

Molekulární přístupy ve šlechtění rostlin

## Aplikovaná genomika

# Genotypování: Využití ve šlechtění a určení identity odrůd

Miroslav Valárik

14.2. 2017



# Šlechtění rostlin:

---

**Cílený výběr a manipulace s genomy rostlinných druhů za účelem získání žádaného fenotypu a genotypu.**

**Šlechtění zahrnuje kontrolované křížení nebo genetické inženýrství a nebo oboje zároveň, následované umělou selekcí potomstva.**

**Šlechtění často vede k domestikaci.**

# Efektivní a rychlejší šlechtění rostlin = využití genomických zdrojů

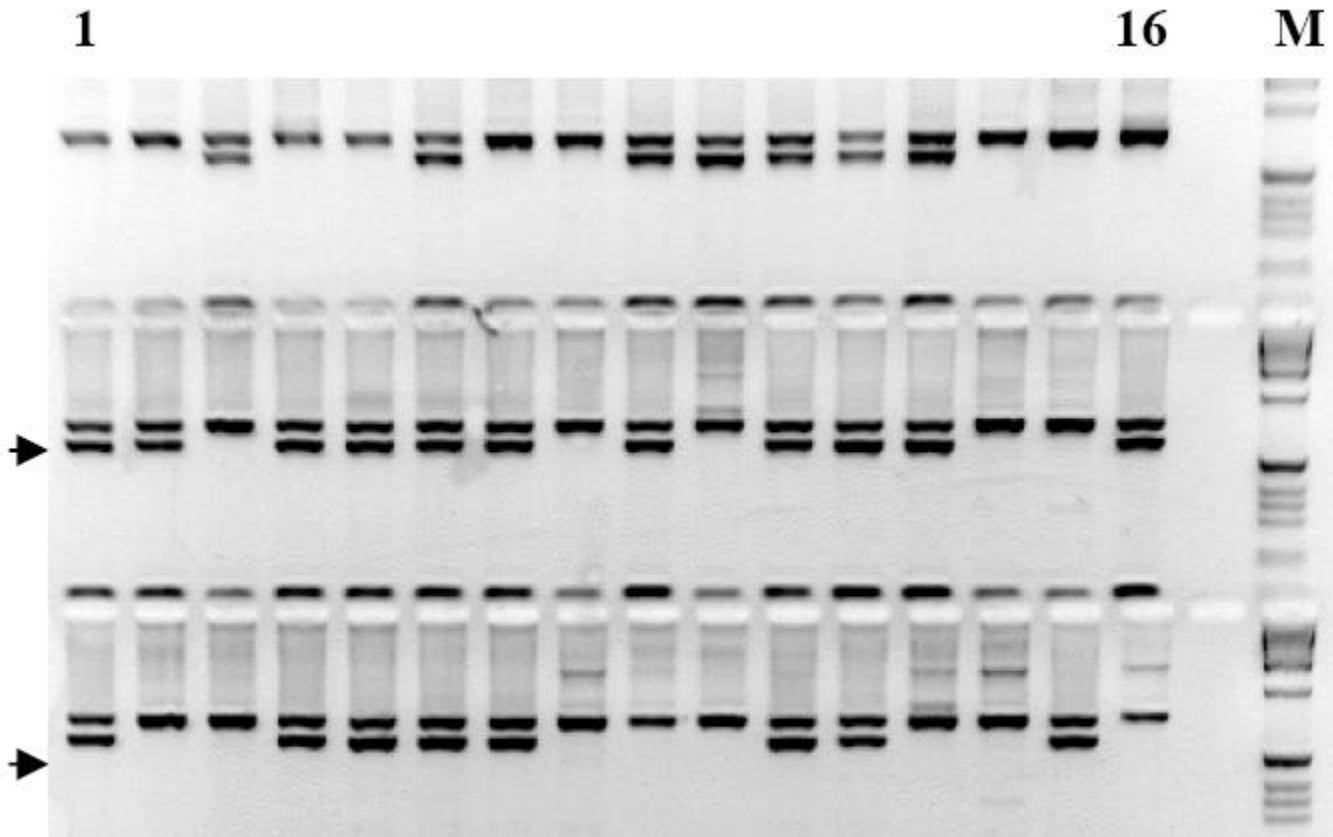
---

Zrychlení a zefektivnění šlechtění vyžaduje selekci nejvhodnějších rodičovských kombinací a efektivní a přesnou selekci potomstva v ranních stádiích vývoje.

Markrem zprostředkovaná selekce „Marker assisted selection (MAS)“ s molekulárními markery v těsné vazbě na sledované znaky a haplotypy = dramatické snížení počtu křížení a linií v potomstvu.

Vyžaduje dostatek markerů se silnou vazbou na sledované znaky a vhodný, ideálně „High-throughput“ genotypovací systém za přijatelné ceny.

