

# Struktura a analýza rostlinných genomů

Jan Šafář

Ústav experimentální botaniky AV ČR, v.v.i  
Centrum regionu Haná pro biotechnologický a  
zemědělský výzkum



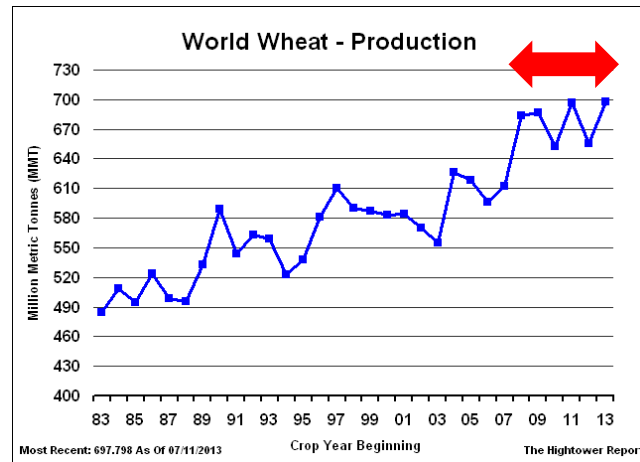
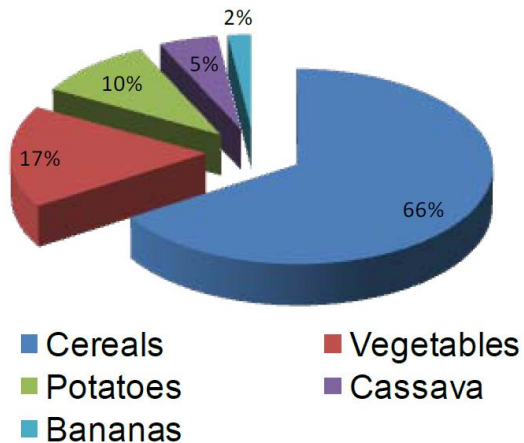
# Proč rostliny? Proč genom?



Norman E. Borlaug

## „Zelená revoluce nese zlatou žeň“ (1969):

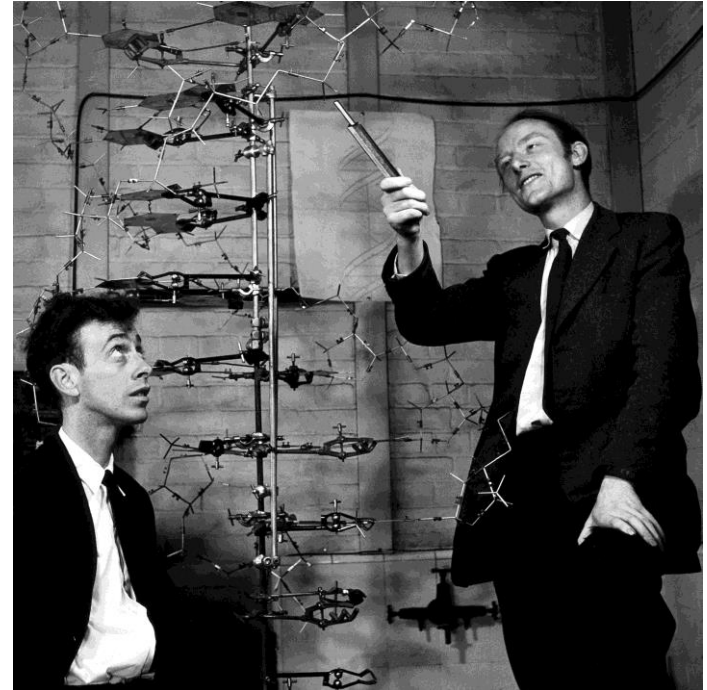
- Současný úspěch se zvýšením produkce pšenice, rýže a kukuřice v zemích Asie poskytuje 20 - 30 let na systematické řešení problému
- Zajištění výživy lidstva vyžaduje zvýšení produkce potravin o 60% v průběhu let 2000 and 2050
- Během 50 let lidstvo vyprodukuje takové množství potravin, jaké spotřebovalo za celou dobu své existence



- **Metody molekulární biologie, biotechnologie a genomiky**

# Dnešní menu

- Co je to genom?
- Z čeho se skládá?
- Jak se vyvíjel?
- Jak ho čteme?



# Co je to genom?

- Hans Winkler (1920): **Gen + ome**  
(„ome“ = velký/úplný soubor jednotek)

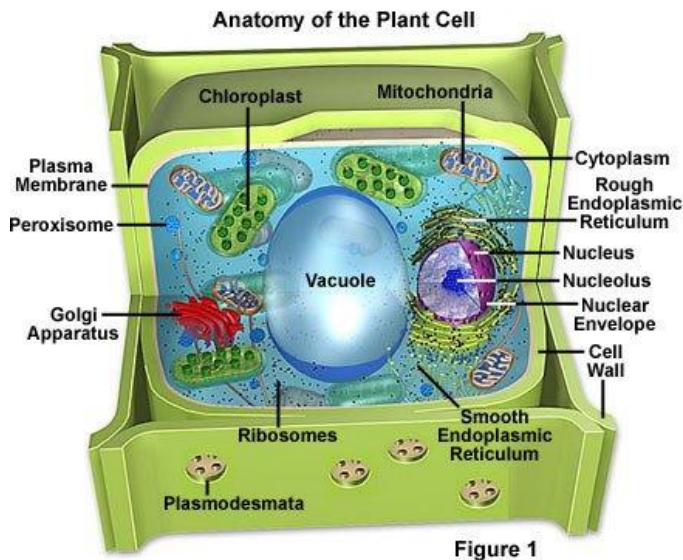
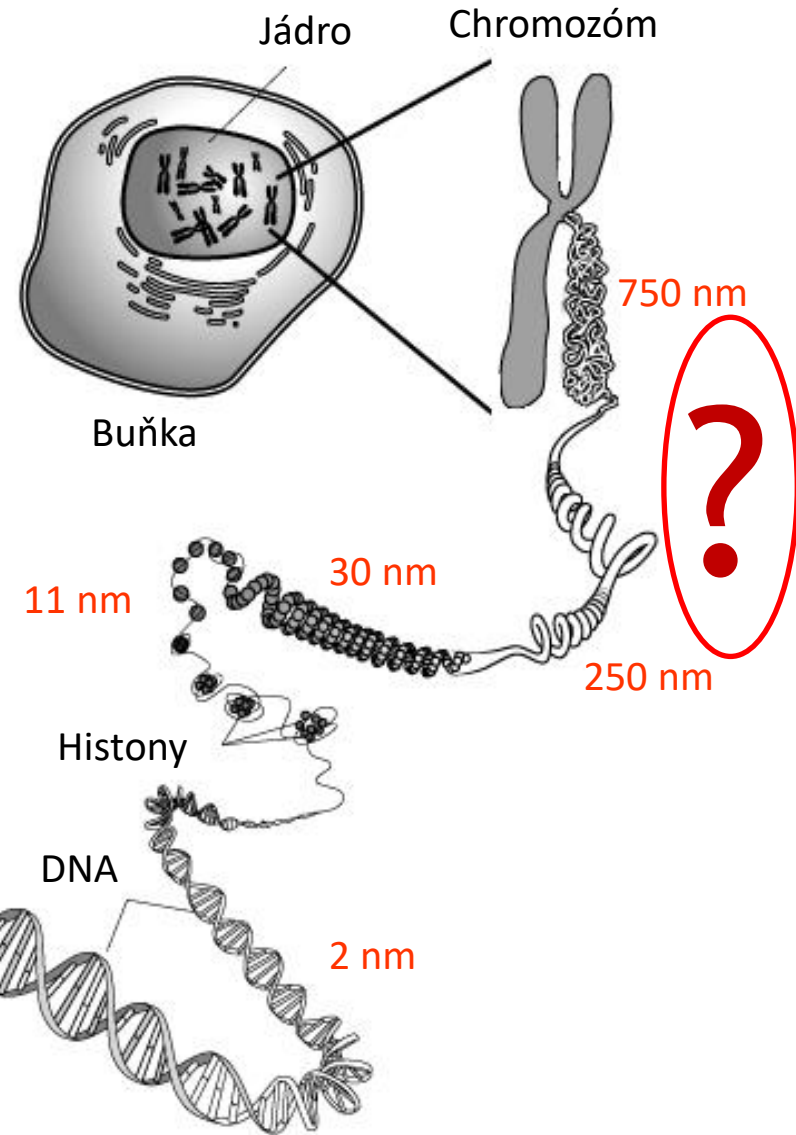
„Navrhuji termín **genom** pro haploidní sadu chromozómů, která spolu s protoplazmou představuje materiální základ druhu.“

