

# Primjena Diversity Arrays tehnologije (DART) i sekvenciranja nove generacije u genotipizaciji hrvatskih kultivara graha - principi i metode

---

Valjak, Nikolina

Undergraduate thesis / Završni rad

2017

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:661875>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-10-20**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)



**SVEUČILIŠTE U ZAGREBU**  
**PRIRODOSLOVNO - MATEMATIČKI FAKULTET**  
**BIOLOŠKI ODSJEK**

**PRIMJENA DIVERSITY ARRAYS TEHNOLOGIJE (DArT) I  
SEKVENCIRANJA NOVE GENERACIJE U GENOTIPIZACIJI  
HRVATSKIH KULTIVARA GRAHA  
– PRINCIPI I METODE**

**APPLICATION OF DIVERSITY ARRAYS TECHNOLOGY (DArT)  
AND NEW GENERATION SEQUENCING IN THE GENOTYPING OF  
CROATIAN BEAN CULTIVARS  
- PRINCIPLES AND METHODS**

**SEMINARSKI RAD**

Nikolina Valjak

Preddiplomski studij Znanosti o okolišu  
Undergraduate Study of Environmental Sciences

Mentor: prof. dr. sc. Zlatko Liber

Zagreb, 2017

## Sadržaj

1. UVOD .....	2
2. GRAH ( <i>Phaseolus vulgaris</i> L.).....	3
3. GENETIČKI BILJEZI .....	5
3.1. Raznolikost duljina PCR umnoženih restrikcijskih fragmenata DNA .....	6
3.2. Mikrosateliti ili ponavljajuće jednostavne sekvence.....	7
3.3. Raznolikost pojedinih nukleotida.....	8
3.3.1. Metode temeljene na hibridizaciji .....	9
3.3.2. Metode temeljene na upotrebi različitih enzima .....	9
3.3.3. Metode temeljene na sekvenciranju DNA.....	12
3.4. DArT biljezi .....	15
3.5. DArTseq biljezi .....	18
4. LITERATURA.....	20
5. SAŽETAK.....	22
6. SUMMARY .....	22

## 1. UVOD

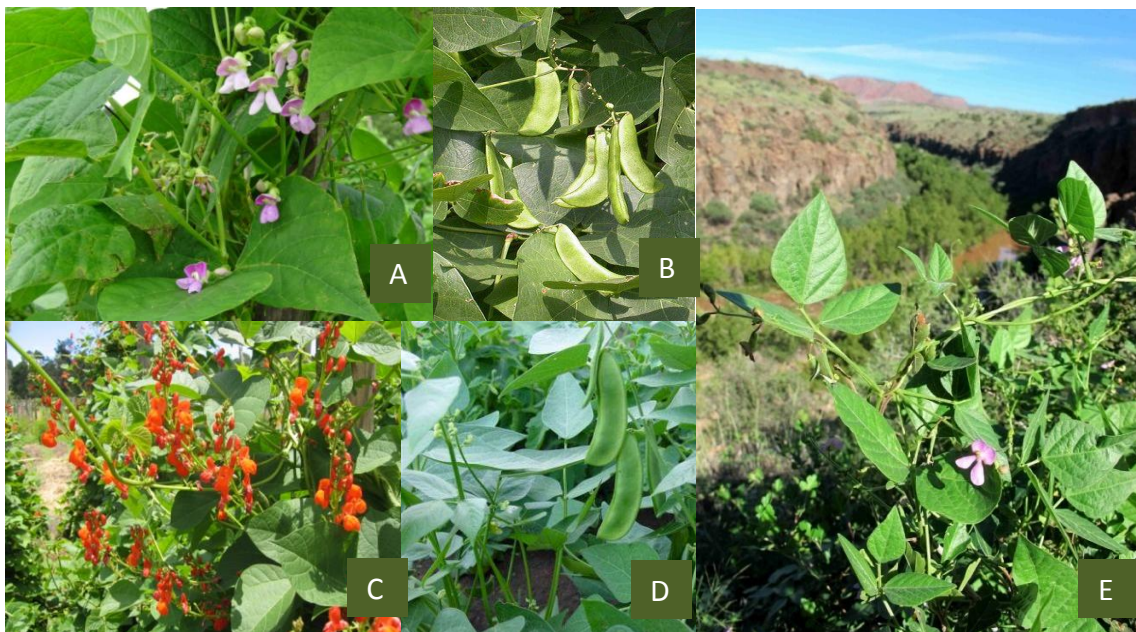
Neishranjenost mineralima jedan je od većih zdravstvenih problema u svijetu. Uzgoj graha bi mogao pomoći u rješavanju ovog problema jer je sjemenka graha važan izvor minerala (P, K, Ca, Mg, Fe, Zn), no ova priča nije tako jednostavna jer je iskorištavanje minerala povezano s udjelom fitinske kiseline u sjemenci graha. Naime, problem je da fitinska kiselina, koja je osnovno skladište fosfora u tkivima biljaka, ima i sposobnost vezanja većine kationa, a nije probavljiva pa povećan unos namirnica bogatih mineralima ne garantira povećanje minerala u ljudskom tijelu. Omjer količine fitinske kiseline i minerala je genetski determiniran te je primjećeno da je različit kod različitih kultivara graha.

Grah u Hrvatskoj ima dugu tradiciju uzgoja, a danas ga uzgajaju mali poljoprivredni proizvođači te on raste na različitim područjima prilagođen lokalnim uvjetima (Carović-Stanko i sur. 2017). U Hrvatskoj je utvrđen veliki broj tradicijskih kultivara koji pokazuju visoki stupanj agronomske i morfološke raznolikosti pa se slična raznolikost očekuje i u količini bioaktivnih hranivih tvari. Isto tako pretpostavka je da je ova raznolikost genetički uvjetovana što je moguće dokazati genotipizacijom upotrebljavajući različite genetičke biljege.

Cilj ovog završnog rada je prikaz osnovnih principa najmodernijih metoda genotipizacije, razmatranih i primjenjenih unutar projekta: „Genetska osnova količine bioaktivnih hranivih tvari hrvatskih tradicijskih kultivara graha“, financiranog od strane Hrvatske zaklade za znanost (<http://beanqual.agr.hr>). Osobita pažnja je obraćena DArTseq metodi detekcije raznolikost pojedinih nukleotida (eng. Single Nucleotide Polymorphism, skraćenica SNP), temeljenoj na sekvenciranju nove generacije (eng. Next Generation Sequencing, skraćenica NGS), a koja se pokazala najpogodnijom za dosezanje glavnih ciljeva projekta.

