

Knihovna náročného čtenáře – čtení v genomu pšenice

Díky pokroku v sekvenačních technologiích byla již přečtena dědičná informace mnoha organismů a může se tedy zdát, že sekvenování se stalo rutinní záležitostí (Živa 2006, 5: 198–200). Ve skutečnosti tomu tak není, protože genomy mnoha rostlinných a živočišných druhů jsou natolik velké a složité, že jejich úplné přečtení (sekvenování) stále představuje jen obtížně překonatelný problém. Do této kategorie se řadí i důležitá zemědělská plodina – pšenice setá (*Triticum aestivum*). Tento příspěvek popisuje potíže spojené se sekvenováním dědičné informace pšenice. Na řešení problematiky, jež má značný praktický význam, se podílí olomoucké pracoviště Ústavu experimentální botaniky AV ČR, v. v. i. – součást Centra regionu Haná pro biotechnologický a zemědělský výzkum.

Pšenici sejeme a sklízíme.

Ale bude to stačit?

Pšenice setá patří k nejdůležitějším světovým plodinám, tvoří základ stravy pro více než třetinu obyvatel planety a jejímu významu ve výživě konkurují pouze rýže a kukuřice. Produkce pšenice se však zvyšuje pomaleji než spotřeba. Studie OSN z r. 2009 odhaduje, že v r. 2050 bude žít na Zemi přibližně 9 miliard lidí. Světové zásoby pšenice se tenčí, za posledních 10 let klesly zhruba na polovinu. Úrodu navíc každoročně ohrožují stále častější a extrémnější výkyvy počasí (horko, sucho, mraz, škůdci a choroby). Tyto vlivy mohou výrazně snížit sklizeň a způsobit nedostatek a zdražování potravin. Proto je žádoucí vytvářet nové odrůdy – odolnější a s vyššími výnosy. Avšak šlechtění pomocí klasických metod již naráží na své limity. Řešení problému vyžaduje nové postupy založené na využití molekulárních technik – pokud bychom měli přesné informace o funkci a poloze jednotlivých genů v genomu, šlechtitel by mohl efektivněji vybírat rostliny s vhodnými kombinacemi

genů a s těmi dále pracovat. Díky molekulárně-genetickým postupům se může proces šlechtění navíc podstatně urychlit, protože nebude vždy třeba čekat, až potomci křížení vyrostou. Na základě analýzy DNA semenáčků bude možné zjistit, zda nesou geny pro námi požadované vlastnosti. U plodin, jejichž genomy již byly přečteny (např. rýže nebo kukuřice), se nové poznatky, zejména objevy nových genů, úspěšně zavádějí do praxe. Z těchto důvodů šlechtitelé podporují a napjatě očekávají výsledky popisující pšeničný genom.

Tvrď oříšek k rozlousknutí

Avšak situace u genomu pšenice není tak jednoduchá, jak bychom si přáli. Jedním z důvodů je samotná velikost genomu této rostliny. Jedna sada chromozomů běžné pšenice (celkem 21 chromozomů) obsahuje 17 miliard bází, tedy přibližně pětkrát více než genom člověka. Pro srovnání, u jiné obilniny – rýže – jde o „směšných“ 370 milionů bází, což znamená, že každý jednotlivý pšeničný chromozom obsahuje více DNA než všechny chromozomy rýže

dohromady. Další skutečností ztěžující sekvenování je vysoký podíl rozptýlených a nekódujících repetitivních sekvencí nukleotidů – u pšenice jejich podíl dosahuje 80 %. Třetí a neméně závažný problém je polyploidní podstata pšeničného genomu. Pšenice setá je alohexaploidní, tzn., že její genom tvoří tři dílčí příbuzné genomy (subgenomy) označované písmeny A, B a D, každý se 7 chromozomy. Moderní pšenice, jak ji známe dnes, totiž vznikla postupným křížením tří druhů trav s příbuznými genomy (donorem genomu A byl druh pšenice *T. urartu*, D mnohoštět *Aegilops tauschii*, u genomu B zřejmě blízký příbuzný druh dnešního *A. speltoides*). Všechny tyto vlastnosti výrazně komplikují, ne-li přímo vylučují, přečtení pšeničného genomu jako celku.