- **Genom** = veškerá DNA, která představuje jednu úplnou kopii dědičné informace organismu
- **Pangenom** = veškerá DNA, která představuje geny všech jedinců daného druhu



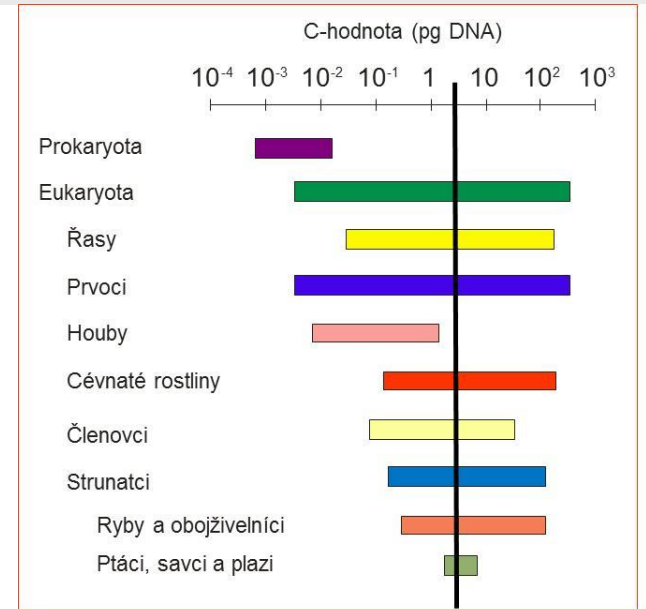
# Většina DNA se nachází v jádře buňky

- **Chromatin**  
(DNA, RNA, proteiny)
- **Chromozómy**  
(v diploidní somatické buňce 2 homologní sady)

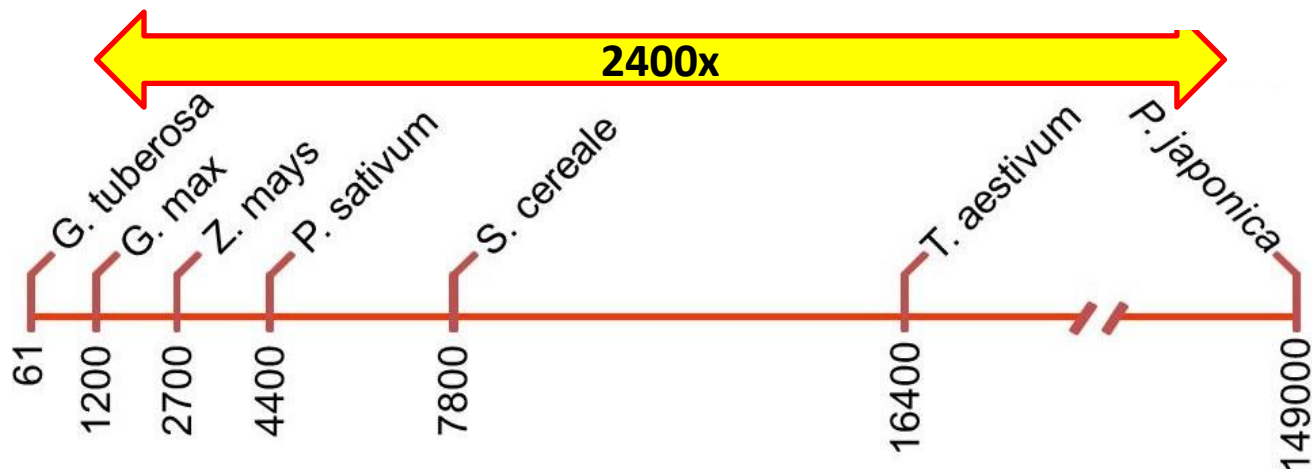


# Záhada C-hodnoty

- Dříve „Paradox C-hodnoty“ (C.A. Thomas, 1971)
- C-hodnota vykazuje slabou korelaci se složitostí organismu
- Krytosemenné:



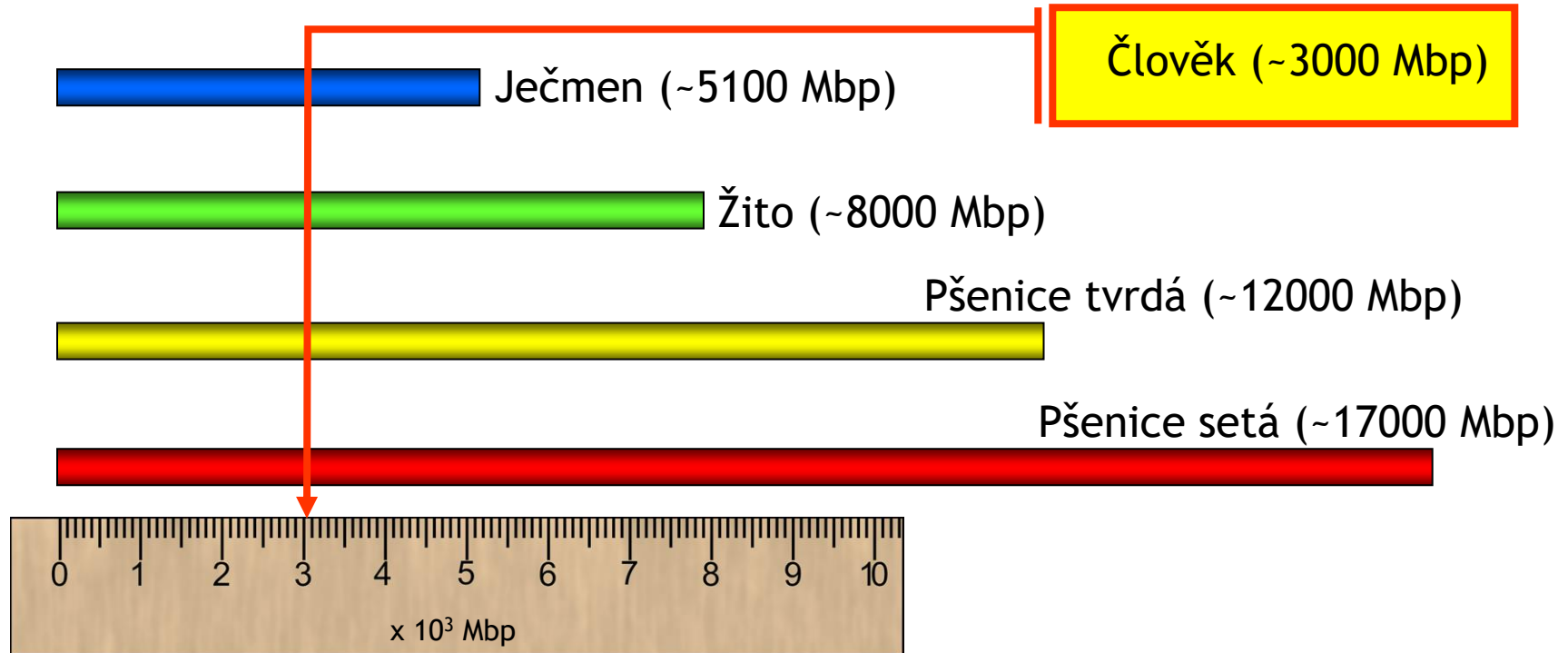
Člověk (~3,5 pg DNA)



Velikost jaderného genomu (1C/Mbp)

# Genomy zemědělských plodin

- Mnoho důležitých plodin má velké a složité genomy



Genom pšenice obsahuje 3000x více písmen dědičného kódu než všechny svazky cyklu o Harry Potterovi



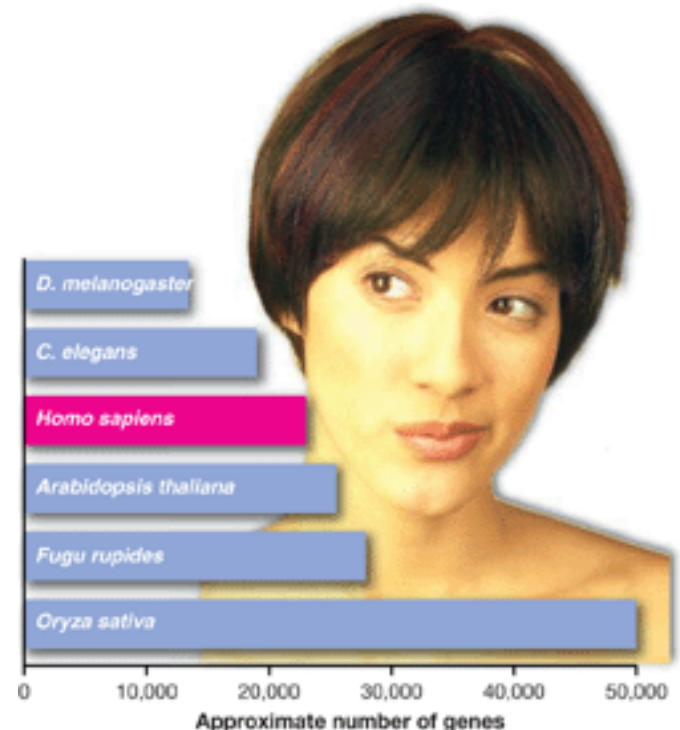
# Z čeho se jaderný genom skládá?

## 1) Geny a genům podobné sekvence

- Strukturní geny (kódující polypeptidy; obvykle 1 - 10 kopií)
- Geny pro funkční RNA (např. rRNA, tRNA; více kopií), miRNA
- Regulační oblasti DNA

## Kolik genů je v genomu?

- Nepříliš přesné odhady:  
25 000 - 50 000 genů
- Počet genů je u všech druhů velmi podobný a představují pouze 1 - 30% genomu
- „Gene space“



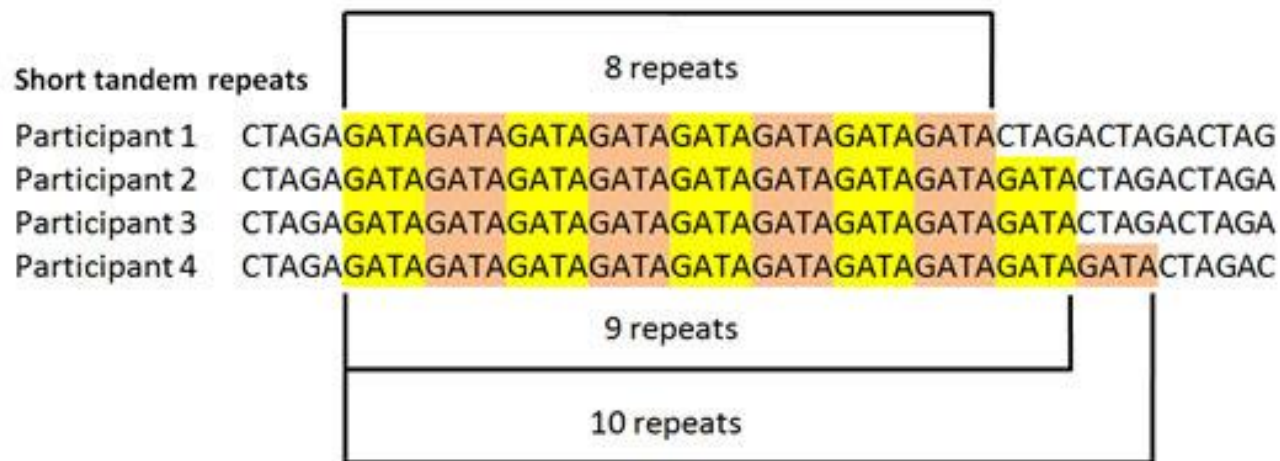


# Co je tedy hlavní složkou genomu?

## 2) Negenové sekvence

- Obvykle tvoří většinu genomu
- Hlavní příčina rozdílů ve velikosti genomu mezi druhy
- V genomu se obvykle mnohokrát opakují  
-> repetitivní sekvence DNA