## 2. GRAH (*Phaseolus vulgaris* L.)

Grah je biljka porodice mahunarki (*Fabaceae*). Mahunarke predstavljaju agronomski važnu porodicu jer su za razliku od npr. žitarica (porodica *Poaceae*) bogati izvor proteina i minerala. Zbog tih svojstava uzgoj mahunarki je u jako velikom porastu. Među mahunarkama grah ima najveću važnost za direktnu upotrebu i jedna je od glavnih vrsta kojom se pokušava iskorijeniti glad u svijetu. Zbog ovih osobina graha nije čudno da su danas vodeće zemlje u uzgoju graha one iz Latinske Amerike i subsaharske Afrike (Akibode i Maredia 2011). U tim zemljama se proizvodi gotovo 2/3 ukupne svjetske proizvodnje (oko 12 milijuna metričkih tona godišnje). Unutar roda *Phaseolus* postoji 70-tak vrsta, ali je samo pet udomaćeno: *P. vulgaris*, *P. dumosus*, *P. coccineus*, *P. acutifolius* i *P. lunatus* (sl. 1). Prema geografskom rasprostranjenju pretpostavlja se da je većina vrsta graha srednjeameričkog podrijetla. Proučavanjem intergenskih jezgrenih sekvenci DNA kao i nekolicine kloroplastnih gena (Delgado-Salinas i sur. 2006.) određena je starost roda na 4-6 milijuna godina, a unutar roda je utvrđeno osam filogenetskih skupina pri čemu je najstarija ona koja u sebi sadrži sve udomaćene vrste osim vrste *P. lunatus*.



Slika 1. Udomaćene vrste graha. A - *Phaseolus vulgaris*, B - *P. dumosus*, C - *P. coccineus*, D - *P. lunatus*, E - *P. acutifolius*

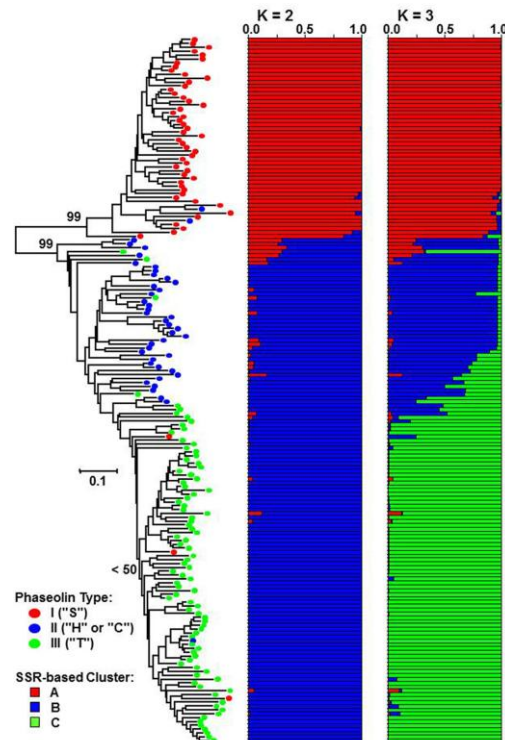
Obični grah (*Phaseolus vulgaris* L.) je ekonomski najvažnija i prema tome najviše uzgajana vrsta graha. Sjemenke ove vrste su izuzetno bogate proteinima, vitaminima, mineralima i vlaknima. Obični grah je samooplodna, diploidna vrsta s 22 kromosoma i veličinom haploidnog genoma između 587 i 637 Mbp (<http://data.kew.org/cvalues>). U prirodi je ova vrsta rasprostranjena od sjevernog Meksika do sjeverozapadne Argentine pri čemu su uočene dvije genetičke skupine: srednjeamerička i andska (sl. 2). Procjenjuje se da su se ove dvije genetičke skupine odvojile prije nekih 165 000 godina (Gaut 2014).



Slika 2. Mjesto i vrijeme diferencijacije dvije različite genske skupine divljeg graha vrste *P. vulgaris* L.: srednjeamerička i andska i njihova domestikacija (Gaut 2014)

Kao dobar biljeg za određivanje pripadnosti sorti ovim genskim skupinama pokazale su se varijacije u nukleotidnom sastavu dijela gena za glavni spremišni protein u sjemenci graha poznat kao fazeolin (Kami i sur. 1995). Obični grah je neovisno udomaćen iz divljih srednjeameričkih i andskih populacija prije oko 8000 godina (Gaut 2014), a nakon otkrića Amerike proširen je na ostale kontinente. Studije fazeolinskih tipova pokazale su da su u Europi prisutna oba genetička skupa s tim da je Andski skup zastupljeniji sa 67%. Osobito je zanimljivo da je izvan Amerike, zbog uzgoja i jednog i drugog genetičkog skupa, došlo prvi put do hibridizacije među njima tako da je u

Europi primjećeno u prosjeku 33 % hibrida (Bellucci et al. i sur. 2014). Slična istraživanja su poduzeta i na hrvatskim kultivarima graha (Carović-Stanko i sur. 2017). U toj studiji je utvrđeno da 68 % analiziranih uzoraka pripada andskom, 27% srednjeameričkom, a samo 5 % hibridnom genetičkom skupu (sl. 3).



Slika 3. Fitch-Margoliash srodstveno stablo temeljeno na udjelu zajedničkih SSR alela između 183 uzorka hrvatskih tradicijskih kultivara. Prosječna pripadnost genetičkim skupovima određena programom STRUCTURE za K2 (srednjeamerički i andski) i K3 (srednjeamerički, andski i hibridni) je prikazana uz srodstveno stablo crvenom, plavom i zelenom bojom za svaku analiziranu jedinku. Klasifikacija prema fazeolinskom tipu je prikazana crvenim, plavim i zelenim krugom na vrhu svake grane stabla. (Carović-Stanko i sur. 2017)

### 3. GENETIČKI BILJEZI

Genetičkim biljgom (eng. genetic marker) smatramo bilo koji dio organizma koji pokazuje neki oblik raznolikosti među istraživanim jedinkama, populacijama, sortama, podvrstama, vrstama i sl. Takav biljeg ne mora nužno biti gen, ali ga se takvim može smatrati s obzirom da se nasljeđuje prema istim načelima. Stoga različite oblike genetičkih biljega zovemo alelima, a njihovo specifično mjesto na kromosomu, lokusom (Šatović 1999). Genetički biljezi se obično dijele na morfološke i molekularne premda se unapređenjem metoda molekularne genetike morfološki biljezi gotovo više ne upotrebljavaju. Unutar molekularnih biljega postoji podjela na proteinske (izoenzimske) i DNA biljege. Razvitkom sve brojnijih i sve efikasnijih DNA biljega proteinski biljezi



se sve rjeđe koriste pa kad govorimo o genetičkim biljezima mislimo ustvari na različite tipove DNA biljega. DNA biljezi otkrivaju genetičku raznolikost izravno na razini DNA. Postoji veliki broj različitih molekularnih tehnika koje su omogućile stvaranje velikog broja lakouočljivih, jeftinih, i ponovljivih DNA biljega. Danas su najčešće upotrebljavani DNA biljezi mikrosateliti ili jednostavne ponavljajuće sekvence (eng. Simple Sequence Repeats, skraćenica SSR) (Moxon i Willis 1999) za dobro proučene genome te raznolikost duljine PCR umnoženih restrikcijskih fragmenata DNA (eng. Amplified Fragment Length Polymorphism – skraćenica AFLP) (Vos i sur. 1995) za slabo proučavane genome. Posljednja znanstvena dostignuća omogućila su detekciju i analizu raznolikosti pojedinačnih nukleotida (eng. Single Nucleotide Polymorphism, skraćenica SNP) na desetke tisuća mjesta unutar cjelokupnog genoma jedinke u istraživanju.