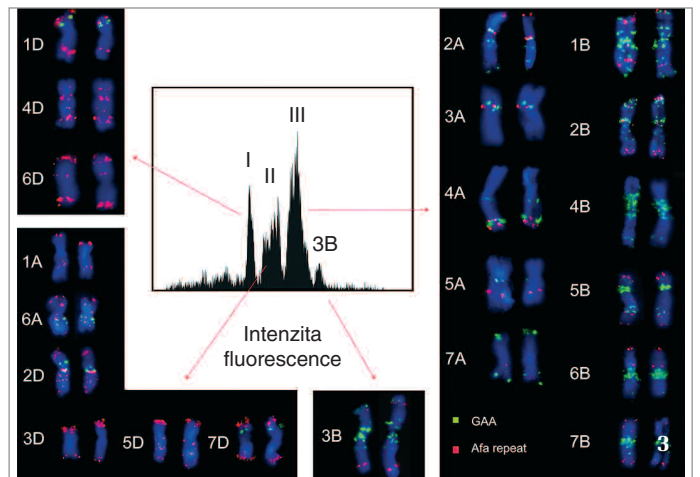
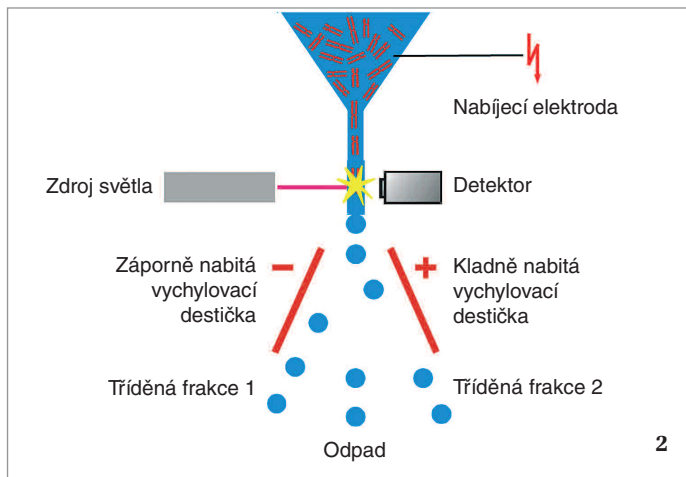
Dvakrát měř a jednou řež. Ale čím?

Během jednoho sekvenování DNA je možno přečíst úsek dlouhý maximálně 500 až 1 000 bází. Rostlinné i živočišné genomy však obsahují zpravidla stovky milionů až miliardy bází. V současné době existují dva přístupy k sekvenování genomů, které řeší tento problém odlišným způsobem. První a modernější se nazývá shotgun sequencing (málo cílený, podobně jako výstřel z brokovnice) a zakládá se na náhodném rozštěpení genomové DNA na malé fragmenty a jejich přímém sekvenování bez nutnosti nakládání s jednotlivými úseky DNA. Tento způsob je rychlejší a levnější, protože celá sekvence genomu se vytvoří v jediné operaci. Moderní přístroje (sekvenátory nové generace) dokáží přečíst až několik miliard písmen (tj. bází) za den, takže genom pšenice by teoreticky mohl být znám za několik dní až týdnů. Problémem není data získat, ale správně je sestavit, abychom z krátkých úseků (zpravidla několik stovek bází) produkovaných sekvenátory získali kompletní pořadí nukleotidů v genomu. Aby se dala sekvence poskládat, je nutné identifikovat přesahy mezi jednotlivými přečtenými úseky, kterých je v případě pšenice několik desítek až stovek milionů. To už je velká výzva pro bioinformatiku zabývající se „skládáním“ DNA. Obtíž totiž nastává v okamžiku, kdy se mají do souvislého „textu“ uspořádat repetitivní sekvence, které se nacházejí na různých místech genomu. Těžko lze též určit, ke kterému ze tří sekvenčně příbuzných subgenomů daná sekvence patří. Z těchto důvodů není uvedený způsob vhodný pro sekvenování celého genomu pšenice najednou.

