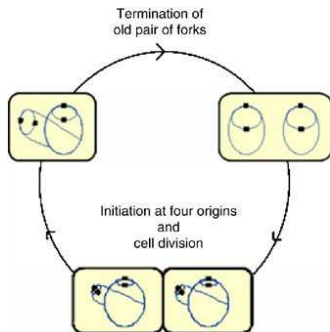


Bacteria and archaea both have circular genomes.

## Ploidy – haploid or diploid?

ploidy ... number of chromosome sets per nucleus

bacteria and ploidy?

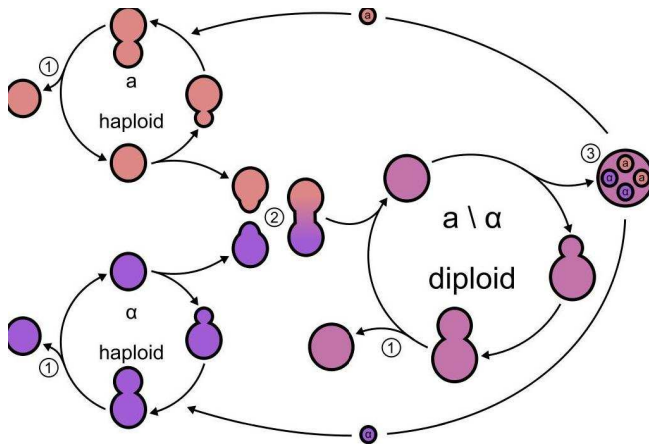


Here, for replicated origins exist before the first of two cell divisions is completed.

Depending on rate of replication and cell division.

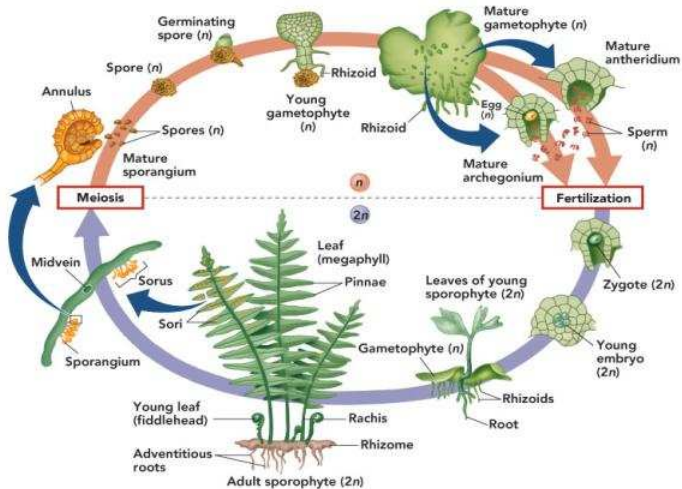
# Ploidy – haploid or diploid?

yeast – haploid and diploid phase in the life cycle



# Ploidy – haploid or diploid?

ferns – haploid and diploid phase in the life cycle



Copyright © 2004 Pearson Education, Inc., publishing as Benjamin Cummings.



# Ploidy – haploid or diploid?

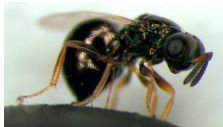
ant *Hymenoptera* – (haplodiploid)

**general**



- ▶ fertilized eggs – diploid females
- ▶ unfertilized eggs – haploid males

**wasp – *Nasonia vitripennis***

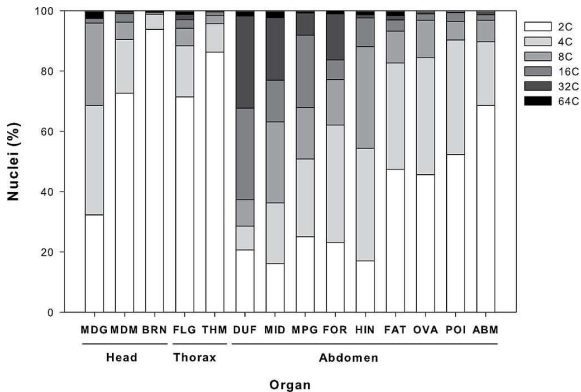


- ▶ some haploids – gynandromorphs  
individuals with female  
(and male) features