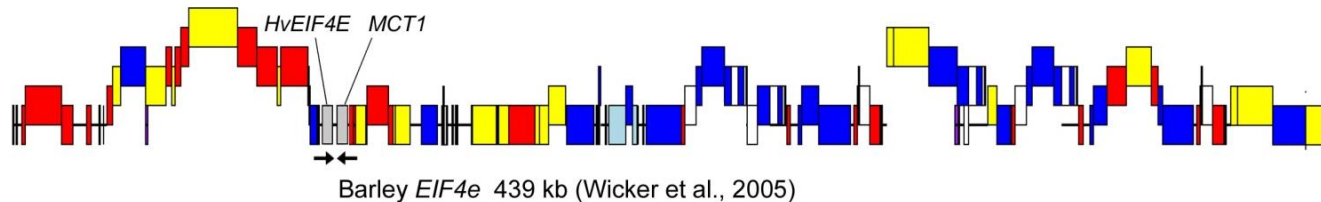
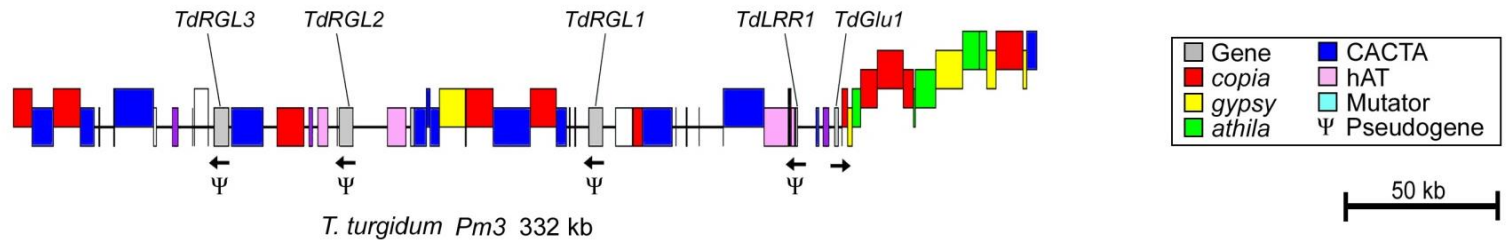
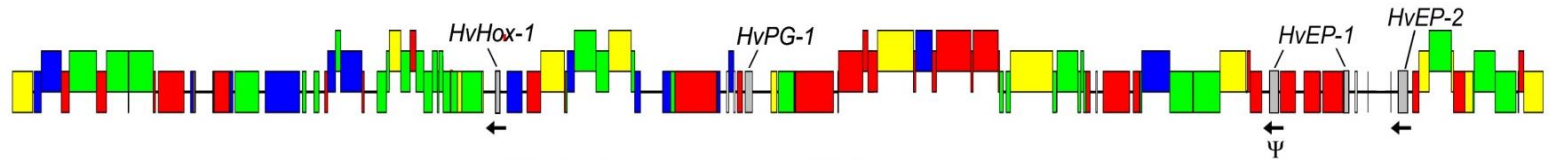
DNA satelity - uspořádané za sebou (tandemově):



Chromozómy 3B  
pšenice seté

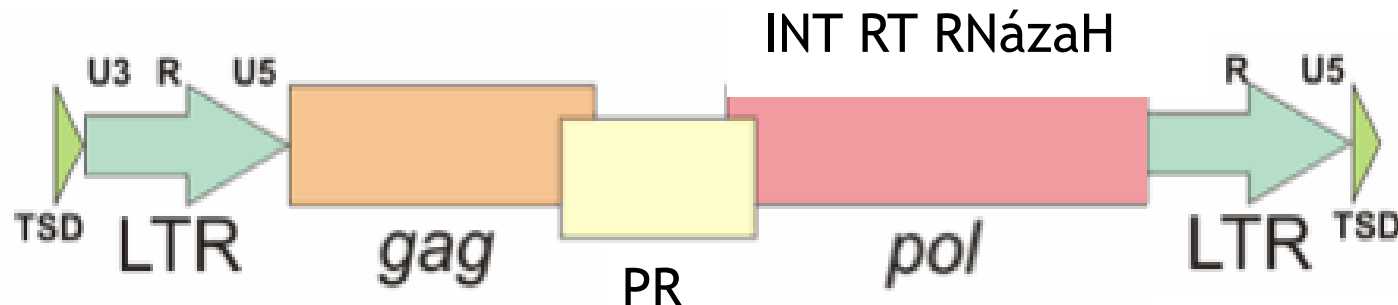
# Rozptýlené negenové sekvence

- Mobilní genetické elementy - transpozóny (DNA transpozóny a retrotranspozóny)
- Tvoří největší část genomu (u rostlin až 80%)
- Inzerce transpozónů do sekvencí jiných transpozónů



# Retrotranspozóny

- Struktura podobná retrovirům, ale netvoří virulentní částice:
  - Délka několik kb, obvykle LTR
  - „LTR retrotranspozóny“ (Ty1-copia, Ty3-gypsy)
  - Bez LTR (LINE, SINE)
- Autonomní i neautonomní elementy
- Regulace aktivity retrotranspozónů: Heterochromatinizace vs. Aktivace
  - modulace exprese genů
  - inaktivace/duplikace/ztráty genů
  - evoluce genomů