Druhý přístup, clone by clone sequencing (sekvenování klon po klonu), spočívá ve vytvoření genomových zdrojů – klonů DNA. Jde o části DNA studovaného genomu začleněné do genomu bakterií. Tyto úseky se replikují spolu s hostitelskou DNA a uchovávají se tak v podstatě natrvalo (viz dále). Do genomu každé bakterie

1 Pšenice setá (*Triticum aestivum*) je celosvětově jednou z nejdůležitějších plodin, avšak její produkce se zvyšuje pomaleji než spotřeba. Znalost sekvence jejího velkého a složitého genomu by umožnila rychleji a efektivněji šlechtit nové odrůdy – s vyšší odolností i výnosem. Foto J. Vrána





je začleněn jeden úsek DNA, obvykle o velikosti kolem 100 tisíc bází. Souboru bakteriálních klonů, které pokrývají celý genom zkoumaného organismu, se říká genomová knihovna. S touto se dále pracuje s cílem seřadit jednotlivé úseky DNA tak, jak se nacházely v původním genomu. V této fázi se však nemůžeme opřít o znalost sekvence těchto fragmentů. Postup je založen na reprodukovatelném rozštěpení DNA každého klonu z knihovny několika restrikčními enzymy (štěpí DNA v určité sekvenci bází) na malé části, jejichž délka se určí pomocí kapilární elektroforézy. Klon je tak možno charakterizovat specifickým profilem fragmentů štěpení (tzv. fingerprint – otisk prstu). Na základě podobnosti těchto profilů můžeme použitím speciálního algoritmu zjistit překryvy klonů a virtuálně je spojit do souvislé řady. Sekvenováním klonů takto vytvořených fyzické mapy pak získáme celou sekvenci genomu daného organismu. Tento přístup byl využit u různých organismů, např. háďátka obecného (*Caenorhabditis elegans*), a z velké části i pro genom člověka. K přečtení složitých polyploidních genomů jako celku nebyl dosud použit.

Sekvenování složitého pšeničného genomu není tedy vůbec jednoduché, a je proto nutné hledat způsoby, jak práci usnadnit. Jednou z alternativ, které se u pšenice nabízejí, je studium genomu jejich diploidních předků nebo jiných blízkých příbuzných. Tento postup se uplatnil např. při izolaci některých důležitých genů nebo u srovnávacích a evolučních studií. Moderní hexaploidní pšenice se však od těchto druhů vývojově oddělila před zhruba 0,5–4 miliony let a za tu dobu její genom prošel mnoha změnami. Ačkoli tedy diploidní pšenice představují v mnoha případech neocenitelné pomocníky, jejich význam pro sekvenování pšenice je minimální.

Když to nejde najednou, tak to půjde po částech – příspěvek z Olomouce

Při hledání způsobu, jak zjednodušit analýzu genomu pšenice, se dnes jako nejslibnější jeví metoda, na jejímž rozvoji se výraznou měrou podílel vedoucí olomouckého pracoviště Ústavu experimentální botaniky AV ČR, v. v. i., Jaroslav Doležel. Kouzlo metody spočívá v rozdělení genomu na funkční a strukturální podjednotky – chromozomy. Každý jednotlivý pšeničný chromozom představuje v průměru 5 %

celkového genomu, což značně zjednodušuje analýzu. Další „výhrou“ je odstranění problému s polyploidii – pracujeme s jednotlivými chromozomy, takže víme, z kterého subgenomu konkrétní sekvence pochází. K dělení genomu na chromozomy se využívá metoda průtokové cytometrie (viz Živa 2005, 1: 46–48). Její výhodou je rychlost, jakou se dají analyzovat a izolovat různé mikroskopické částice, zejména buňky nebo jejich části (řádově až desítky tisíc částic za sekundu). Průtokové cytometry našly uplatnění v mnoha vědních disciplínách a také v klinické medicíně. Průtoková cytometrie (obr. 2) je založena na detekci optických parametrů mikroskopických částic (rozptyl světla a fluorescence), které se pohybují v úzkém proudu vodního roztoku a procházejí jedna za druhou intenzivním světelným paprskem (nejčastěji laserovým). Jejich optické vlastnosti se vyhodnocují detektory. Podle intenzity signálu může přístroj částice fyzicky izolovat – třídit. Generuje mikroskopické kapky tekutiny, a pokud se v kapce v okamžiku jejího vzniku vyskytuje částice, kterou chceme získat, cytometr ji nabije elektrickým nábojem. Nabitá částice se při průchodu silným elektrostatickým polem vychýlí do strany, kde ji můžeme jímat a získáme čisté frakce částic, v našem případě chromozomů.