# MAS



**Použití PCR markeru SR4R04 pro selekci sterilních rostlin u žita.  
Fertilní rostliny mají dva proužky a sterilní jeden.**

# DNA polymorfismus = Marker

---

## DNA polymorfismus:

- Různé mechanismy mohou vést ke vzniku, více či méně komplexních variant DNA sekvence. Přítomnost jedné nebo vícero takovýchto variant je popisována jako polymorfismus.
- Polymorfismus je většinou detekován jako odlišné profily proužků, změna sekvence, barva, síla, pozice nebo přítomnost/nepřítomnost signálu.
- DNA polymorfismus je přítomen bez ohledu na vnější podmínky nebo vývojové stádium organismu.

## Mechanismy produkující DNA polymorfizmy:

- 1) Bodové mutace
- 2) Indels
- 3) Přestavby

# Markery - použití

---

1. Identita genotypu
2. Hodnocení genových zdrojů
3. Studium diverzity
4. Mapování genomů
5. Fylogenetické a evoluční studie

# Detekce DNA polymorfizmů a znalost sekvence genomů

---

## Nevyžadují znalost sekvence genomu

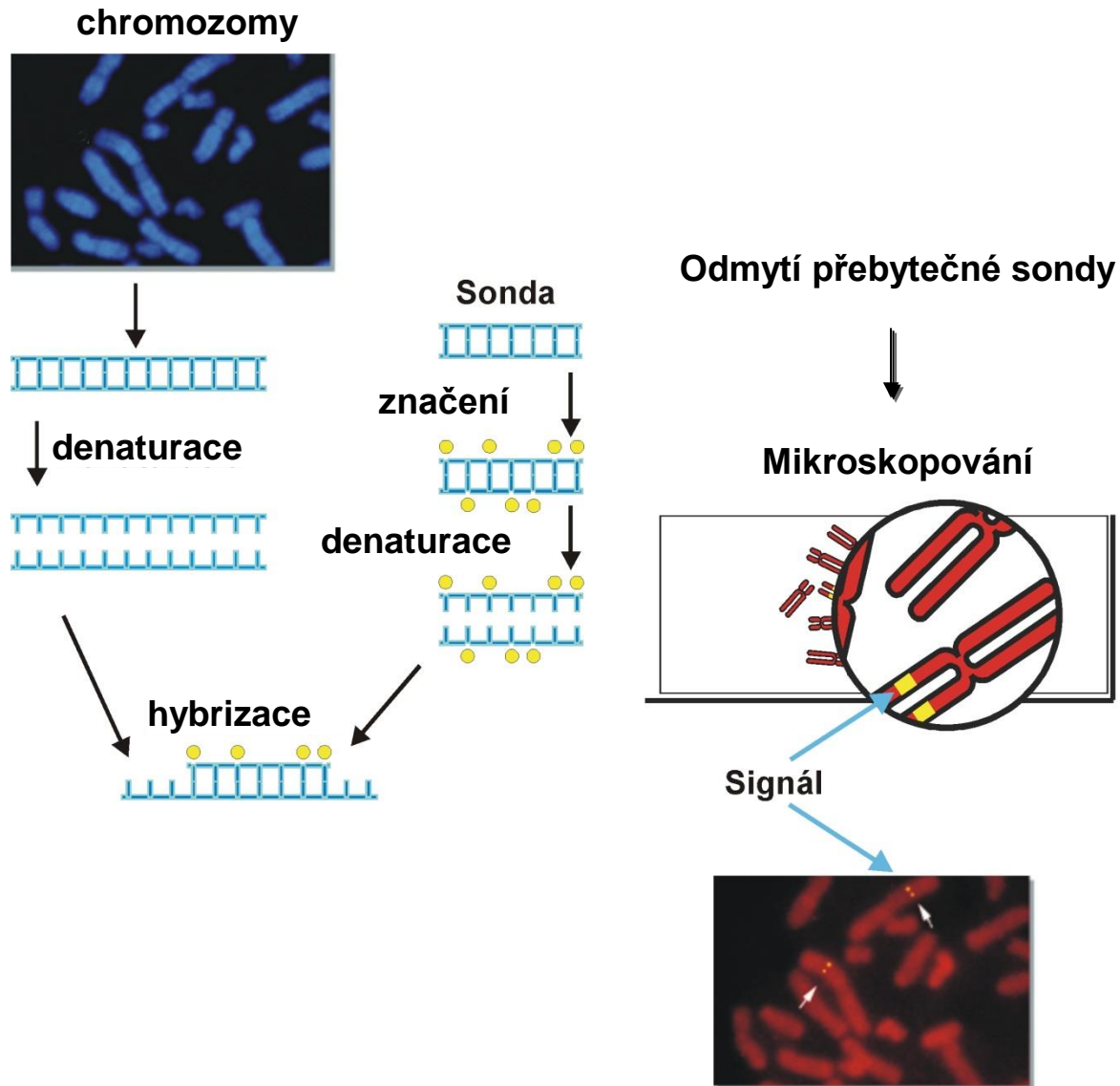
- ❖ Cytogenetické (*In situ*)
- ❖ Random Amplification of Polymorphic DNA (RAPD)
- ❖ Amplified fragment length polymorphisms (AFLP)
- ❖ **DArT / DArTSeq**