### **3.1. Raznolikost duljina PCR umnoženih restrikcijskih fragmenata DNA**

Raznolikost duljina PCR umnoženih restrikcijskih fragmenata DNA (eng. Amplified Fragment Length Polymorphism, skraćenica AFLP) je selektivna PCR amplifikacija fragmenata nastalih digestijom genomske DNA restrikcijskim enzimima. Postupak se sastoji od tri faze. U prvoj fazi dolazi do restrikcije DNA pomoću dvije vrste restrikcijskih enzima, onih kojima nastaje veći broj fragmenata manje molekularne težine (npr. restrikcijski enzim *MseI*) i onih kojima nastaje manji broj fragmenata veće molekularne težine (npr. restrikcijski enzim *EcoRI*). Nakon toga se u prvoj fazi na krajeve nastalih fragmenata dodaju oligonukleotidni adapteri poznatog slijeda nukleotida. Druga faza je preselektivna i selektivna PCR-amplifikacija restrikcijskih fragmenata temeljena na činjenici da su sekvence adaptera poznate pa se na osnovi njih mogu konstruirati komplementarne sekvence PCR-početnica. Uz to se u preselektivnoj amplifikaciji na 3' krajeve PCR početnica dodaje jedan nasumično izabran nukleotid, a u selektivnoj još dva nasumično izabrana kako bi se PCR-om umnožio samo dio od ukupnog broja restrikcijskih fragmenata. Naime, restrikcijskih fragmenata je u startu tako mnogo da se ne mogu razdvojiti i samim tim detektirati kapilarnom elektroforezom. Treća faza podrazumijeva elektroforezu fragmenata AFLP gdje oni segregiraju po veličini zbog različite pokretljivosti u električnom polju (Mate 2009). Raznolikost duljina PCR umnoženih fragmenata DNA kao metoda je vrlo pouzdana,



ponovljiva, jeftina i lako dostupna, ne zahtijeva prethodno znanje o DNA sekvenci istraživanog organizma, vrlo je informativna zbog mogućnosti istovremene analize velikog broja polimorfnih lokusa, a ko-migrirajući AFLP produkti se uglavnom smatraju homolognima i lokus-specifičnima. Glavni nedostatak ove metode je dominantna ekspresija biljega, pri čemu se dobiva uvid samo u prisutnost ili odsutnost alela, bez utvrđivanja razlika između dominantnih homozigota i heterozigota (Greguraš 2013).

### **3.2. Mikrosateliti ili ponavljajuće jednostavne sekvence**

Mikrosateliti su kratke, uzastopno ponavljajuće jedinice od 1-6 nukleotida koje se ponavljaju od nekoliko pa do nekoliko stotinu puta, stvarajući na taj način DNA sekvencu dugačku od nekoliko desetaka do nekoliko stotina nukleotida. Poznati su pod nazivima jednostavne ponavljajuće sekvence (eng. Simple Sequence Repeats, skraćenica SSR) ili kratka uzastopna ponavljanja (eng. Short Tandem Repeats, skraćenica STR). Nalazimo ih kod svih dosad ispitivanih organizama pa čak i u najmanjim bakterijskim genomima te u jezgrenom, kloroplastnom i mitohondrijskom genomu. Zbog brojnih mutacija koje utječu na broj ponavljajućih sekvenci imaju visoku razinu polimorfizma. Prednosti mikrosatelita pred drugim molekularnim biljezima su što se SSR-aleli mogu otkriti na jednom lokusu jednostavnom provjerom na temelju lančane reakcije polimerazom. Da bi se ti biljezi koristili potrebno je znati nukleotidnu sekvencu na krajevima ponavljajuće regije kako bi se konstruirale specifične PCR-početnice. Broj ponavljanja mikrosatelitnog motiva na nekom lokusu može jako varirati na interspecifičnoj, intraspecifičnoj i intrapopulacijskoj razini. Uzrok tome je visoka stopa mutacije (između  $10^{-2}$  i  $10^{-6}$  po lokusu po generaciji). Pretpostavlja se da mutacijski mehanizmi mogu biti vrlo kompleksni te da još uvijek nisu dovoljno razjašnjeni. Replikacijsko klizanje je najčešći uzrok varijabilnosti, a uzrokuje male varijacije od 1 do 2 ponavljanja. Tijekom replikacije dolazi do prolaznog razdvajanja i proklizavanja DNA lanca što dovodi do krivog sparivanja lanaca, a to za posljedicu ima stvaranje omče s jednim ili dva ponavljanja 'viška' na jednom lancu. U sljedećim replikacijskim ciklusima omča može poslužiti kao kalup; ukoliko se takva omča nalazi na rastućem lancu, rezultat će biti povećanje broja ponavljanja (eng. forward slippage), a ukoliko se nalazi na lancu kalupu, rezultat će biti smanjenje broja ponavljanja (eng.

backward slippage)(Greguraš 2013). Mikrosateliti imaju mnoge osobine idealnih genetičkih biljega: lako su dostupni, visoko polimorfni i stoga vrlo informativni, ponovljivi, pouzdani, lako mjerljivi, kodominantni, neutralni i dobro pokrivaju cijeli genom. Kao i svi genetički biljezi i mikrosateliti imaju neke nedostatke. Najveći nedostatak je taj što su pretežno „vrsno-specifični“ biljezi. Naime, primjenjivi su ne usko srodnim vrstama, a za izradu PCR-početnica potrebna je znatna laboratorijska stručnost, te značajan utrošak vremena i novca (Greguraš 2013).

### **3.3. Raznolikost pojedinih nukleotida**

Raznolikost pojedinih nukleotida (eng. Single Nucleotide Polymorphism, skraćenica SNP) nastaje kao rezultat brojnih točkastih mutacija diljem genoma uz pretpostavku da na jednom SNP mjestu mogu biti samo dva od četiri moguća nukleotida. Za ovaj tip polimorfizma je poznato da je odgovoran za većinu genetičkih varijacija. Naime, pretpostavlja se da 80% svih genetičkih polimorfizama odgovornih za fenotipske razlike između individua su točkaste mutacije jednog nukleotida. Zbog svojih osobina SNP biljezi predstavljaju novu generaciju molekularnih biljega s širokim spektrom upotrebe. Ono što SNP biljege čini pogodnima za istraživanja je to što su učestalo i ravnomjerno raspoređeni po genomu, evoluiraju po dobro poznatim, jednostavnim mutacijskim modelima (eng Simple Mutation Model), kao što je model beskonačnih lokacija (eng. Infinite Site Model) (Greguraš 2013). SNP lokusi imaju nisku stopu mutacije ( $10^{-8}$  do  $10^{-9}$  po generaciji) zbog čega je i homoplazija mnogo manje izražena. Radi se o kodominantnim biljezima, no za razliku od mikrosatelita, oni su bialelni. To znači da njihova moć razlučivanja ne proizilazi iz velikog broja alela, nego iz velikog broja lokusa korištenih u analizi. Značajan je napredak postignut u zadnjih desetak godina u povećanju točnosti, efikasnosti i automatizaciji otkrivanja SNP biljega. Upotreba SNP biljega podrazumijeva dobru istraženost genoma istraživane vrste. Osobito je vrijedno za detekciju velikog broja SNP biljega imati potpuno sekvencionirani genom istraživane vrste. Danas postoji veliki broj metoda otkrivanja SNP-raznolikosti. Sve te metode su kombinacija različitih metoda razdvajanja alela te različitih metoda detekcije signala. Danas se najčešće korištene metode za razdvajanja i detekciju SNP-alela mogu podijeliti na: metode temeljene na hibridizaciji, metode temeljene na upotrebi različitih

enzima te metode temeljene na sekvenciranju DNA. Neke od ovih metoda zahtijevaju prethodnu PCR amplifikaciju DNA dok druge upotrebljavaju ukupnu genomsku DNA.