# Regulace genové exprese retrotranspozóny

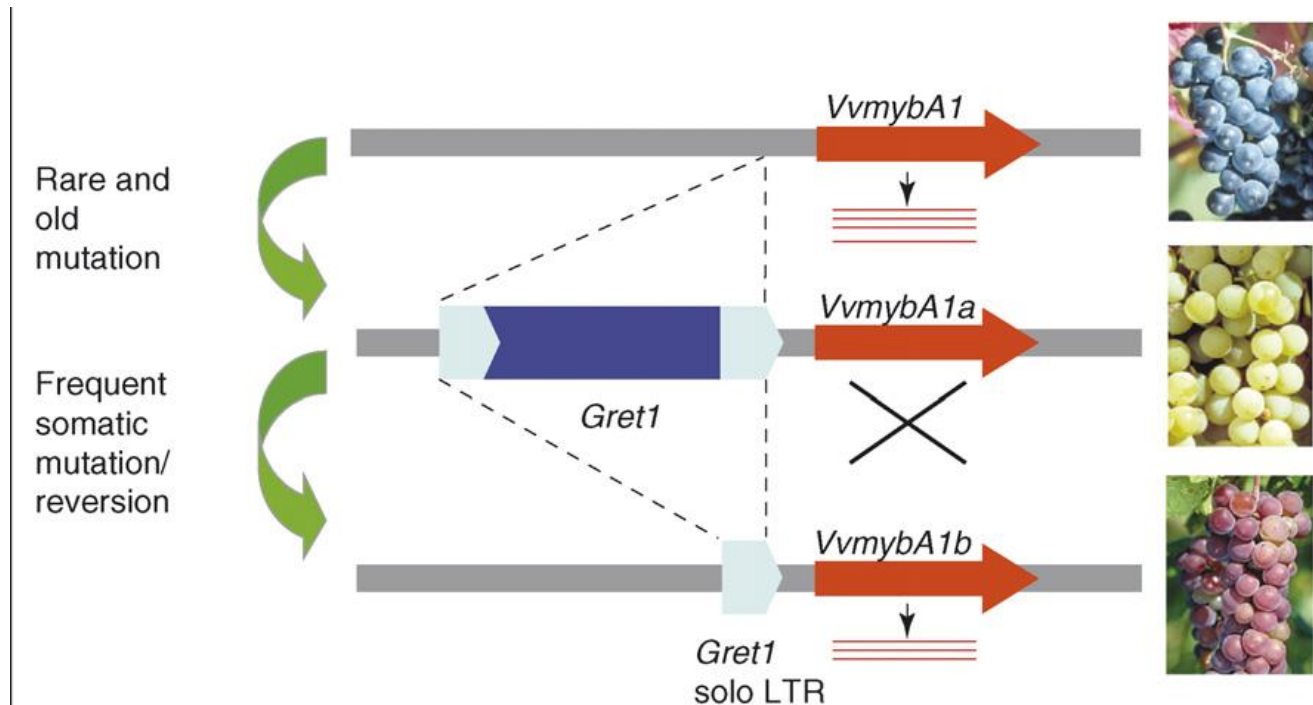
- Význam při domestikaci/šlechtění rostlin:

Příklady:

pomeranč: gen Ruby regulovaný inzercí copia RT (Butelli et al. 2012, Plant Cell)

réva: inaktivace exprese TF VvmybA1 přítomností gypsy RT (Kobayashi et al. 2004, Science)

kukuřice inaktivace genu *CCT* inzercí CACTA-like elementu rozšíření pěstování do mírného pásma (Yang et al., PNAS, 2013)



# Co udělaly 3 transpozóny s genomem rýže

*Oryza sativa* (390 Mbp)

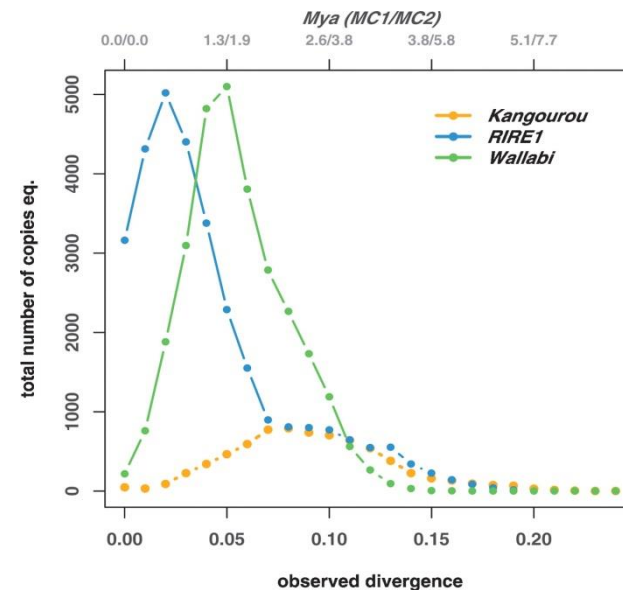


*Oryza australiensis* (965 Mbp)



*Oryza australiensis*

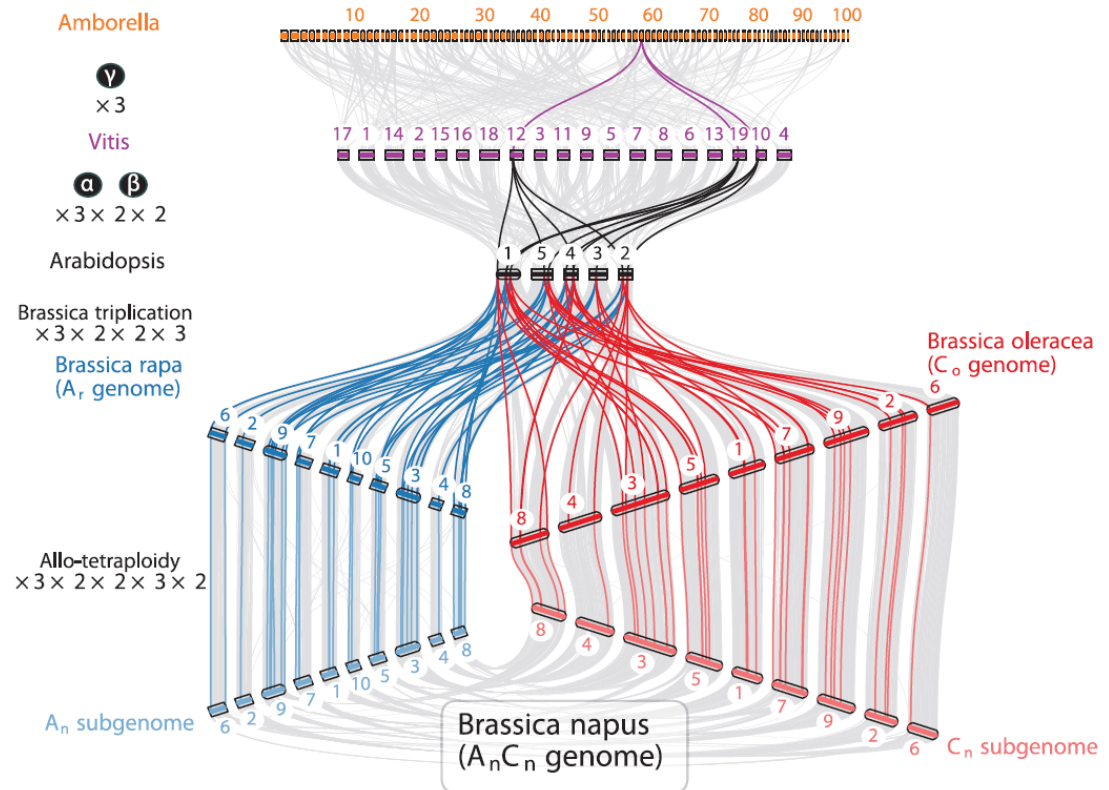
Transpozón	Kopii	DNA
<i>RIRE1</i>	40000	265 Mbp
<i>Kangourou</i>	10000	90 Mbp
<i>Wallabi</i>	39000	250 Mbp
Společně		605 Mbp