## Vyžadují znalost sekvence genomu

- ❖ Microsatelity (SSRs, STMS or SSRPs) a Inter-simple sequence repeats (ISSRs)
- ❖ Sequence-tagged sites (STS)
- ❖ Cleaved amplified polymorphic sequence (CAPS)
- ❖ **Single nucleotide polymorphism (SNP)**
- ❖ **ISBP / RJM (Polymorfizmus na bázi inzerčního místa transposonu)**
- ❖ **Genotypování sekvenováním (GBS)**

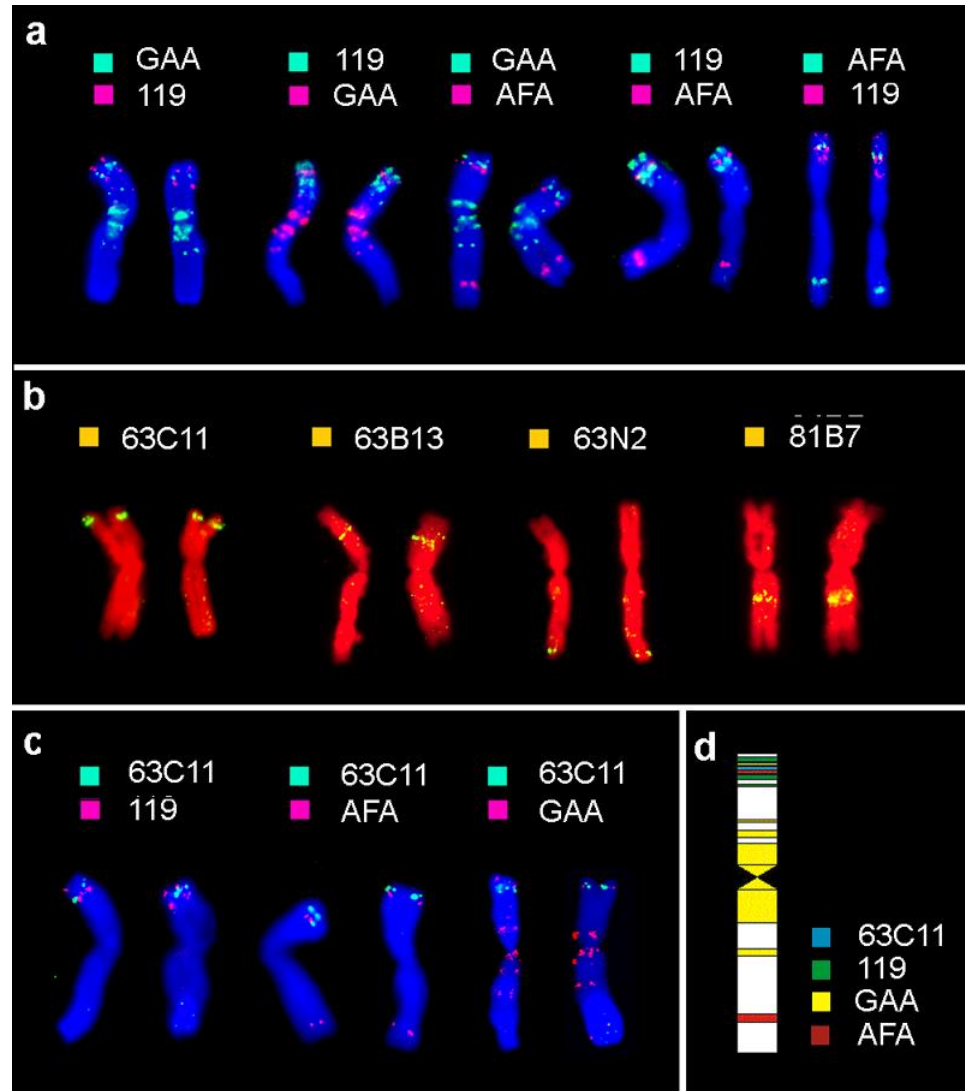
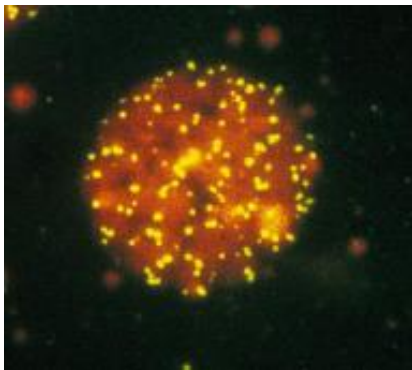
**High-throughput**

# Přestavby, karyotyp: *In situ* na chromosomech





# In situ na jádrech a chromosomech



# DArTseq (diversity arrays technology)

---

Adaptace DArT přístupu na NGS technologie. Paralela s Genotyping-by-Sequencing (GBS).