### **3.3.1. Metode temeljene na hibridizaciji**

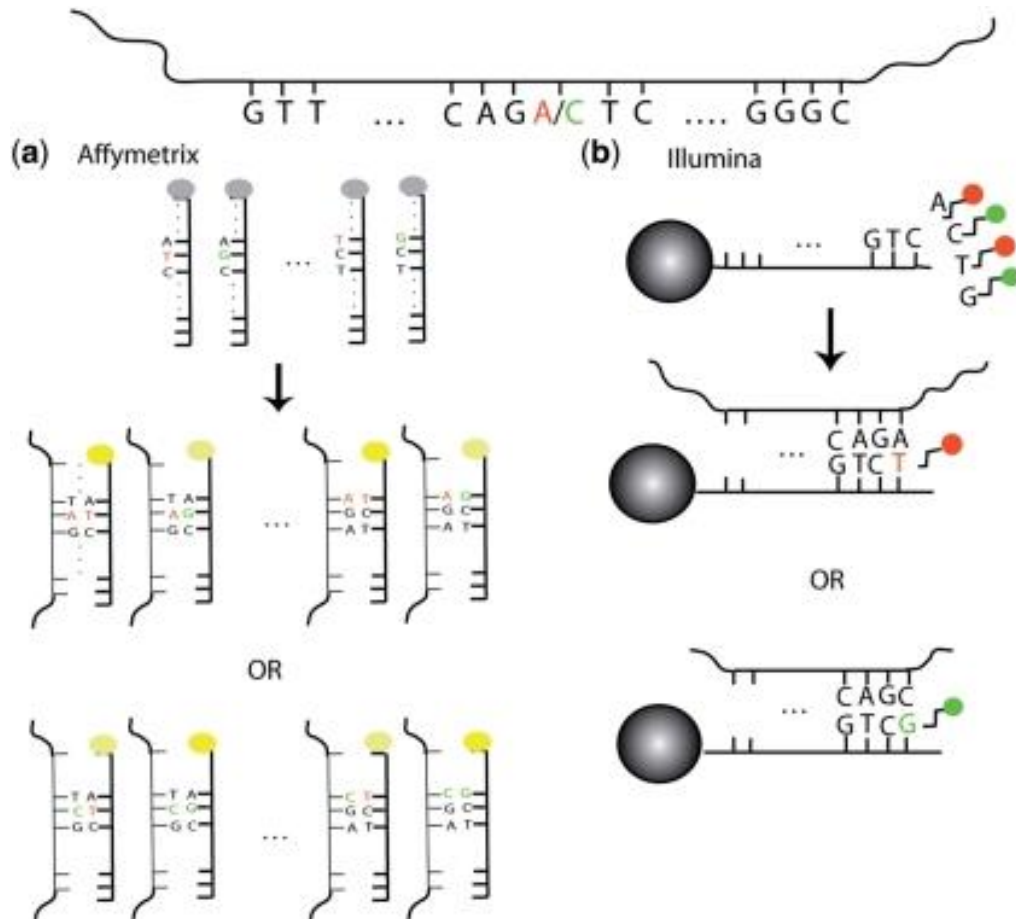
Postoji veliki broj metoda koje se osnivaju na hibridizaciji komplementarne DNA probe na SNP mjesto (npr. Dynamic allele-specific hybridization (DASH), Molecular beacons, SNP microarrays itd.).

U zadnje vrijeme, prije svega zbog generiranja velikog broja podataka na brojnim SNP mjestima istovremeno, „SNP microarrays“ (SNP mikromatrica ili SNP mikročip) pristup je sve češći. Ovaj pristup se osniva na hibridizaciji DNA na mikromatrice (čipove) na kojima su pričvršćene stotine tisuća SNP proba iz cijelog genoma istraživane vrste. Danas najčešće platforme na kojima se osniva detekcija SNP biljega su one od tvrtke Affymetrix ili tvrtke Illumina (sl. 4). Tako Affymetrix SNP 5.0 GeneChip omogućuje istovremenu analizu preko 500000 SNP mjesta diljem ljudskog genoma. Mogućnost detekcije ovako velikog broja SNP biljega omogućuje njihovu upotrebu u ranoj detekciji genetičkih bolesti kao što su reumatoidni artritis, rak prostate, neonatalni dijabetes i sl.. Problem s DNA čipovima je skupoća njihove konstrukcije, ali kada su jednom napravljeni i komercijalno dostupni na tržištu, cijena analize SNP alela strmoglavno pada i postaje zasigurno najjeftinija po pojedinom SNP lokusu. Illumina BARCBean6K\_3 BeadChip za detekciju 6000 polimorfnih SNP lokusa u genomu graha razvijen je 2015. godine (Song i sur. 2015), međutim još uvijek nije komercijalno dostupan pa ga se nije moglo iskoristiti za potrebe istraživanja projekta „Genetska osnova količine bioaktivnih hranivih tvari hrvatskih tradicijskih kultivara graha“.

### **3.3.2. Metode temeljene na upotrebi različitih enzima**

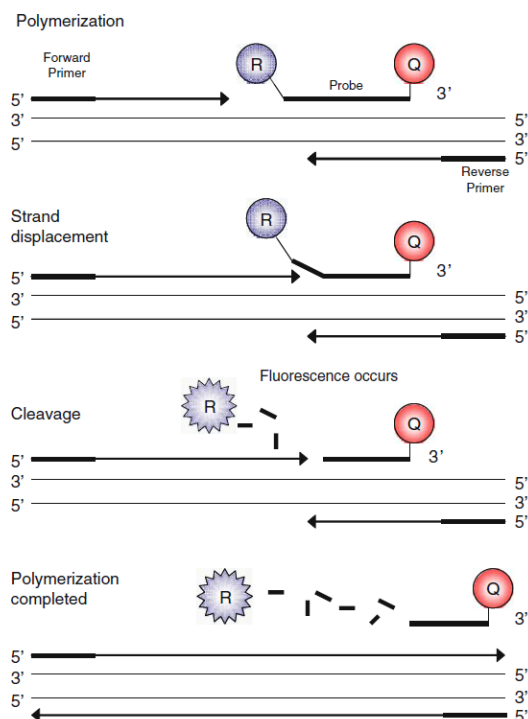
Različiti tipovi enzima (restriksijske endonukleaze, DNA ligaze, DNA polimeraze, DNA nukleaze, Flap endonukleaze itd.) su do sada upotrebljavani kako bi se generirali SNP biljezi. Jedna od često korištenih metoda, a koja se temelji na upotrebi 5' nukleazne aktivnosti Taq DNA polymerase, je TaqMan<sup>®</sup> metoda. Ova metoda je idealna za istraživanje koje se temelji na upotrebi manjeg broja SNP lokusa i velikog broja DNA uzoraka koje treba analizirati. TaqMan metoda je PCR reakcija kod koje uz Taq DNA

polimerazu sudjeluju obje PCR početnice te obje jednolančane oligonukleotidne DNA probe (SNP probe) svaka s karakterističnom bazom za SNP mjesto te florescencijskom bojom na jednom kraju i utišivačem (eng. quencher) na drugom kraju.



