

## STUDY OF GENOME SIZE EVOLUTION

Čegan R.<sup>1, 2</sup>, Obšivačová V.<sup>1, 2</sup>, Kubeková H.<sup>2</sup>, Kejnovský E.<sup>2</sup>, Šafář J.<sup>3</sup>, Vyskot B.<sup>2</sup>, Hobza R.<sup>1, 2</sup>

<sup>1</sup>Department of Plant Biology, Faculty of Agronomy, Mendel University of Agriculture and Forestry in Brno, Zemedelska 1, 613 00 Brno, Czech Republic

<sup>2</sup>Department of Plant Developmental Genetics, Institute of Biophysics, Academy of Sciences of the Czech Republic, Kralovopolska 135, 612 65 Brno, Czech Republic

<sup>3</sup>Department of Molecular Cytology and Cytometry, Institute of Biophysics, Academy of Sciences of the Czech Republic, Kralovopolska 135, 612 65 Brno, Czech Republic

E-mail: cegan@ibp.cz, hobza@ibp.cz

---

### ABSTRACT

The aim of this study is detection of abundance of repetitive elements and their proportion in size of the *Silene vulgaris* genome. Furthermore we intend to compare representation and localization of individual repetitive elements in *S. vulgaris* and *S. latifolia*. Observed data should help to answer why is variability between *S. latifolia* and *S. vulgaris* genome so vigorous. Both species are diploid and have the same chromosome number ( $2n=24$ ), nevertheless *S. latifolia* has 2.5x larger genome than *S. vulgaris*. Preliminary data showed that some repetitive sequences (retrotransposons and tandem repeats) are unique for first or second species, respectively. On the basis of acquired data we decipher how quick is evolution of particular sequences regarding their structure and relative and absolute proportion in the genome. Our data can help shed a light on the question, which concrete DNA sequences causes quick genome expansion of many species from genus *Silene*. Particular representation of repetitive sequences in individual sexes of dioecious species can further help us to understand their role in reproductive strategies evolution and in sex chromosomes formation.

**Key words:** *S. vulgaris*, transposable elements, genome size evolution, speciation, sex chromosomes

**Acknowledgments:** This research is supported by IGA MUAF (DP 7/2009)

## ÚVOD

Velikosti genomů vyšších rostlin jsou značně variabilní. Mezi druhy s nejmenšími (*Arabidopsis*) a největšími (*Fritillaria*) genomy je rozdíl více než tři řádů. Mezi základní procesy, které formují velikost genomu u rostlin patří polyploidie a zastoupení jednotlivých repetitivních sekvencí v genomu. U většiny rostlinných rodů se oba tyto mechanismy evoluce genomů prolínají a spolupůsobí. Rod *Silene* je unikátní tím, že téměř všichni jeho zástupci jsou diploidní a mají stejný počet chromozomů ( $2n=24$ ). Dva nejvýznamnější druhy z hlediska využití v základním a aplikovaném výzkumu jsou *S. vulgaris* a *S. latifolia*. *S. vulgaris* je gynodioecický druh, jež je převážně využíván jako modelový organizmus pro studium rezistence rostlin k těžkým kovům (van Hoof *et al.*, 2001) a pro studium evoluce cytoplasmatické samčí sterility (Taylor *et al.*, 2001). *S. latifolia* je jeden z nejkoumanějších rostlinných druhů z hlediska vzniku dvoudomosti a evoluce pohlavních chromozomů (Vyskot a Hobza, 2004). Ačkoli oba dva druhy jsou blízce příbuzné, liší se velikost jejich genomů přibližně 2,5krát. V nedávné době bylo publikováno mnoho dat, která se týkají evoluce repetitivních sekvencí u *S. latifolia*. Bylo ukázáno, že některé retrotranspozony jsou schopny přenášet a rozšiřovat tandemové repeticce v genomu (Kejnovsky *et al.*, 2006a). Dále bylo prokázáno, že některé sekvence centromerického původu jsou značně akumulovány na nerekombinující části chromozomu Y (Hobza *et al.*, 2007). Některé invertované repetitivní sekvence u *S. latifolia* se v genomu značně rozšířily, zvláště pak v centromerách a na chromozomu Y (Hobza *et al.*, 2006). Podrobná analýza organelové DNA ve frakci jaderného genomu ukázala masivní transfer chloroplastových genů do jádra během nedávné evoluce druhu (Kejnovsky *et al.*, 2006b). V nedávné době byla provedena podrobná studie sledující početní zastoupení a distribuci nejpočetnějších DNA sekvencí v genomu (Cermak *et al.*, 2008). Narozdíl od *S. latifolia*, u *S. vulgaris* nejsou k dispozici žádná data, týkající se charakterizace zastoupení jednotlivých repetitivních sekvencí v genomu.