1) *Redukce genomu* = vytvoření genomových „reprezentací“

- štěpení dvěma restrikčními endonukleázami – 1. často štěpící (*Bst*NI)  
2. vzácně štěpící (*Pst*I)
- ligace sequenčních adaptéru jen k *Pst*I místům
- Illumina sekvenace konců – identifikace SNP/indel polymorfizmů = **DArTseq markery**
- výrazně vyšší počet markerů (**tisíce**)

Také Genotyping-by-sequencing (GBS), konkurenční skupina.

# Microsatelity

---

1. Microsatelity jsou tandemové repetice (1-10 bp) často a relativně rovnoměrně se vyskytující v genomu.
2. Aby mohli být použity jako markery musí být jejich přítomnost v genomu identifikována
3. Polymorfismus v repetitivním regionu může být detekována pomocí PCR z primerů v hraničních oblastech. **Je možná jistá míra automatizace.**

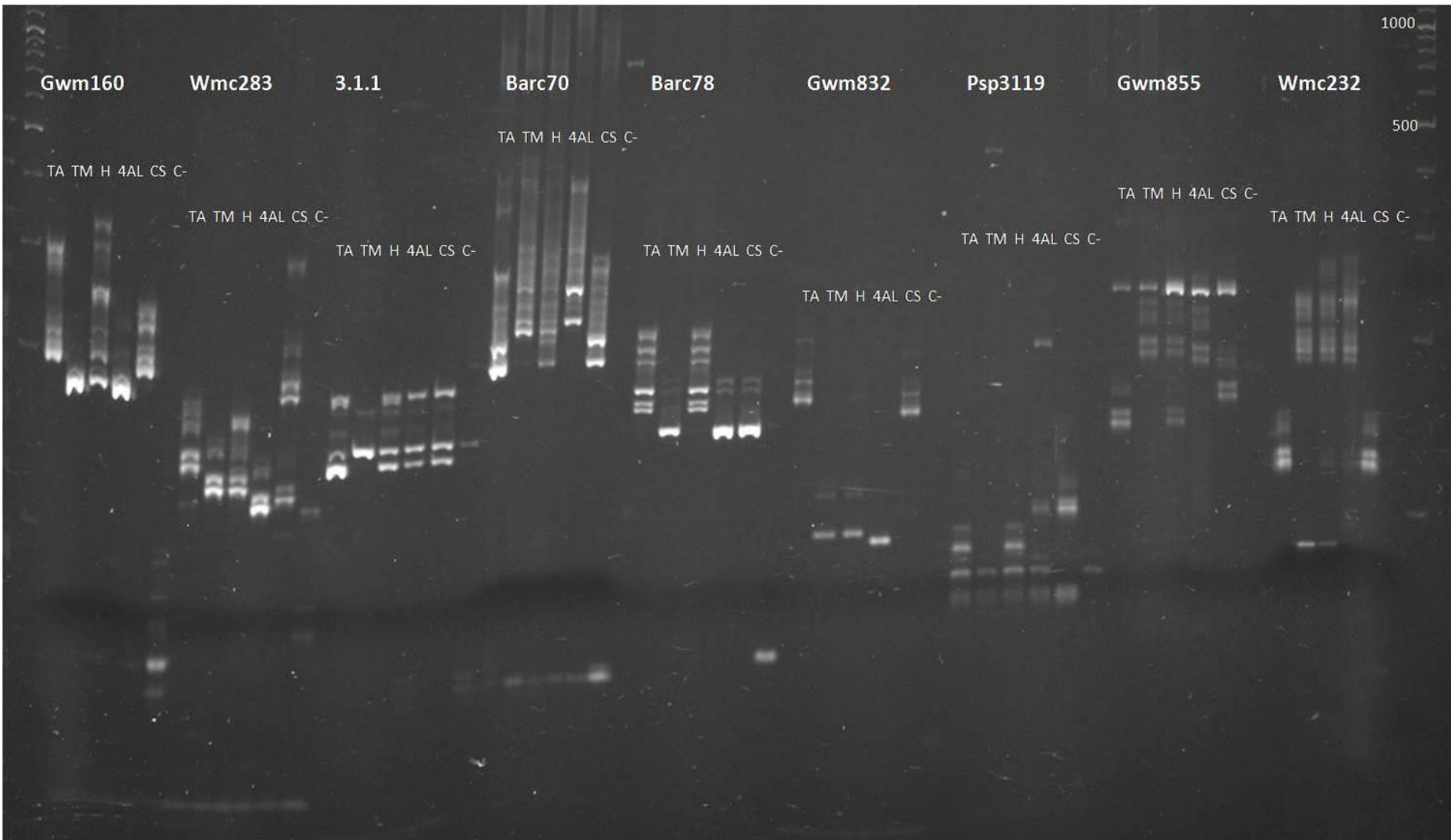
gtgggattgatggatgaggatgcttggagagagagagagagagagagagagagagagattcacacttgggtgtacagttcaattcagg