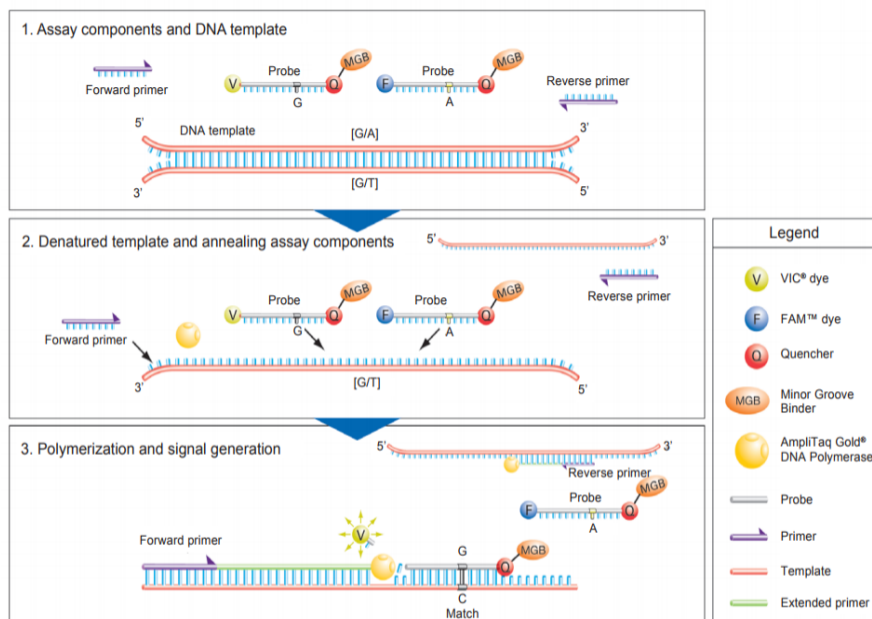
Slika 4. Prikaz SNP microarray tehnologije. Na vrhu je prikazan fragment DNA koji ima A/C SNP mjesto koje se nastoji istražiti ovim pristupom. (a) Affymetrix metoda upotrebljava 25-merne probe komplementarne jednom ili drugom alelu. Fragment DNA se veže s obje probe bez obzira koji alel nosi, ali signal florescentne boje je puno jači ako je svih 25 nukleotidnih baza komplementarno (tamno žuto). (b) Kod Illumina metode 50-merna komplementarna sekvenca (proba) tik do određenog SNP alela je pričvršćena na čip. Nakon što se na ovu sekvenču hibridizira fragment istraživane DNA dolazi do produženja 50-merne sekvenče na SNP alelu s florescentno označenim nukleotidima. Detekcijom oslobođene florescencije odredi se SNP alel na tom mjestu. (preuzeto iz LaFramboise 2009)

Blizina utišivača reducira florescencijski signal boje. Kada Taq DNA polimeraza prolazi mjesto vezanja potpuno komplementarne SNP probe, svojom 5' nukleaznom aktivnošću odvaja florescencijsku boju od utišivača pri čemu dolazi do specifične florescencije za tu bazu (SNP alel)(sl. 5).



Slika 5. PCR-amplifikacija i detekcija uz pomoć fluorescentne probe i 5' nukleazne aktivnosti Taq DNA polimeraze. Dvije boje fluorescencijska (R) i utišavač (Q) su zakačene na oligonukleotinu probu potpuno komplementarnu segmentu DNA koji se amplificira između PCR početnica. Kada su obje boje vezane na probu nema fluorescencije, međutim kada Taq DNA polimeraza dođe do mjesta na kojem je proba, svojom 5' nukleaznom aktivnošću oslobađa fluorescencijsku boju što dovodi do fluorescencijskog signala (preuzeto iz Livak 2003)

Jedan od ključnih momenata TaqMan metode je činjenica da probe nekomplementarne u jednoj bazi imaju nižu „melting“ temperaturu ( $T_m$ ) koja im pri optimalnom izboru temperature vezanja početnice i temperature ekstenzije pri PCR reakciji ne dozvoljava vezanje na DNA kalup. U TaqMan metodi se obično koriste FAM i VIC fluorescencijska boja, svaka specifična za jedan odnosno drugi SNP alel. Dodatak tzv. minor groove binder (MGB) dodatka na utišivač povisuje  $T_m$  temperaturu i time omogućuje upotrebu kraćih oligonukleotidnih proba i bolju diskriminaciju dva SNP alela (Slika 6.). TaqMan® metoda je bila jedna od opcija detekcije SNP biljega za potrebe istraživanja na projektu „Genetska osnova količine bioaktivnih hranivih tvari hrvatskih tradicijskih kultivara graha“ međutim cijena provedbe ove metode, s obzirom da se zahtijevala analiza velikog broja SNP biljega i velikog broja DNA uzoraka, višestruko je premašila financijske mogućnosti projekta.

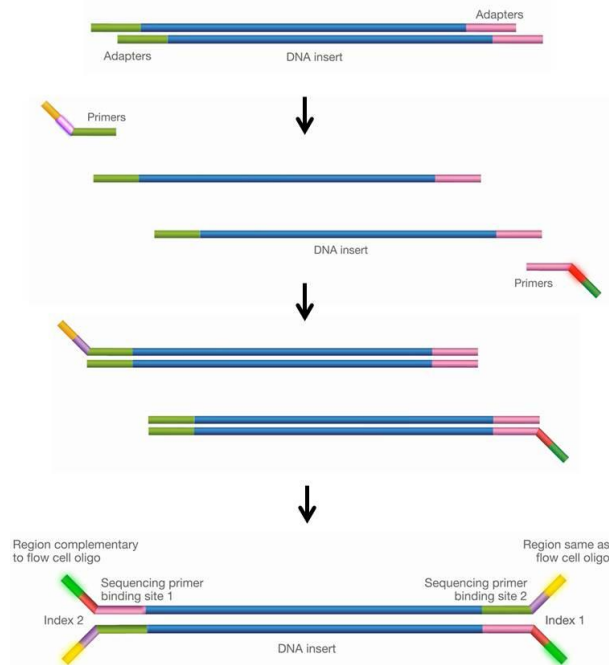


Slika 6. TaqMan<sup>®</sup> SNP Genotyping Metoda. Razlučivanje SNP alela se postiže selektivnim vezivanjem jedne od dvije TaqMan MGB probe. (preuzeto iz TaqMan<sup>®</sup> SNP Genotyping Assays – Applied Biosystems Product Bulletin)

### 3.3.3. Metode temeljene na sekvenciranju DNA

Sekvenciranje DNA je zlatni standard u detekciji SNP polimorfizma (Kwok i Duan 2003). Do sada najčešći pristup je bio direktno sekvenciranje fragmenta DNA umnoženog lančanom reakcijom polimerazom koristeći tzv. „dye-terminator“ kemiju i kapilarnu elektroforezu na automatskim DNA sekvencerima. Međutim, s pojavom sekvenciranja nove generacije (eng. Next Generation Sequencing, skraćena NGS) i većim brojem više ili manje sličnih platformi (npr. Roche GS/454, Illumina, Ion Torrent, Nanopore itd.) sve više se ove tehnologije upotrebljavaju za pronalaženje SNP lokusa i detekciju SNP alela. Bez obzira na platformu prvi korak u svim NGS-sekvenciranjima je konstrukcija DNA biblioteka iz DNA ili RNA molekula izoliranih iz nekog biološkog uzorka. Budući da su izolirane molekule DNA ili RNA predugačke da bi se odmah počele sekvencirati potrebno ih je prvo fragmentirati. Fragmentacija se može obaviti različitim tehnikama kao što su sonikacija, nebulizacija ili tretmanom restriktivnim enzimima. Nakon fragmentacije obično slijedi skupljanje fragmenata željene veličine. Ključni korak u pripremi DNA biblioteke za NGS-sekvenciranje je ligacija adaptera na oba kraja svakog fragmenta DNA. Adapteri su umjetno sintetizirane sekvence DNA koje u sebi imaju različite dijelove nužne u procesu sekvenciranja kao

što su: (1) univerzalne početnice koje iniciraju početak reakcije sekvenciranja, (2) PCR početnice za povećanje broja istih DNA fragmenata, (3) sekvence nužne za prihvaćanje na podlogu reakcijske pločice ili reakcijske kuglice (eng. bead), (4) sekvence barkoda kada se sekvencira istovremeno više DNA ili RNA uzoraka (Slika 7).