## MATERIÁL A METODIKA

### *Rostlinný materiál a izolace DNA*

Rostliny *Silene vulgaris* byly pěstovány v kultivační místnosti za standardních podmínek (24 °C, 16h světlo/8h tma). Genomická DNA byla izolována z mladých listů semenáčků pomocí DNAeasy Plant Mini Kitu (Qiagen).

### *Konstrukce a analýza („screening“) knihovny krátkých inzertů*

Vyizolovaná DNA byla sonikací (7s) nalámána na požadovanou délku 700-1200 párů bází. Nalámané konce byly následně ošetřeny T4 DNA polymerázou a fosforylovány pomocí T4 polynukleotid kinázy. Získané úseky byly přečištěny gelovou elektroforézou. Přečištěné fosforylované krátké inzerty byly ligovány pomocí kitu Smart Cloning Kit (Lucigen) do plazmidového vektoru pSMART<sup>®</sup> LCamp a následně transformovány do kompetentních buněk *E. coli* 10G. Takto získané transformované bakterie u nichž byla pomocí PCR ověřena přítomnost inzertů byly použity na konstrukci knihovny krátkých inzertů. Tato knihovna byla vytvořena v Laboratoři molekulární cytogenetiky a cytometrie Ústavu experimentální botaniky AV ČR, v.v.i. v Olomouci využitím robota firmy Genetix. Knihovna celkově obsahuje 7720 klonů a je uložena ve 20 mikrodestičkách.

Pro hybridizační analýzy je „natištěna“ na 3 membrány.

Pro analýzu knihovny krátkých inzertů *S. vulgaris* byla použita sonda genomické DNA *S. vulgaris*. Značení sond bylo provedeno pomocí Prime-It II Random Labeling Kitu (Stratagene) a jako izotop byl použit  $\alpha$ - $^{32}\text{P}$  dATP, podle standardního protokolu. Membrány byly hybridizovány při 60°C po dobu 16 hodin a odmyvány v 0,3x SSC/0,1 % SDS 20 minut a 0,1x SSC/0,1 % SDS 20 minut. Signály byly detekovány autoradiografií a vyhodnoceny podle manuálu dodavatele BACové knihovny.

#### *PCR protokol*

Positivně hybridizující klony byly z knihovny vypíchnuty a byla na nich provedena PCR. Standardní PCR podmínky byly 95 °C po dobu 3 min, následovány 25 cykly 94 °C/30s, 60 °C/30 sec min, 72 °C/1 min a finální inkubace 72 °C/5 min. Pro amplifikaci byly použity primery SL1: 5'–CAG TCC AGT TAC GCT GGA GTC–3' a SR2: 5'–GGT CAG GTA TGA TTT AAA TGG TCA GT–3'.

#### *Sekvenování a počítačová analýza získaných dat*

Amplifikované PCR produkty byly přečištěny pomocí ExoSAP, naznačeny pomocí BigDye® Terminator Cycle Sequencing Kitu dle manuálu výrobce a následně purifikovány pomocí Agencourt® CleanSEQ® kitu. Purifikované a naznačené vzorky byly sekvenovány v Laboratoři molekulární cytogenetiky a cytometrie Ústavu experimentální botaniky AV ČR, v.v.i. v Olomouci, pomocí 96 kapilárního sekvenátoru ABI 3730xl dle manuálu výrobce.