P1

P2



# Microsatelity / ISSR



# Cleaved amplified polymorphic sequence (CAPS)

MAG2931\_F3/R4\_digestion\_Ddel  
 MAG974\_digestion\_HaeIII\_28\_3\_2013

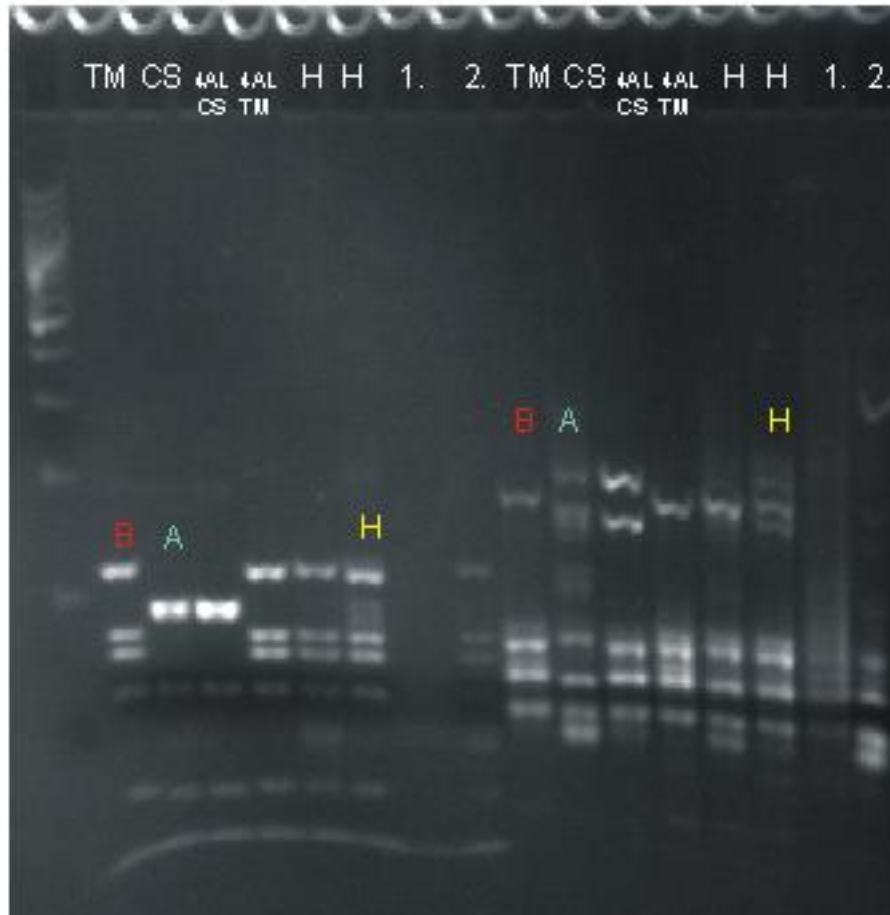
1 L

348

MAG974 4ALTM

MAG2931

MAG974



----- 21 F  
CC (N) 12 ^NNNN  
 ----- 81 \* B  
 ----- 81 \* B  
 ----- 81 \* B  
 ----- 84 F  
 ----- 84 C  
 ----- 84 H 500  
 ----- 98 \* F  
 ----- 98 \* M 200  
 ----- 98 \* H  
 ----- 99 \* B 100  
 ----- 106 \* F  
 ----- 110 N  
 ----- 149 B  
 ----- 173 \*# B

CNNR (N) 9 ^NNNN

NN^ (N) 9AC (N) 5CTCC (N) 7 NNN^

CCTC (N) 6 N^

C^CG C

G CG^C

R GCGC^Y

GGCGGA (N) 9 NN^

C^CNNG G

^CCWGG

-^CCNGG

--^CCNGG

--CC (N) 12 ^NNNN

--CCDG (N) 10 ^NNNN

--GCN |NGC

# SNP markery - Systém KASPAR

V lidském genomu (3Gb) je možné najít až 10 000 000 SNP

- založen na kompetitivní alelově specifické PCR – nasednutí alelově specifického primeru na stoprocentně komplementární sekvenci

genomická DNA    ..... [T/C] .....

**Vhodné pro velký počet vzorků a malý počet markerů**

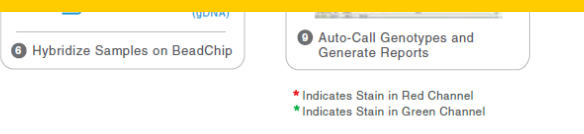
Univerzální FRET kazety  
- komplementární k univerzální sekvenci 1 a 2

nebo 1536-jamkovém formátu.  
Díky univerzálním FRET kazetám výrazně levnější než TaqMan.