Slika 7. Konstrukcija DNA biblioteka iz DNA ili RNA molekula izoliranih iz nekog biološkog uzorka. Fragmentacija DNA i ligacija adaptera na oba kraja fragmenta DNA je ključni korak u pripremi DNA biblioteka za NGS-sekvenciranje

Od svih metoda sekvenciranja nove generacije Illumina platforma trenutno je najčešće korištena. Srž sekvencijske tehnologije koju upotrebljava ova platforma su fluorescencijski označeni nukleotidi s reverzibilnim terminatorom (Bentley et al 2008). Budući da je ova tehnologija u stvari samo oblik sekvenciranja sintezom, vrlo je slična klasičnom Sangerovom sekvenciranju, ali s jednom važnom razlikom, terminalni nukleotid samo privremeno zaustavlja elongaciju u lančanoj reakciji polimerazom. Naime, nakon ugradnje odgovarajućeg nukleotida i očitavanja njegove specifične fluorescencije CCD kamerom, dio nukleotida odgovoran za fluorescenciju i zaustavljanje elongacije se izrezuje te dolazi do elongacije za jedan novi nukleotid. Cijela procedura se ponavlja do kraja DNA fragmenta koji se nastoji sekvencirati (Slika 8). Illumina-sekvenciranje nije ništa drugo do milijuni PCR reakcija na milijunima fragmenata DNA istovremeno, pri čemu se minijaturnom CCD kamerom (Slika 9) snime milijuni fluorescencijskih signala prilikom ugradnje svakog novog nukleotida. Cijela



konstrukcija pločice na kojoj se odigravaju milijuni ovih reakcija (Illumina sequencing flow cell) osmišljena je tako da se ugradnja nukleotida i snimanje fluorescencije odvija istovremeno na svim lancima tj. u svakoj od milijunskih PCR reakcija. U stvarnosti jedan mali dio lanaca unutar brojnih skupina s istim fragmentima DNA (eng. clusters) gubi usklađenost i stvara šum u očitavanju signala. Kako se ponavljaju sekvencijski ciklusi tako ovaj šum postaje sve veći i veći. Postepeni gubitak usklađenosti s povećanjem broja sekvencijskih ciklusa je glavni razlog zašto se duljina čitanja po svakom lancu kreće samo između 125 i 300 nukleotidnih baza (ovisno o modelu uređaja /HiSeq ili MiSeq/ i podešenoj kvaliteti očitavanja). Nakon završetka očitavanja fluorescencijskih signala svih nukleotida u svim lancima se obrađuju software-ski izgrađujući milijune lanaca koji se mogu i međusobno preklapati. Ukoliko je sintetiziran jako veliki broj ovih lanaca, do mjere da se nalaze na stotine gotovo istih lanaca, moguće je uz pomoć računalnih programa sklopiti cijeli genom analizirane vrste (eng. De Novo Sequencing). Postotak greške kod Illumina-sekvenciranja nove generacije je najniži od svih do sada poznatih platformi i iznosi nešto manje od 1 %. Za sekvenciranje nove generacije s Illumina-platformom se najčešće koriste dva uređaja HiSeq i MiSeq (<https://www.illumina.com/systems/sequencing-platforms.html>). Koristeći HiSeq sustav uz podešenu najveću kvalitetu očitavanja prosječno po DNA fragmentu se očita 125 baza. MiSeq sustav očitava nešto duže fragmente, u prosjeku od oko 250 baza. Prednost HiSeq sustava je u očitavanju većeg broja fragmenata u odnosu na MiSeq sustava. Kod HiSeq sustava se očita do 8 milijardi fragmenata (više od Tb baza) i to s oba kraja fragmenta (eng. Paired-End), dok MiSeq učita 50-tak milijuna fragmenata ukupne veličine od oko 15 Gb baza. Postoje tri koraka u generiranju podataka tijekom Illumina-sekvenciranja. Prvo se nakon svakog ciklusa sinteze jedne baze, uz pomoć kontrolnog računalnog programa, analiziraju slike uhvaćene CCD kamerom (Slika 9) na način da se prvo lociraju klasteri s istim fragmentima DNA, a zatim se utvrdi intenzitet signala i šum u svakom klasteru. Izlazni podaci iz ovog koraka ulaze u sljedeći gdje Real Time računalni program (RTA) određuje nukleotidne baze, računa mjeru kvalitete sekvenciranja te izbacuje iz analize loša očitavanja. U trećem koraku određene baze se prebacuju i sažimaju u FASTQ podatke uz pomoć računalnog programa CASAVA. Različiti uzorci DNA se mogu indeksirati i sekvencirati zajedno (tzv. multiplex pristup). Nakon završenog sekvenciranja razdvajanje različitih DNA uzoraka (demultiplexing) se odvija u trećem koraku generiranja podataka i sastavni su dio FASTAQ datoteke (Wang 2016).

Metoda temeljena na sekvenciranju DNA i poznata pod nazivom DArTseq ([www.diversityarrays.com](http://www.diversityarrays.com)) pokazala se po financijskim pokazateljima, ali i informativnošću te brojem identificiranih SNP biljega, kao najpogodnija za dosezanje postavljenih ciljeva projekta: „Genetska osnova količine bioaktivnih hranivih tvari hrvatskih tradicijskih kultivara graha“. Ova metoda je spoj dobrih osobina hibridizacije DNA na mikromatrice (metoda poznata kao DArT) i brojnih dobrih osobina Illumina sekvenciranja nove generacije, zbog čega u nazivu postoji dodatak seq (DArTseq).