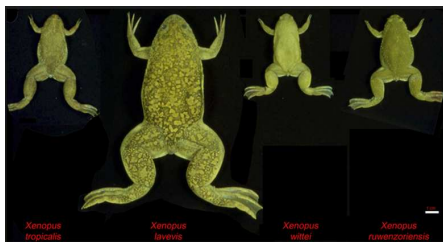
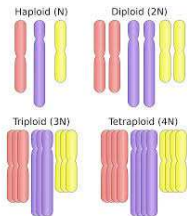
# Ploidy – haploid or diploid?

## ant – *Dinoponera australis*

- ▶ organ-specific endopolyploidy



# Polyploidy



- ▶ *Xenopus tropicalis* –  $2n$  – diploid (normal)
- ▶ *Xenopus laevis* –  $4n$  – tetraploid
- ▶ *Xenopus wittei* –  $8n$  – octaploid
- ▶ *Xenopus ruwenzoriensis* –  $12n$  – dodecaploid

# Polyploidy

Food and crops are often selected to be polyploid to increase the yield.

- ▶ black mulberry – natural docosaploid  $22n$
- ▶ sequoia (redwood) – natural  $6n$
- ▶ broccoli –  $4n$
- ▶ wheat –  $2n$ ,  $4n$  or  $6n$
- ▶ seedless watermelon –  $3n$
- ▶ kiwifruit –  $6n$
- ▶ certain strawberries –  $10n$

## Number of Chromosomes

species	2n number of chromosomes
jack jumper ant	2(1)
fruit fly	8 (2 sex-)
mouse	40
human	46 (2 sex-)
chimp	48
zebrafish	50
platypus	52 (10 sex-)
horse	64
wolf	78
chicken	78 (10 macro-, 10 inter-, 2 sex-, 56 mini-)
carp	104
field horsetail	216
ciliates	10 in Micronucleus 15,000 in Macronucleus

Minichromosomes < 20Mb are frequent in amphibia and sauropsida, particularly aves.

Macrochromosomes > 40Mb.