# „High-throughput“ SNP genotypování - Illumina

- ❑ Vhodné pro asociační mapování = určení diverzity a selekce vhodných linií pro křížení

**Vhodné pro velký počet vzorků a velký počet markerů**





# Genotypování sekvenováním (GBS)

---

V lidském genomu (3Gb) je možné najít až 10 000 000 SNP

- sekvenováním je možné získat informace o genotypu v tisícičkách markerových lokusů najednou (SNP, *presence/absence variation=PAV*, *copy number variation=CNV*)

**Vhodné pro velký počet vzorků a malý počet markerů**

- **RAD** (*restriction-site associated DNA* - krátké sekvence obklopujících rekogniční místa restričních endonukleáz)
- k tvorbě genomové reprezentace mohou být využity i metylačně citlivé RE obohacení o geny: **DArTseq**,

# Sekvenační technologie nové generace

---

(= masivně paralelní sekvenování)

	Run Time	Read Length	Quality	Total nucleotides sequenced	Cost /MB
454 Pyrosequencing	24h	700 bp	Q20-Q30	1 GB	\$10
Illumina Miseq	27h	2x300bp	> Q30	15 GB	\$0.15
Illumina Hiseq 2500	1 - 10days	2x250bp	>Q30	3000 GB	\$0.05
Ion torrent	2h	400bp	>Q20	50MB-1GB	\$1
Pacific Biosciences	30m - 4h	10kb - >40kb	>Q50 consensus >Q10 single	500 - 1000MB /SMRT cell	\$0.13 - \$0.60

Další technologie – **Sangerovo sekvenování** (ABI) – 1700 \$ /1Mb, délka – do 1000 bp  
**Oxford nanopores** – Dlouhé čtení, problémy s kapacitou

# Praktická ukázka Genotypování Banánovníku pomocí SSR markerů



3730xl DNA analyzer

## • Pomocí PCR v nezávislých PCR reakcích

Fluorophore	Marker	Motif	Primers nucleotide sequence		Annealing temperature	Reference	GenBank accession
			FWD (5'-3')	REV (5'-3')			
6-FAM	mMaCIR01	(GA) <sub>20</sub>	TTAAAGGTGGGTTAGCATTAGG	TTTGATGTCACAATGGTGTTC	55°C	Lagoda et al. 1998	X87262
6-FAM	mMaCIR03	(GA) <sub>10</sub>	TGACCCACGAGAAAAGAAGC	CTCCTCCATAGCCTGACTGC	55°C	Lagoda et al. 1998	X87263
NED	mMaCIR07	(GA) <sub>13</sub>	AACAACCTAGGATGGTAATGTGTGGAA	GATCTGAGGATGGTCTGTTGGAGTG	53°C	Lagoda et al. 1998	X87258
VIC	mMaCIR08	(TC) <sub>6</sub> N <sub>24</sub> (TC) <sub>7</sub>	ACTTATCCCCCGCACTCAA	ACTCTCGCCCATCTTCATCC	55°C	Lagoda et al. 1998	X87264
PET	mMaCIR13	(GA) <sub>16</sub> N <sub>76</sub> (GA) <sub>8</sub>	TCCCAACCCTGCAACCACT	ATGACCTGTGGAACATCCTTT	53°C	Lagoda et al. 1998	X90745
PET	mMaCIR24	(TC) <sub>7</sub>	ATCTTTCTTATCCTTCTAACG	ATTAGATCACCGAAGAACTC	48°C	Lagoda et al. 1998	Z85972
VIC	mMaCIR39	(CA) <sub>5</sub> GATA(GA) <sub>5</sub>	AACACCGTACAGGGAGTCAC	GATACATAAGGCAGTCACATTG	52°C	Lagoda et al. 1998	Z85970
6-FAM	mMaCIR40	(GA) <sub>13</sub>	GGCAGCAACAACATACTACGAC	CATCTTACCCCCATTCTTTTA	54°C	Lagoda et al. 1998	Z85977
6-FAM	mMaCIR45	(TA) <sub>4</sub> CA(CTCGA) <sub>4</sub>	TGCTGCCTTCATCGCTACTA	ACCGCACCTCCACCTCCTG	57°C	Lagoda et al. 1998	Z85968
VIC	mMaCIR150	(CA) <sub>10</sub>	ATGCTGTCAITGCCTTGT	GAATGCTGATACCTCTTTGG	54°C	Hippolyte et al. 2010	AM950440
6-FAM	mMaCIR152	(CTT) <sub>18</sub> , (CT) <sub>17</sub> , (CA) <sub>6</sub>	CCACCTTTGAGTTCTCTCC	TTTCCCTCTTCGATTCTGT	54°C	Hippolyte et al. 2010	AM950442
VIC	mMaCIR164	(AC) <sub>14</sub>	AAGACAAGTCCATTGCTTG	GTTCCGGGCTTTCGGT	55°C	Hippolyte et al. 2010	AM950454
NED	mMaCIR196	(TA) <sub>4</sub> , (TC) <sub>17</sub> , (TC) <sub>3</sub>	GCTCCAAACCTCCCTTT	CGATGCCACACTGGAC	55°C	Hippolyte et al. 2010	AM950462
NED	mMaCIR214	(AC) <sub>7</sub>	CCATTGAGAGATCAACCC	CTATTTGACGTTGGTGGTC	53°C	Hippolyte et al. 2010	AM950480
NED	mMaCIR231	(TC) <sub>10</sub>	GCAATAGTCAAGGGAATCA	ACCCAGGTCTATCAGGTCA	55°C	Hippolyte et al. 2010	AM950497
PET	mMaCIR260	(TG) <sub>8</sub>	GATGTTTGGGCTGTTTCTT	AAGCAGGTGAGATTGTTCC	55°C	Hippolyte et al. 2010	AM950515
6-FAM	mMaCIR264	(CT) <sub>17</sub>	AGGAGTGGGAGCCTATTT	CTCCTCGGTCAGTCCTC	53°C	Hippolyte et al. 2010	AM950519
NED	mMaCIR307	(CA) <sub>6</sub>	AGACTTGATCGCTTGATAA	ACGCTGCACCAGTCAA	54°C	Hippolyte et al. 2010	AM950533
PET	Ma-3-90	(CT) <sub>11</sub>	GCACGAAGAGGCATCAC	GGCCAAATTTGATGGACT	53°C	Crouch et al. 1998	n/a