### 3.4. DArT biljezi

Diversity Arrays Technology (DArT) je metoda koja se osniva na hibridizaciji. DArT biljezi su dominantni biljezi. Ova metoda je razvijena prije nešto više od petnaest godina (Jaccoud i sur. 2001). Protokol je optimiziran od strane tvrtke Diversity Arrays Technology u Canberra-i, Australija ([www.diversityarrays.com](http://www.diversityarrays.com)). DArT protokol se sastoji od tri glavna koraka (Slika 10):

#### 1) Izgradnja DNA-matrice (eng. microarrays ili DNA chip)

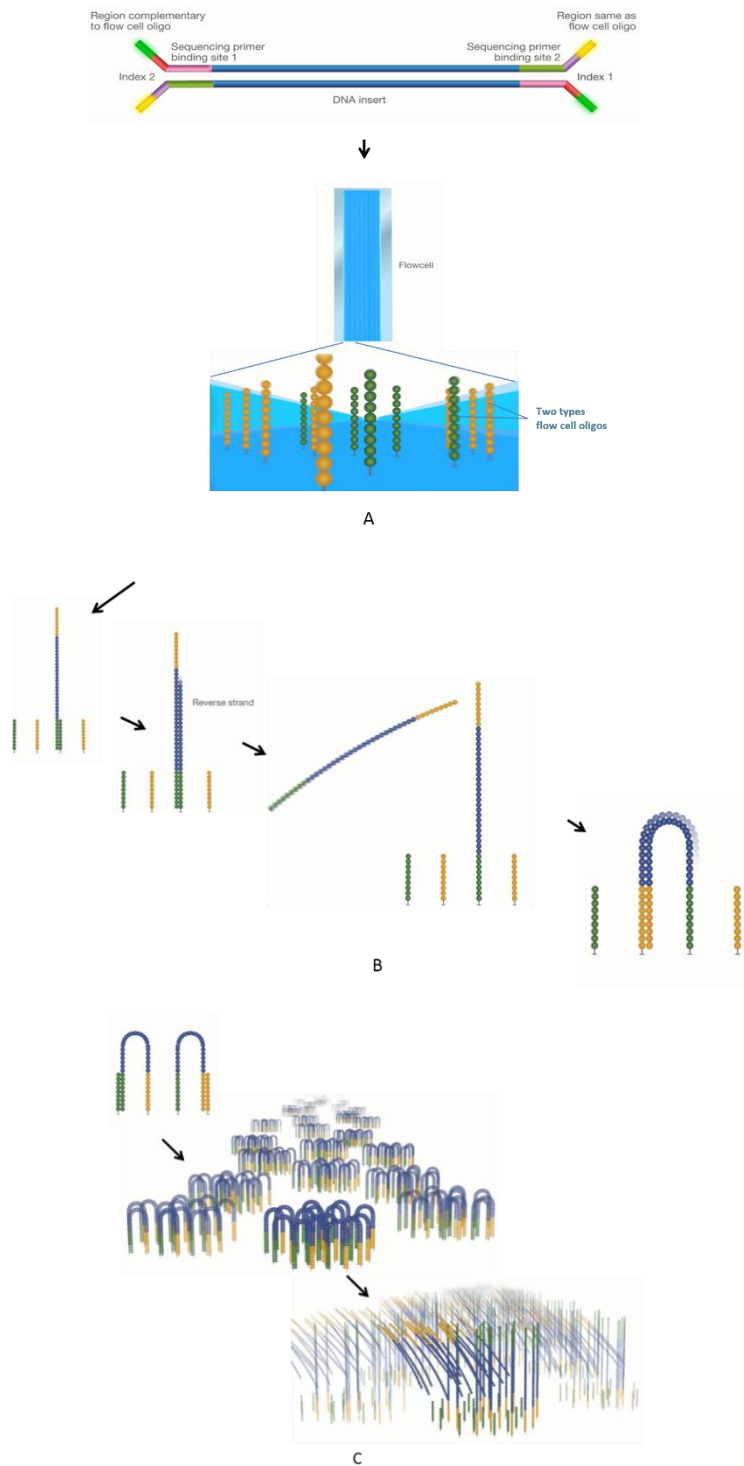
- konstrukcija genomske (DNA) biblioteke neke svojte (genomic representation) na način da se nastoji smanjiti kompleksnost cjelokupnog genoma. Smanjenje kompleksnosti se postiže tako da se genomska DNA reže obično s dva ili tri tipa restriksijskih enzima. Nakon toga se na dobivene krajeve fragmenata ligiraju sintetski DNA adapteri od kojih samo neki imaju redosljed nukleotida komplementaran PCR-početnicama u budućoj lančanoj reakciji polimerazom.
- utiskivanje genomske biblioteke na staklene ploče Nakon provedene lančane reakcije polimerazom dobiveni fragmenti (10000 – 20000) se prevode u jednolančane DNA i pričvršćuju na staklene pločice (eng. microarrays).

#### 2) Genotipizacija

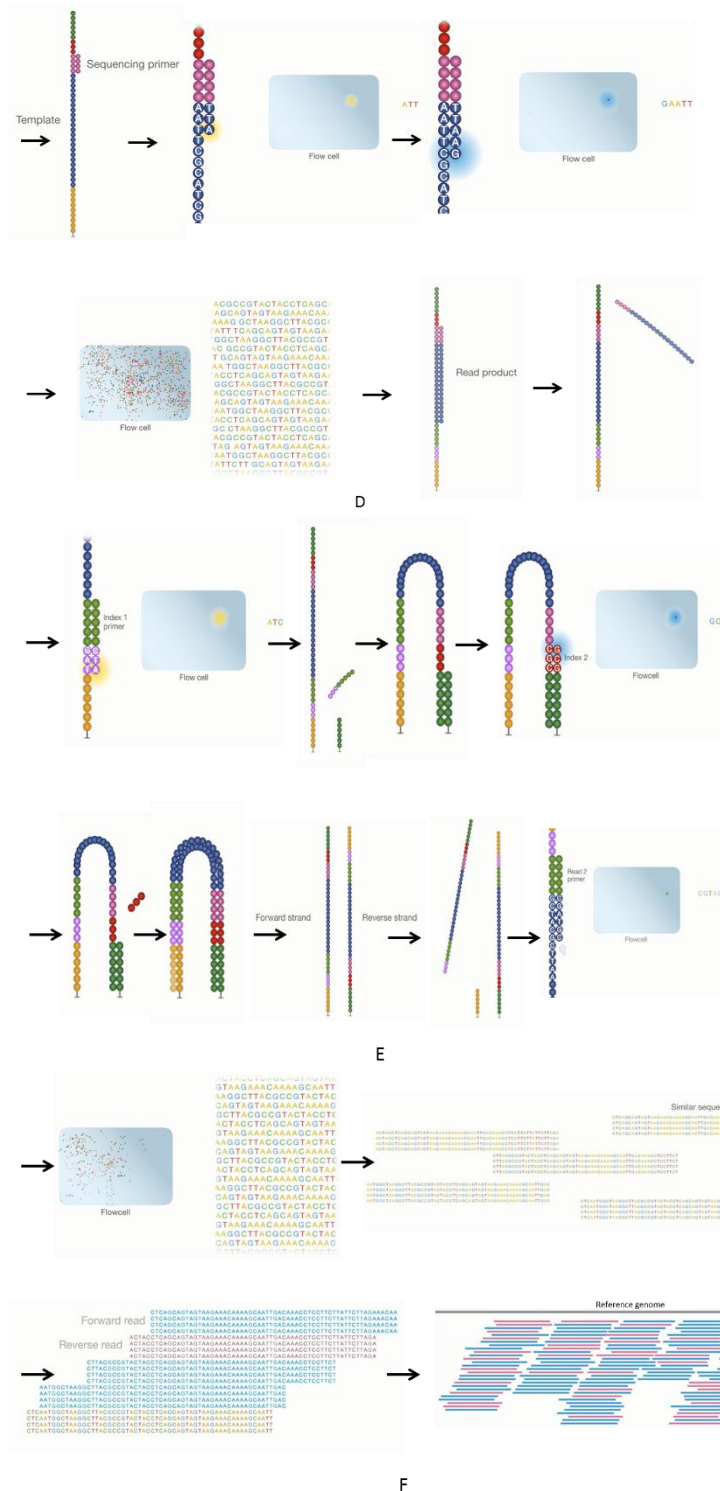
- fragmentiranje DNA jedinki koje se genotipiziraju, označavanje DNA fragmenata jedinki koje ulaze u analizu, hibridizacija tih fragmenata na DNA-fragmente na mikromatrici te ispiranje i skeniranje mikromatrice