Získané sekvence byly počítačově analyzovány. Ke složení sekvencí byl použit program Geneious, DNA baser, MAFFT a CLUSTAL W. Pro zjištění sekvencí homologií mezi jednotlivými klony byl využit program JDotter, pro homologie s již známými sekvencemi BLAST (Basic Local Alignment Search Tool) a homologie k repetitivním sekvencím (Repbase) byly zjištěny pomocí CENSORu (compares and mask nucleotide repeats). Tandemové repetice byly identifikovány pomocí Tandem Repeats Finder 4.0. Otevřené čtecí rámce byly identifikovány pomocí ORF Finder, FGENESH, GeneScan a konzervativní proteinové domény pomocí CD SEARCH (NCBI) nebo RPSBLAST.

#### *FISH (Fluorescence in situ hybridization)*

Mitotické chromozomy *S. vulgaris* a *S. latifolia* byly připraveny z kořenových špiček ošetřených podle Lengerové *et al.* (2004) s drobnými úpravami. Denaturace prób byla 10 minut při 75 °C. Do hybridizační reakce bylo vzato 100-400 ng denaturované próby a hybridizace probíhala 16 hodin při 37 °C ve vlhké komůrce. Jako sondy byly použity Cy3 značené oligonukleotidy (SV STAR left: CGAACGATAAGGAGTGACTA a SV STAR right: CACACAACGAGTCCAAAAGACTAA). Chromozomy byly barveny DAPI (4', 6'-diamidino-2-fenylindole). Pozorování mitotických preparátů bylo provedeno fluorescenčním mikroskopem Olympus AX70 a snímky byly zpracovány pomocí programu ISIS.

## **VÝSLEDKY A DISKUZE**

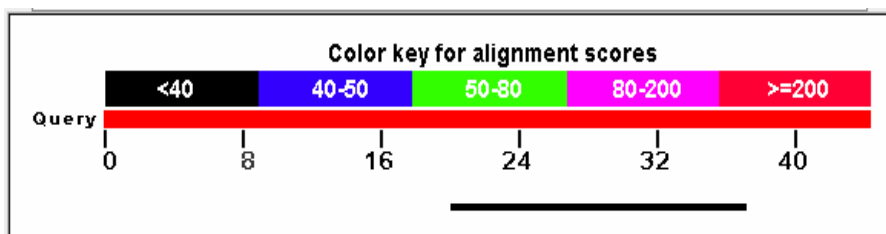
První částí experimentu byla konstrukce knihovny krátkých inzertů *Silene vulgaris* s počtem klonů 7720. Dále byly provedeny radioaktivní hybridizace této knihovny, s genomickou DNA *Silene vulgaris*. Nejsilněji hybridizující klony byly vybrány (282) a v současnosti probíhá jejich sekvenace a počítačová analýza těchto sekvencí. Z dosud sekvenovaných 130 klonů 56 vykazuje homologii k sekvenci

gb|EU646284.1| *Silene latifolia* clone L141\_6C14 satellite STAR sequence a 65 těchto sekvencí vykazuje homologii s gb|EU646384.1| *Silene latifolia* clone L141\_8J2 putative Athila-like retrotransposon, complete sequence, and satellite STAR sequence.

*STAR (Silene tandem repeat)*

Ze získaných sekvencí SV STAR, byla vytvořen contig dlouhý 1170 bp a v něm byla pomocí programu Tandem Repeat Finder nalezena tandemová repetice dlouhá 43 bp s počtem opakování 16 (CGAACGATAAGGAGTGACTACACACAACGAGTCACAAAGACTAA). Porovnáním této tandemové repetice se STAR C *Silene latifolia* jsme zjistili, že je homologní (obr.1) pouze v jedné části (right) a druhá část (left) je pro *S. vulgaris* unikátní. Na tyto společné a unikátní sekvence byly navrženy Cy3 značené primery sloužící jako sondy pro FISH.