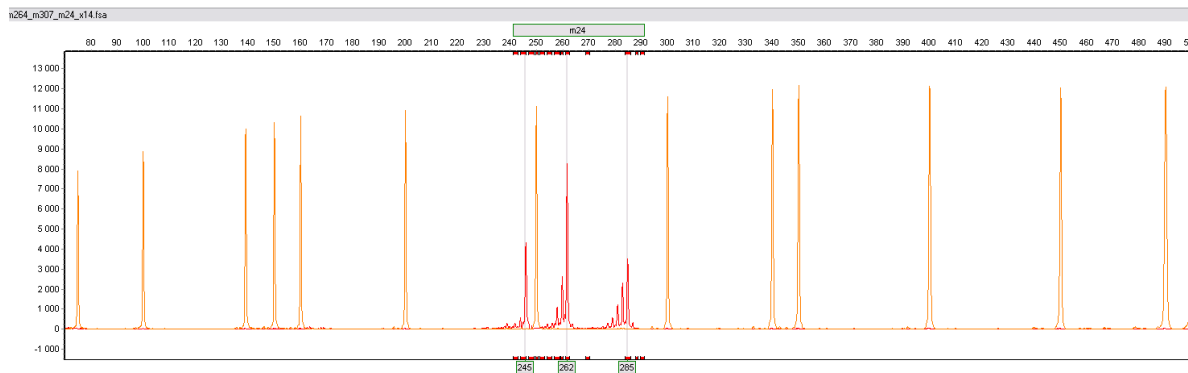
# Praktická ukázka Genotypování Banánovníku pomocí SSR markerů

## • Analýza SSR profilů programem GeneMarker



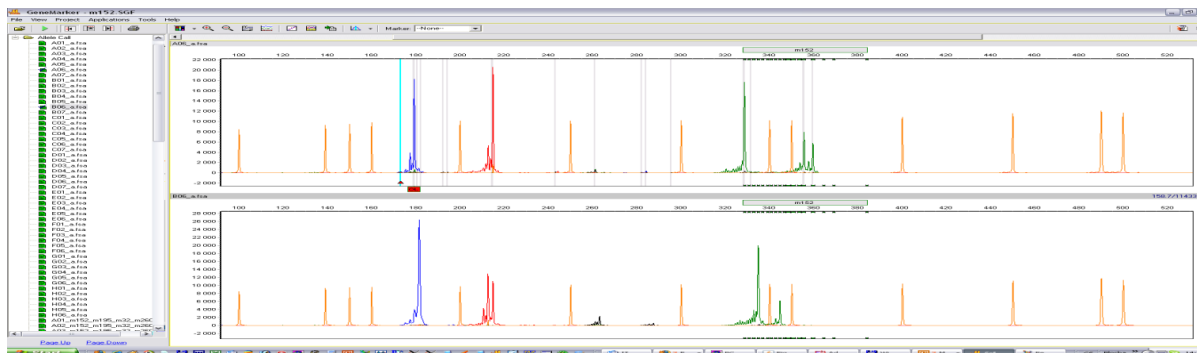
3730xl DNA analyzer

Přesná identifikace velikosti alel s vysokým rozlišením



SSR profil  
genotypu Sport of  
Silk (AAB)  
pro marker  
mMaCIR24  
(červené píky)