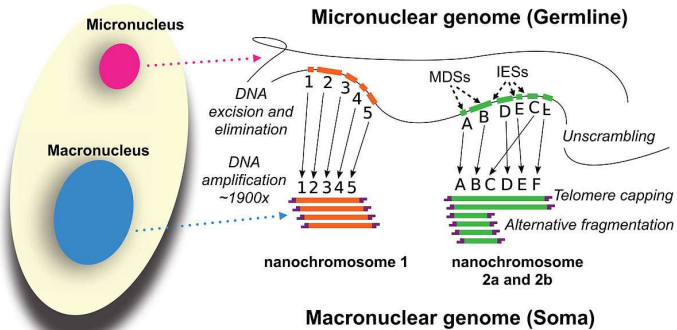
# The Ciliate Genomes



Oxytricha



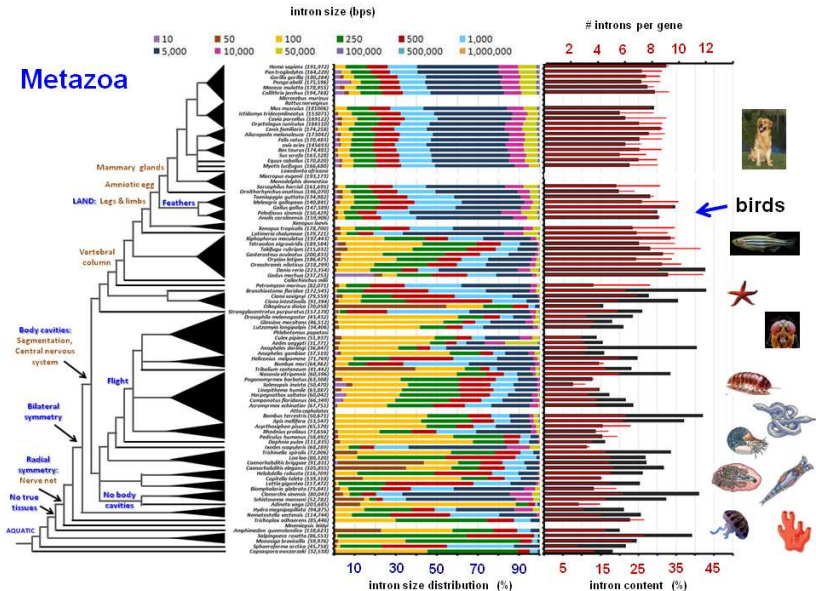
Tetrahymena



## Some genome statistics - human

- ▶ less than 1.5% coding sequence
- ▶ 26% introns
- ▶ 8-20% regulatory sequence
- ▶ 50% repetitive sequence
- ▶ 8% simple repeats
- ▶ 80% transcribed! (ENCODE)

# Some genome statistics - intron number and size





# Sequences Missing From a Sequenced Genome

- ▶ telomere sequence
- ▶ centromer sequence
- ▶ microsatellites
- ▶ ...

# Breaking a Central Assumption

“the genome is the same in all cells of a multicellular organism”

- ▶ ciliates
- ▶ lempray (germline and soma different)
- ▶ human immune cells recombine DNA (V(D)J-joining)
- ▶ DNA amplification (Drosophila salivary glands)
- ▶ cancer cells (copy number variation)

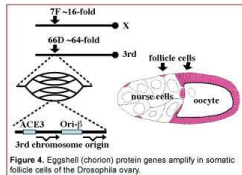
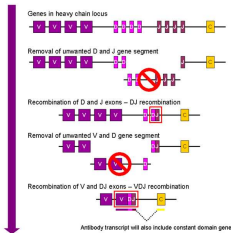
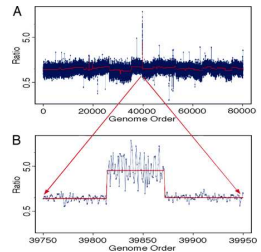


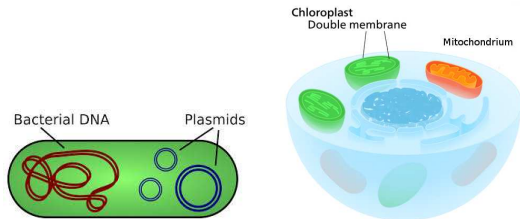
Figure 4. Eggshell (chorion) protein genes amplify in somatic follicle cells of the Drosophila ovary.



# Other genomes in the genome/cell/organism

## Sources of DNA other than the nuclear genome

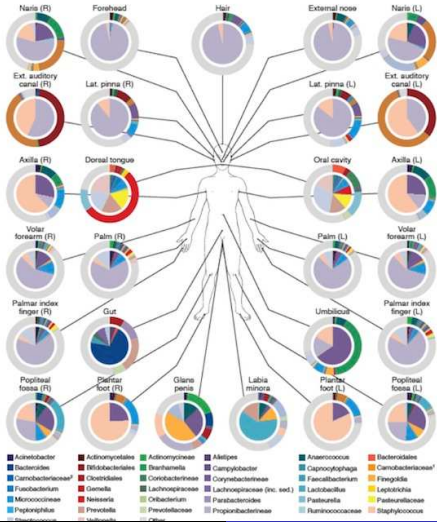
- ▶ many different plasmids in bacteria
- ▶ organelles in eukarya: mitochondria, plastides
- ▶ i.e. endosymbionts (assimilated over time)
- ▶ current symbionts, parasites, viruses, selfish genes,...?



Sequencing the whole organism? →

# Microbiome – Metagenome

is "the ecological community of commensal, symbiotic, and pathogenic microorganisms that literally share our body space."



# Literature