# Opakované cykly polyploidizace

- Krytosemenné rostliny jsou starobylí polyploidi
- Polyploidizační události byly následovány přestavbami karyotypu
- Vyšší adaptabilita a zvýšená tolerance polyploidů vůči nepříznivým podmínkám



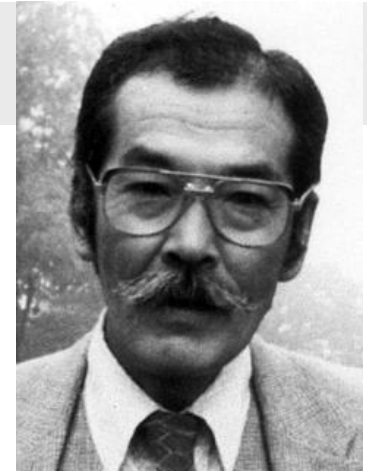
- *Amborella trichopoda* je nejbazálnější žijící zástupce krytosemenných rostlin (výskyt: Nová Kaledonie)
- Typická oblast genomu *A. trichopoda* odpovídá až 72 oblastem genomu aloteraploidního druhu *Brassica napus*

# Evoluce genovou duplikací

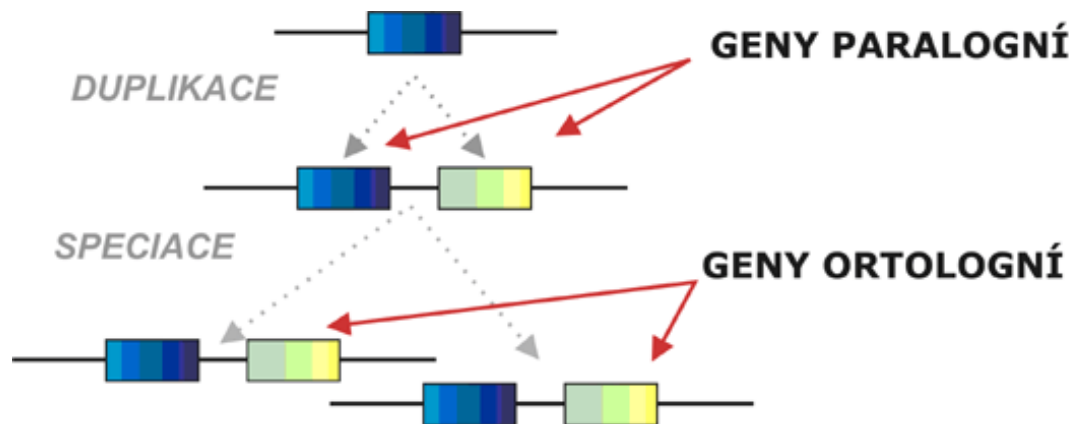
„Je snazší tvořit nové geny ze starých, než je vytvářet *de novo*“

## Důsledky polyploidizace

- Genomický šok (přestavby genomu a změny genové exprese)
- Uvolněný selekční tlak na duplikované geny a akumulace mutací



(S. Ohno, 1970)



Geny jednoho druhu, které vznikly duplikací ancestrálního genu (mohou se navzájem lišit funkcí)