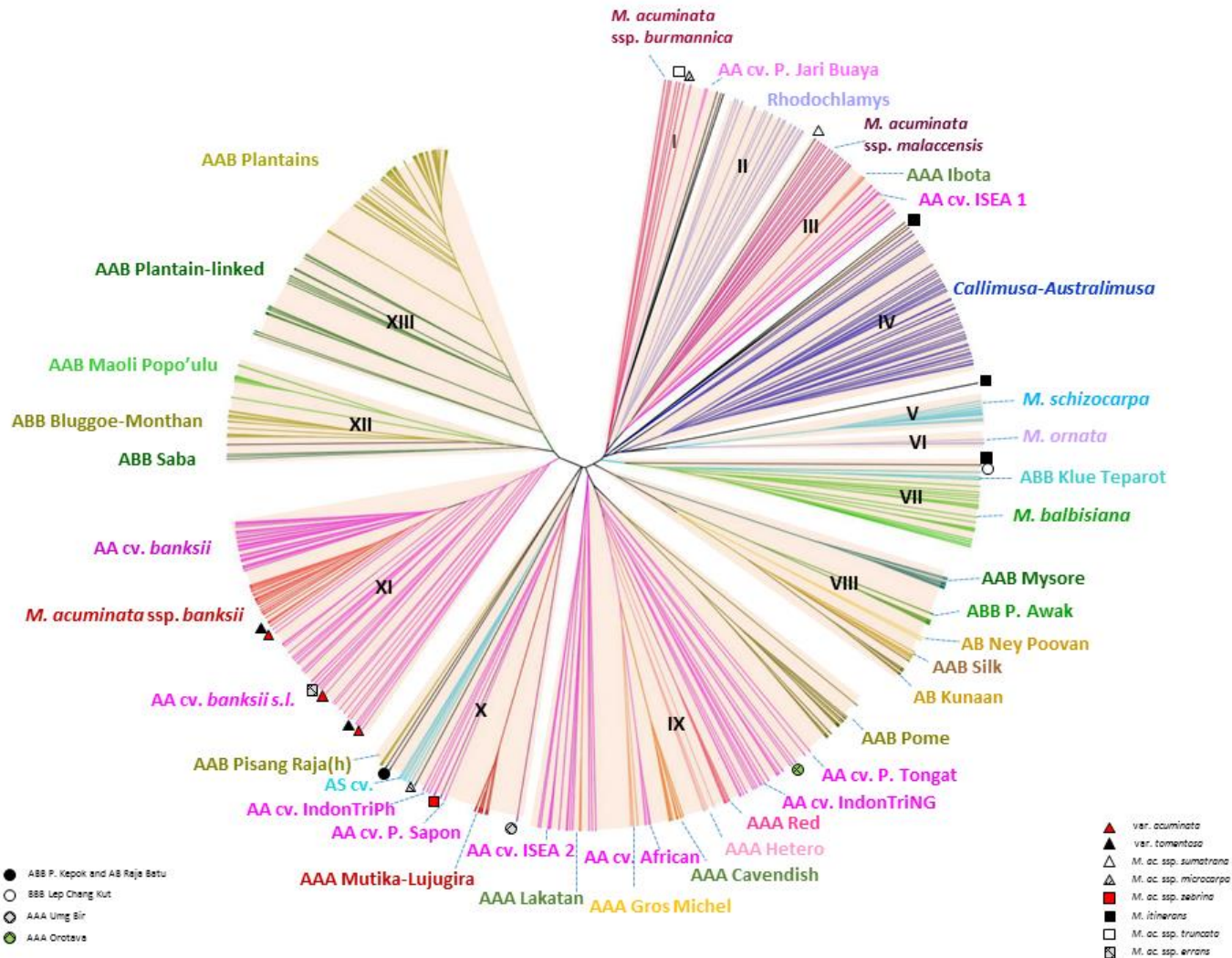
High-throughput, multiplexing



Multiplexing SSR profilů  
genotypu Mbwarzirume  
(AAA)  
pro SSR markery  
mMaCIR152 (zelený)  
mMaCIR195 (modrý)  
mMaCIR260 (červený)

# Praktická ukázka Genotypování Banánovníku pomocí SSR markerů

Binární matice v současné době obsahuje: **933** položek (2x,3x,4x)



# Shrnutí

---

Pokud potřebujete výsledky z malého počtu markerů malého počtu rostlin:

Udělejte si to sami nebo v naší Aplikační laboratoři - ideálně: SSR STS, CAPS, SNP

Pokud potřebujete výsledky z malého počtu markerů velkého počtu rostlin:

Využijte naší Aplikační laboratoř - ideálně: SNP použitím KASPAR nebo GBS systému

Pokud potřebujete výsledky z velkého počtu markerů velkého počtu rostlin:

Poradte se s námi a pomůžeme vám vybrat optimální ganotypovací platformu.



**Děkuji za pozornost**