Jak už jsme si ale u pšenice zvykli, nic nedostaneme zadarmo. Nejinak je to s přípravou vzorků pro průtokovou cytometrii. U rostlin není lehké získat suspenzi kompaktních chromozomů. Buňky většiny pletiv se dělí velmi neochotně a pokud ano, tak ne synchronně (jednotlivé buňky nevstupují do dělení současně). Nejlepším materiálem pro přípravu suspenze chromozomů jsou kořenové špičky, protože obsahují hodně dělicích se buněk, jsou karyologicky stabilní (nedochází ke změně v počtu a struktuře chromozomů) a lze je získávat v relativně velkém množství z naklíčených semen. Dělení buněk se synchronizuje pomocí inhibitorů syntézy DNA, např. hydroxymocoviny, která zastaví všechny buňky ve stejné fázi buněčného cyklu. Po odstranění hydroxymocoviny se začíná většina buněk dělit ve stejný okamžik. Chromozomy izolujeme v metafázi, kdy jsou nejvíce kondenzované a kompaktní. Akumulace buněk v této fázi mitotického dělení se docílí přidáním činidla zabraňujícího tvorbě dělicího vřeténka (např. oryzalinu). Poté lze chromozomy

2 Zjednodušené schéma průtokového cytometru. Mikroskopické částice procházejí jedna za druhou světelným paprskem, detektory poté vyhodnotí jejich optické vlastnosti. Třídění částic je umožněno nabitím tvořících se kapek elektrickým nábojem a následným vychýlením v elektrostatickém poli. Moderní přístroje dokáží třídit až 6 různých populací částic současně při rychlostech až desítek tisíc částic za sekundu. Orig. J. Doležel

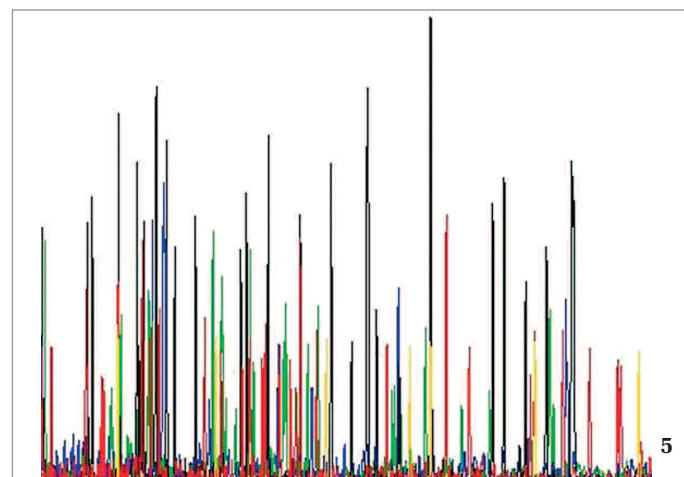
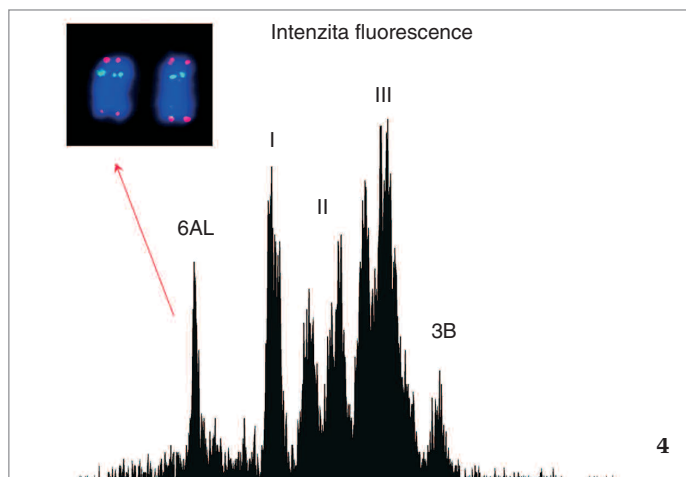
3 Histogram distribuce relativního obsahu DNA (flow karyotyp, uprostřed) a karyotyp pšenice seté. Chromozomy z jednotlivých vrcholů byly vytříděny a poté identifikovány v mikroskopu metodou fluorescenční hybridizace *in situ* (FISH) s fluorescenčně značenými sondami pro krátké repetitivní sekvence GAA (zeleně značené) a Afa (červeně značené). Největší chromozom (3B) vytváří samostatný vrchol a lze ho získat ve vysoké čistotě. Ostatní chromozomy tvoří tři vrcholy, z nichž každý obsahuje několik chromozomů o podobné velikosti.