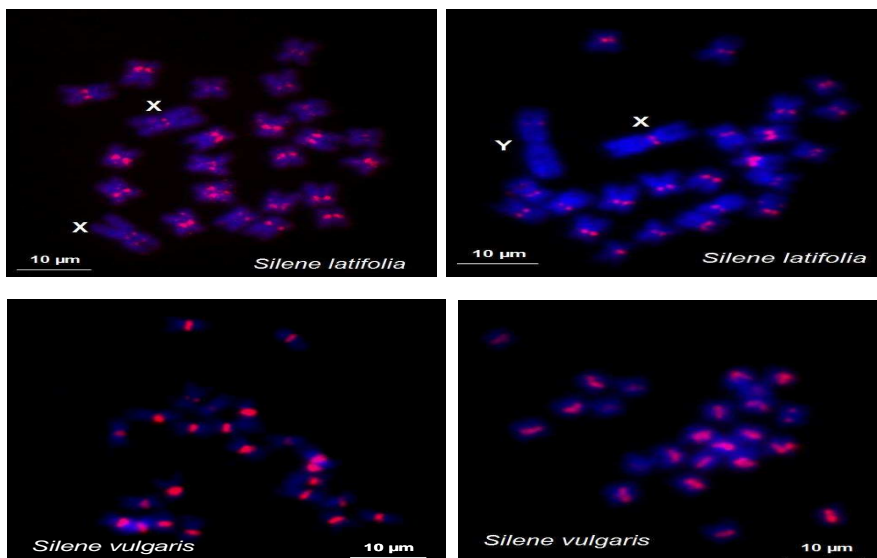
Obr.1 Alignment STAR tandemové repetice *S. vulgaris* a *S. latifolia*



>*S.vulgaris* STAR (44bp) CGAACGATAAGGAGTGACTACACACAACGAGTCACAAAGACTAA

>*S. latifolia* STAR C (43bp) TAATGAACAAGGACAAATGATTACACACAACGAGTCACAAATG

Obr.2 FISH se sondami STAR left a right na mitotických chromozomech *S.vulgaris* a *S. latifolia*



## Další experimenty

Aby jsme získali další informace potřebné k zjištění zastoupení repetitivních elementů v genomu *S. vulgaris*, k porovnání s daty získanými u *S. latifolia* a k zodpovězení původních otázek, budeme dále hybridizovat knihovnu se sondami obsahující konzervativní části jednotlivých repetitivních elementů. K přípravě sond pro analýzu knihovny krátkých inzertů budou použity tyto degenerované primery pro části repetitivních elementů: pro konzervativní části LINE endonukleázy (Noma *et al.*, 1999), gypsy-like reverzní transkriptázy (Friesen *et al.*, 2001), copia-like reverzní transkriptázy (Flavell *et al.*, 1992), Alu SINE (Fawcett *et al.*, 2006), CACTA transpozázy (Staginnus *et al.*, 2001), mariner transpozázy (Feschotte and Wessler 2002) a mutator transpozázy (Lisch *et al.*, 2001). konzervativní části LINE endonukleázy (Noma *et al.*, 1999), gypsy-like reverzní transkriptázy (Friesen *et al.*, 2001), copia-like reverzní transkriptázy (Flavell *et al.*, 1992), Alu SINE (Fawcett *et al.*, 2006), CACTA transpozázy (Staginnus *et al.*, 2001), mariner transpozázy (Feschotte and Wessler 2002) a mutator transpozázy (Lisch *et al.*, 2001). S PCR produkty jednotlivých elementů bude provedena hybridizace a následně nejsilnější hybridizující klony budou sekvenovány. Nejreprezentativnější klony zastupující jednotlivé elementy budou poté vybrány k fluorescenční *in situ* hybridizaci pro určení paternu a zastoupení jednotlivých elementů v genomu *S. vulgaris*.