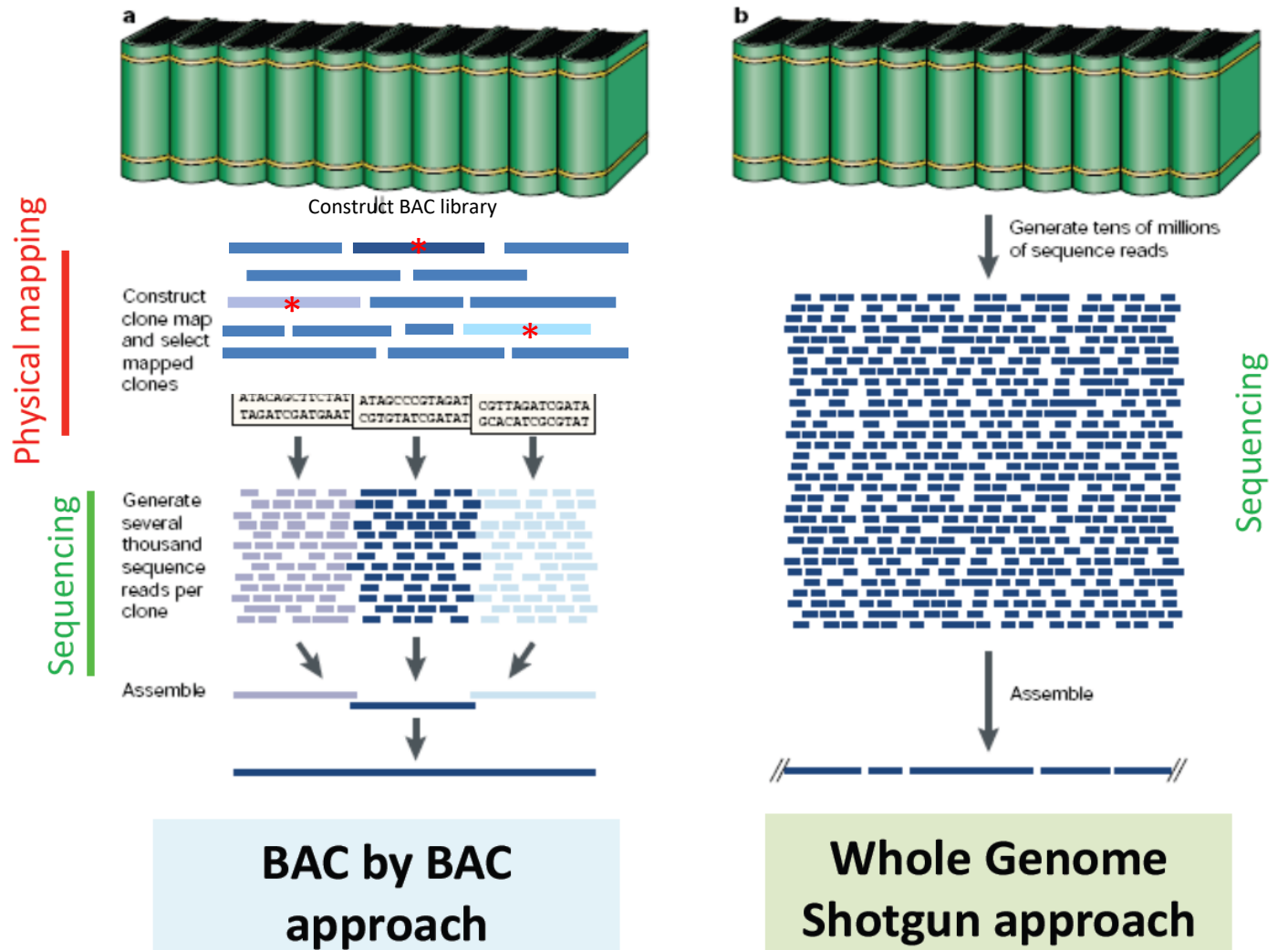
# Analýza genomu - sekvenování

proces určování primárního pořadí nukleotidů v molekule DNA

```
TTGTTTTTCTGCCGATATTTTTGGTAGCAATCTTCACATGCCCGTTCAGTGCTCAATGTTG
GCGACACTTAGGAATAGAGTGGAATCAAATTTGGAGTTCTTTCAGATGATCACTCTAGC
AAGACTCAAGTTTGCTTGGAGGGGCTTTCTGGAGGTTTTCTTCTTAGCAGCATGGCACA
TTTGGAAACAACGAAATAGTTTAATCTTCCAGAATATTCAGCCCTCCTTCCAAGCTTGGAA
CCCACCACCTGCACCACCTCCAAACCCACTGCCACCACCCATTCCACCTCCTCCGAGTT
CGCCCCCACCACCAACTCCACCTCCTGCACCACCTCCAAACCCACCGCCACCACCCAT
TCCACCTCCTCCGAATCTGCCCCCACCACCAACTCCACCTCCTGCACCAGCACCTCCTC
CGAAACCGCCACCAGCACCAATGCCACCACCACCCCGGCACCACCACCAATCCACC
GCCACCACCCGCCCTACTCCACCTCCACCCCTGAGACCACCACCGGCACCTACACCC
CCTCCGCTTCCGAATCCAGCACCAACACCACCGCCAACCTCCACCTCCACCTCCACCTCC
AAACCCGCCACTGCCACCCATGCCTCCTCCACCAAGACCACCACTCCCTCCTAAGCCTC
CGCCTCCACCTCCACCTCCACCCAAACCACCTGCAGCCCCACCGCCAGTGCCACCACC
CACGCCGCCACCTACACCAATGGTAAAAAAAATATTTTTGGTAGCAAACCGAATGAACAC
CCTAGCTTATTGGCATTGCCAAAATTTTTCTTTTTAGCATAATTAAGAAGGGAGATAGACCA
AGCCCTTCATTCCAATATTCAGCTAAAAGACCCATGGTGATATCCAAAACCTACAATCAAG
AAAGAAAGCAGATTGAGTATAGTGGAATAGGCGAACGAGCCTCTAACTGAGTTTGATTA
GGTGGCTCTAATAACACTCCTCAGGTATTGGTTCGACTGCGAGTTTCAGGCACACACTAA
AGCACAGTTCTAAGGCTAAATCCCCTCGTCTATCACACGCTAAAGCACAATTCTAAGGCTT
AGCTCAAGTCGTTGGTCGTTTTTCACACGGGTTACGCTGCTGCTGTGTAAGCAAATAACCT
AGGGTTATCTTAC
```

# Strategie sekvenování

A: redukční B: hierarchické C: celogenomové neuspořádané

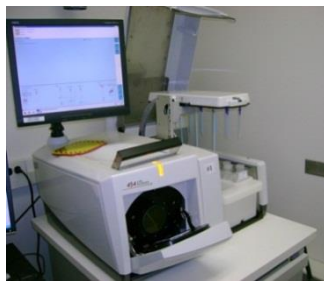




# Sekvenování (čtení) genomů rostlin

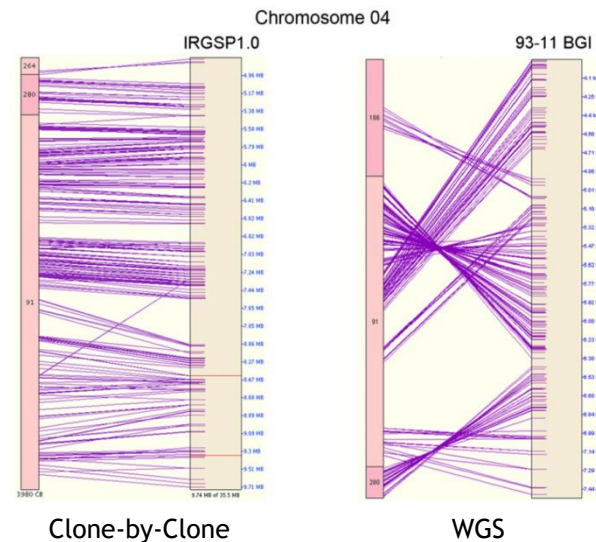
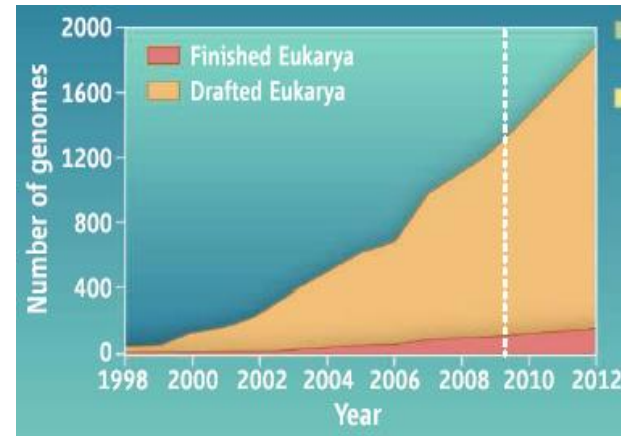
## Problém s velkými genomy

- NIKOLIV kapacita sekvenátorů
- ALE potíže se sestavováním mnoha krátkých sekvenačních čtení s převahou repetitivní DNA (... a polyploidie)
  - Délka čtení by musela přesahovat 200kb aby bylo možné sestavovat úseky genomů bohaté na repetitivní DNA
  - Současné metody „next generation sequencing“ takových parametrů zdaleka nedosahují



# Kvalita *de novo* sekvenování genomů

- Potřeba vysoké kvality sekvence referenčního genomu (šlechtitelská) u plodin
- Akceptovatelná kvalita *de novo* sekvenovaných genomů klesá
- Důsledek masivního používání NGS
- ... a podcenění složitosti rostlinných genomů