4 Příklad linie pšenice, u níž chybí krátké rameno chromozomu 6A. Díky tomu je na histogramu vlevo dobře oddělený vrchol odpovídající dlouhému rameni chromozomu 6A (6AL), které je znatelně kratší než ostatní chromozomy. Orig. J. Vrána a M. Kubaláková (obr. 3 a 4)

5 Restrikční profil (tzv. fingerprint) jednoho klonu BAC (vložený fragment DNA v umělém klonovacím vektoru – bakteriálním chromozomu). Fragmenty byly získány rozštěpením DNA klonu pomocí směsi restrikčních enzymů, obarveny fluorescenčními značkami a analyzovány kapilární elektroforézou. Orig. J. Vrána

z buněk pomocí mechanického homogonizátoru (mixeru) převést do izolačního pufru a suspenze se přefiltruje, aby se odstranily velké shluky a buněčné úlomky. Chromozomy se obarví fluorescenčním barvivem specificky se vážícím na DNA a vzorek je tak připravený pro analýzu průtokovým cytometrem.

Výsledkem analýzy chromozomů na základě jejich fluorescenčního profilu je histogram odpovídající relativnímu obsahu DNA jednotlivých typů chromozomů – flow karyotyp (obr. 3, uprostřed). V ideálním případě by každý vrchol (peak) na histogramu reprezentoval jeden chromo-



zom. K tomu by došlo tehdy, kdyby se jednotlivé chromozomy výrazně lišily svou velikostí. Ale už nás snad ani nepřekvapuje, že u pšenice to tak není. Na obr. 3 jsou patrné čtyři vrcholy, mikroskopickou analýzou jsme určili, které chromozomy jsou v daném vrcholu přítomny. Z výsledků vyplývá, že ze standardního pšeničného karyotypu lze získat s vysokou čistotou pouze jeden chromozom, označovaný 3B (chromozom 3 z genomu B). Bylo tedy nutné vyřešit, jak rozlišit a roztrždit zbývající typy chromozomů. A tady máme konečně štěstí, protože snad jediná pozitivní vlastnost pšenice související s polyploidii je vysoká tolerance k aneuploidii. To znamená, že pšenice může postrádat část, nebo dokonce i celý chromozom, a její životaschopnost není nijak ohrožena. Chromozomy z dalších dvou subgenomů mohou totiž ten ztracený funkčně nahradit, jelikož obsahují podobnou genetickou informaci. Díky tomu mohly být v minulosti experimentálně vytvořeny stabilní pšeničné linie, z nichž každá postrádá vždy jedno rameno některého chromozomu (obr. 4). Tyto linie pokrývají celé spektrum pšeničných ramen. My je s úspěchem využíváme pro získání co největšího množství chromozomově specifické DNA pro další kroky směřující k sekvenování kompletního genomu – tvorbu knihoven DNA a fyzických map pro jednotlivá ramena chromozomů.

Jak již bylo řečeno, knihovny DNA jsou soubory bakteriálních klonů, do nichž byly vloženy fragmenty DNA určitého organismu. Soubor klonů může obsahovat celý genom nebo třeba jediný chromozom zkoumaného organismu. Pro jejich tvorbu je nezbytná DNA o velikosti až milionů párů bází, kterou můžeme enzymaticky rozštěpit na úseky o velikosti 100–300 tisíc bází. Získané fragmenty se vloží do klonovacích vektorů. V současné době je nejpoužívanějším vektorem umělý bakteriální chromozom (bacterial artificial chromosome – BAC). Vektor BAC s cizorodou DNA se vloží do bakterie *Escherichia coli*. Jednotlivé klony bakterií nesoucí různé úseky DNA se mohou namnožit a uchovávat. Pro tvorbu knihovny je obvykle potřeba 10–20 µg DNA. Takové množství bývá snadno dostupné, pokud knihovnu sestavujeme z celkové jaderné DNA. V případě chromozomů je ale situace jiná, průměrný pšeničný chromozom obsahuje méně než 1 ng DNA, u ramen je to ještě méně (zhruba o polovinu). Tvorba chro-