## ZÁVĚR

Nejčastější repeticí genomu *S. vulgaris* je tandemově se opakující sekvence centromerického původu STAR, podobně jako v genomu *S. latifolia* jsou nejpočetněji zastoupeny tandemové repeticí X-43.1 (subtelomerická repeticí využívána k odlišení pohlavních chromozomů) a STAR (*Silene* tandem repeat). Přestože je STAR lokalizací v genomu poměrně konzervativní, sekvenční divergence mezi jednotlivými druhy je značná. Genom *S. latifolia* je téměř 2.5x větší než genom *S. vulgaris* díky expanzi různých typů transpozonů.

## LITERATURA

Cermak T, Kubat Z, Hobza R, Koblizkova A, Widmer A, Macas J, Vyskot B & Kejnovsky E (2008) Survey of repetitive sequences in *Silene latifolia* with respect to their distribution on sex chromosomes. *Chromosome Research* 16: 961-976.

Fawcett JA, Kawahara T, Watanabe H, Yasui Y (2006) A SINE family widely distributed in the plant kingdom and its evolutionary history. *Plant Molecular Biology* 61: 505-514.

Feschotte C, Wessler SR (2002) Mariner-like transposases are widespread and diverse in flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* 99: 280-285.

Flavell JA, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992) *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Research* 20: 3639-3644.

Friesen N, Brandes A, Heslop-Harrison JS (2001) Diversity, origin, and distribution of retrotransposons (gypsy and copia) in conifers. *Molecular Biology and Evolution* 18: 1176-1188.

- Hobza R, Lengerova M, Svoboda J, Kubekova H, Kejnovsky E & Vyskot B (2006) An accumulation of tandem DNA repeats on the Y chromosome in *Silene latifolia* during early stages of sex chromosome evolution. *Chromosoma* 115: 376-382.
- Hobza R, Kejnovsky E, Vyskot B, Widmer A (2007) The role of chromosomal rearrangements in the evolution of *Silene latifolia* sex chromosomes. *Molecular Genetics and Genomics* 278: 633-638.
- Kejnovsky E, Kubat Z, Macas J, Hobza R, Mracek J, Vyskot B (2006a) Retand: A novel family of gypsy-like retrotransposons harboring an amplified tandem repeat. *Molecular Genetics and Genomics* 276: 254-263
- Kejnovsky E, Kubat Z, Hobza R, Lengerova M, Sato I, Tabata S, Fukui K, Matsunaga S & Vyskot B (2006b) Accumulation of chloroplast DNA sequences on the Y chromosome of *Silene latifolia*. *Genetica* 128: 167-175.
- Lengerová M., Kejnovský E., Hobza R., Macas J., Grant S.R. et. Vyskot B. (2004): Multicolor FISH mapping of the dioecious model plant *Silene latifolia*. *Theor Appl Genet*, 108: 1193-1199.
- Noma K, Ohtsubo E, Ohtsubo H (1999) Non-LTR retrotransposons (LINEs) as ubiquitous components of plant genomes. *Molecular and General Genetics* 261: 71-79.
- Staginnus Ch, Huettel B, Desel Ch, Schmidt T, Kahl G (2001) A PCR-based assay to detect *En/Spm*-like transposon sequences in plants. *Chromosome Research* 9: 591-605.
- Taylor DR, Olson MS, McCauley DE (2001) A quantitative genetic analysis of nuclear-cytoplasmic male sterility in structured populations of *Silene vulgaris*. *Genetics*. 158: 833-841.
- van Hoof NA, Hassinen VH, Hakvoort HW, Ballintijn KF, Schat H, Verkleij JA, Ernst WH, Karenlampi SO, Tervahauta AI (2001) Enhanced copper tolerance in *Silene vulgaris* (Moench) Garcke populations from copper mines is associated with increased transcript levels of a 2b-type metallothionein gene. *Plant Physiology* 126: 1519-1526.
- Vyskot B, Hobza R (2004) Gender in plants: sex chromosomes are emerging from the fog. *Trends in Genetics* 20: 432-438.