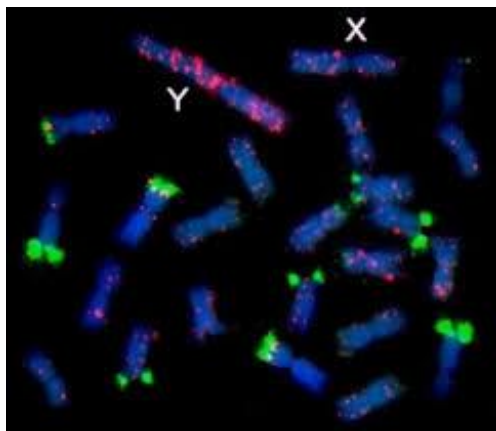
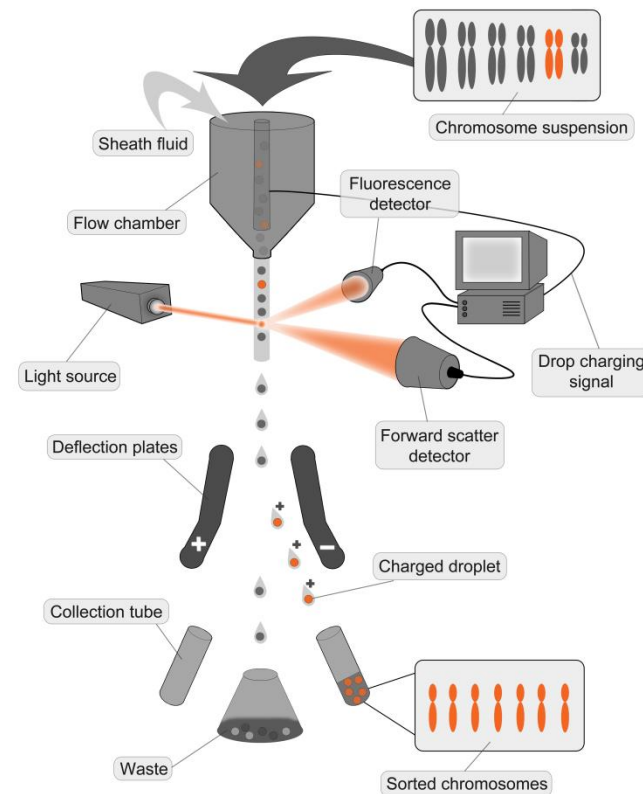
Pan et al., Plant J. 77:795-805, 2014



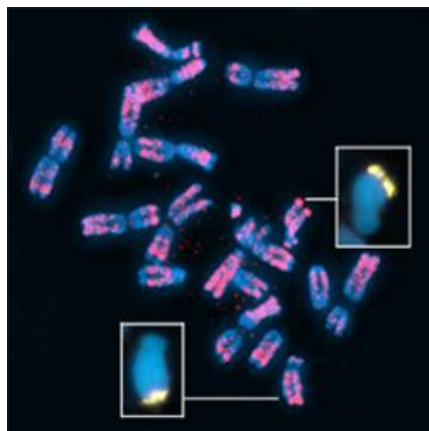
# Jde to i jinak aneb chromozómově-centrický přístup

## Nabízí možnost analyzovat

- Pohlavní chromozómy
- Přidatné / nadbytečné (B) chromosomy
- Specifické (aberrantní) chromozómy
- Allele phasing
- Strukturní chromozómovou heterozygotnost



Pohlavní chromozómy knotovky bílé



B chromozómy žita

# A co dál ?

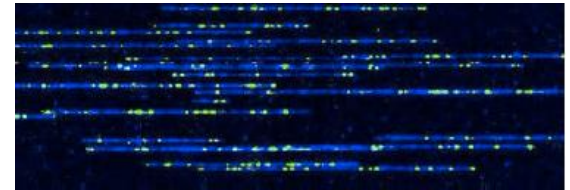
- **Strukturní analýza**

  - Lepší NGS technologie (PacBio, Oxford Nanopore)

  - Lepší bioinformatické nástroje (NRGene)

  - Optické mapování

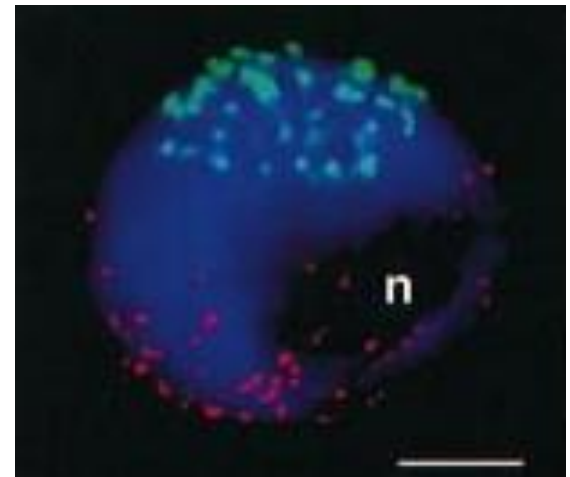
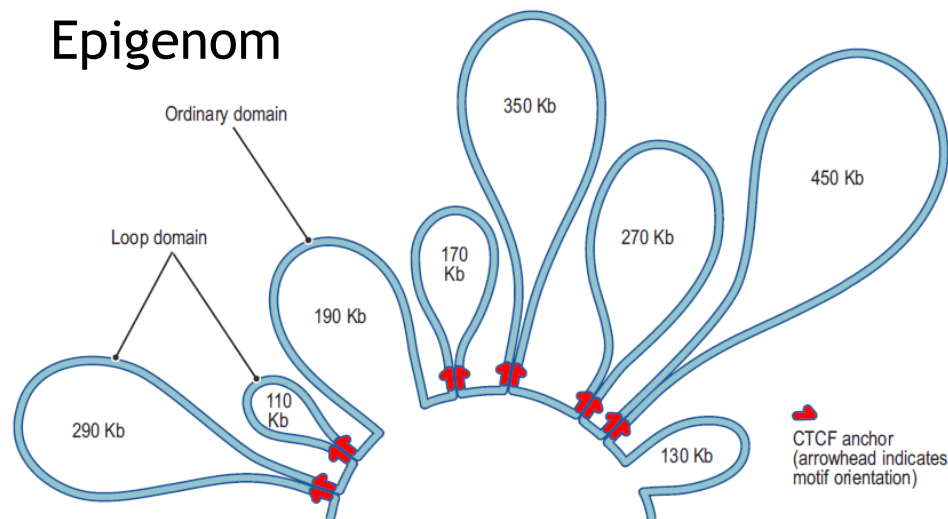
  - Sekvenování dalších kultivarů/variet



- **Funkční anotace** (identifikace a projev genu/alely)

  - 3D struktura jádra (domény, smyčky, exprese)

  - Epigenom







**Děkuji vám za pozornost**