mozomově specifických knihoven DNA tedy představuje další výzvu, tentokrát z hlediska počtu chromozomů, které musíme roztrždit. Limitující je množství chromozomů ve vzorku. I při použití dvou přístrojů současně představuje denní limit přibližně 500 tisíc kopií jednoho chromozomu (ramene). Proto byl optimalizován protokol tvorby chromozomově specifické knihovny, kdy k její přípravě stačí 5 µg DNA, což představuje několik milionů až desítek milionů kopií chromozomů/ramen (v závislosti na jejich velikosti). První taková specifická knihovna vznikla v r. 2004 na olomouckém pracovišti Ústavu experimentální botaniky pro chromozom 3B. Pokud víme, jde o první BAC knihovnu z tříděných chromozomů na světě. Brzy následovaly knihovny pro další pšeničné chromozomy a ramena a předpokládáme, že do konce r. 2012 bude genom pšenice knihovnami pokryt. Časová i finanční náročnost je důvodem, proč tento proces trvá relativně dlouho. Třídění dostatečného množství chromozomů a příprava vysokomolekulární DNA zabere obvykle tři až pět týdnů, tvorba knihovny přibližně dva týdny a uspořádání klonů do speciálních plastových destiček mezi třemi až čtyřmi týdny. Jde o značné přístrojové a materiální náklady, takže bez financování z veřejných zdrojů bychom se neobešli.

Již jsem uvedl výše, že jednou z nejdůležitějších aplikací knihoven DNA je tvorba fyzických map, tedy rekonstrukce pořadí dlouhých úseků chromozomů s využitím BAC fingerprintingu (obr. 5). Zde mají chromozomově specifické knihovny pšenice výhodu oproti celogenomovým, protože se porovnává podstatně menší počet klonů (35–100 tisíc proti cca dvěma milionům). Pak je tu již zmíněné úspěšné vyřešení problému s příbuznými subgenomy. První fyzickou mapu pšeničného chromozomu (opět 3B) vytvořili francouzští kolegové z Université Blaise Pascal v Clermont-Ferrand na základě našeho postupu (Paux a kol. 2008).

Tříděné chromozomy jsou potenciálně vhodné i pro přímou metodu sekvenování (shotgun sequencing). Ale stejně jako tvorba knihoven DNA je tato metoda náročná na množství vstupního materiálu (µg až desítky µg DNA). Tento fakt, spolu s finanční náročností a pracností sestavování výsledné sekvence velkých genomů (které navíc obsahují velké procento repetitivních sekvencí), zřejmě stojí za tím, že se tříděné

chromozomy při přímém sekvenování neuplatnily. S ohledem na bouřlivý vývoj na poli nových sekvenačních technologií se ale dá předpokládat, že se to v blízké budoucnosti změní. Bylo publikováno již několik prací využívajících tříděných chromozomů pšenice, žita a ječmene k tzv. survey sekvenování, což se dá volně přeložit jako „nahlédnutí“ do genomů těchto plodin. Stručně řečeno, DNA jednotlivých zkoumaných chromozomů a ramen byla za pomoci enzymů namnožena (výchozí množství cca 50 ng) a sekvenována. Bylo nalezeno velké množství sekvencí odpovídajících genům. Ty se porovnávají s databázemi sekvencí jiných zástupců čeledi lipnicovitých (*Poaceae*), jako jsou čirok (*Sorghum*), rýže (*Oryza*) a válečka (*Brachypodium*). Na základě srovnání vznikly mapy příbuznosti (syntenie) mezi jednotlivými druhy na úrovni chromozomu. Díky tomu se mohlo následně sestavit i předpokládané pořadí genů na jednotlivých zkoumaných chromozomech. Získané poznatky jsou velice důležité a jejich význam při studiu genomu pšenice může předčít jen samotné přečtení celého jejího genomu.

Rozdělit a vládnout

Zmiňované vytvoření fyzické mapy pro chromozom 3B ukázalo nové možnosti při analýze pšeničného genomu, které rozdělení celku na jednotlivé části nabízí, a to mezinárodní spolupráci. Knihovny pro jednotlivá ramena se přidělí různým laboratořím, které mohou vytvářet fyzické mapy a sekvenovat souběžně, čímž se celý proces urychlí a finanční náklady rozdělí. Tato strategie se stala páteří Mezinárodního konzorcia pro sekvenování pšeničného genomu (International Wheat Genome Sequencing Consortium, IWGSC). Naše laboratoř je klíčovým členem tohoto konzorcia právě s ohledem na tvorbu chromozomově specifických genomových zdrojů. IWGSC sdružuje a koordinuje laboratoře z více než 15 zemí. Na jeho internetových stránkách se průběžně zveřejňují pokroky (<http://www.wheatgenome.org>). Do čtyř až pěti let by měla být dokončena kompletní sekvence pšeničného genomu, která se poté poskytne vědecké i šlechtitelské komunitě.

Tato práce byla podporována MŠMT ČR a Evropským fondem pro regionální rozvoj (OP VaVpI č. ED0007/01/01).