#### 3) Analiza dobivenih podataka

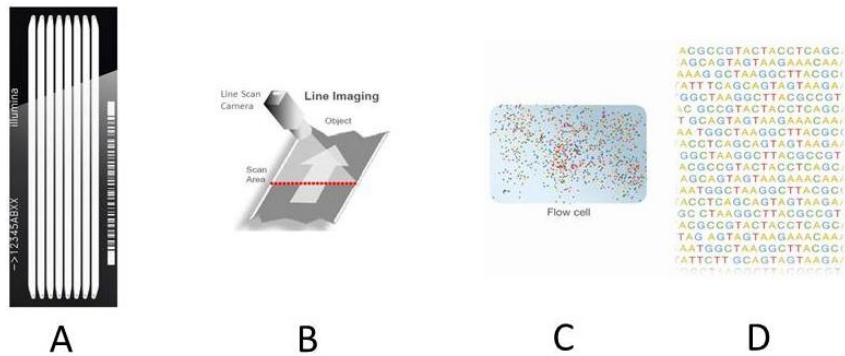
- detekcija fluorescencije, analiza polimorfizma (stotine i tisuće polimorfnih DNA biljega po genomu), primjena dobivenih podataka (npr. konstrukcija genetičkih karata)



Slika 8 A, B, C. Osnovni principi Illumina-sekvenciranja A. Fragmenti DNA s dodanim adapterima dodaju se u kanale Illumina Flow cell pločice na koju su pričvršćena u milijunima kopija dva tipa jednolančanih nukleotida komplementarna dijelu adaptera na oba kraja fragmentirane DNA; B. Hibridizacija jednolančanih fragmenata DNA na oligonukleotidne adaptere i PCR sinteza novog lanca koji je kovalentno pričvršćen na oligonukleotid u kanalu pločice. Lanac koji je služio kao kalup u PCR sintezi se ispire, a nosintetizirani lanac se savija i hibridizira sa svojim drugim krajem sa susjednim oligonukleotidom na pločici formirajući strukturu nazvanu most (eng. bridge) ; C. Nakon sinteze novih lanaca dolazi do njihova razdvajanja te ponovne hibridizacije s okolnim slobodnim oligo nukleotidima na pločici u brojnim ciklusima PCR-umnožavanja (eng. bridge amplification). Na ovaj način umnože se brojni fragmenti DNA dobiveni izgradnjom biblioteke na način da se formiraju klasteri istih fragmenata. Na kraju ovog postupka selektivno se zadrže samo fragmenti pričvršćeni na jedan tip oligonukleotida (prilagođeno <https://www.youtube.com/watch?v=9YxExTSwgPM>)



Slika 8 (nastavak) D, E, F. Osnovni principi Illumina-sekvenciranja D. Uz pomoć sekvencijske početnice i kalupa sintetizira se novi lanac dodavanjem fluorescencijski označenih nukleotida s reverzibilnim terminatorom DNA polimeraze. Fluorescencija svakog dodanog nukleotida se bilježi CCD kamerom na milijunima mjesta na Illumina Flow cell pločici. Nakon završene sinteze i detekcije novostvoreni lanac se ispiri s Illumina Flow cell pločice; E. Sekvenciranje jednog od ukupno dva barkoda na svakom prisutnom fragmentu DNA (jako važno kada se sekvencira istovremeno više DNA uzoraka). Nakon ispiranja sekvence barkoda jednolančana DNA prikvačena na oligonukleotid Illumina pločice stvara most povezivanjem drugog kraja lanca s drugim tipom oligonukleotida pločice. U sljedećem koraku sintetizira se komplementarni lanac i oslobađa lanac kalupa kako bi se izvršilo sekvenciranje barkoda na drugoj strani lanca i sekvenciranje iz drugog smjera (Paired-End sekvenciranje); F. Uz pomoć računalnog programa se formiraju klasteri sličnih sekvenci te ukoliko postoji referentni genom moguće je u potpunosti složiti genom istraživane vrste (de novo sequencing) ((prilagođeno <https://www.youtube.com/watch?v=9YxExTSwgPM>))



Slika 9. A. Illumina „flow cell“ je posebno dizajnirana staklena pločica koja se sastoji od osam reakcijskih kanala (tzv. fluidic channels). DNA biblioteke pripremljene za sekvenciranje se nanose u kanale i hibridiziraju na već usidrene jednolančane DNA pričvršćene unutar kanala. U svakom koraku sekvencijskog postupka stara DNA polimeraza i neiskorišteni modificirani nukleotidi se ispumpaju van kanala, a nova polimeraza i modificirani nukleotidi se upumpavaju u kanal kroz otvore na krajevima pločice; B. CCD kamera snima fluorescencijske signale u reakcijskim kanalima; C. Prilikom ugradnje svakog novog nukleotida CCD kamera snimi milijune fluorescencijskih signala u brojnim PCR reakcijama; D. Svaki ugrađeni nukleotid fluorescira drugom bojom što se računalno prevodi u slijed nukleotida u fragmentu DNA.

### 3.5. DArTseq biljezi

DArTseq biljezi su spoj dobrih osobina hibridizaciji DNA na mikromatrice (DArT biljezi) i brojnih dobrih osobina Illumina sekvenciranja nove generacije (seq). Kao što je objašnjeno u prethodnom poglavlju DArT biljezi se temelje na redukciji kompleksnosti cjelokupnog genoma upotrebljavajući restriksijske endonukleaze te na hibridizaciji tih restriksijskih fragmenata na DNA-mikromatrice (microarrays ili DNA chip) kako bi se detektirale tisuće biljega po genomu. Kod DArTseq biljega redukcija kompleksnosti genoma je istog tipa, ali se detekcija osniva na Illumininoj platformi sekvenciranja nove generacije (NGS).

Kod DArT genotipiziranja koje se temelji na NGS sekvenciranju koriste se dvije metode redukcije kompleksnosti genoma: PstI\_adapter/TaqI/HpaII\_adapter ili PstI\_adapter/TaqI/HhaI\_adapter pri čemu TaqI restriksijski enzima ima ulogu eliminatora dijela PstI -HpaII tj. PstI-HhaI fragmenta. PstI restriksijska mjesta imaju adaptere u kojima se nalaze barkodovi (nukleotidi) karakteristični za određenu jedinku u analizi te redosljed nukleotida komplementaran sekvencijskim početnicama i redosljed komplementaran oligonukleotidima pričvršćenim na „Illumina Flow cell pločice“. Daljnji korak u DArTseq proceduri je Illumina NGS-sekvenciranje detaljnije opisano u poglavlju 1.3.3.3. *Metode temeljene na sekvenciranju DNA*. Nakon sekvenciranja dobiveni FASTQ file-ovi se filtriraju po kvaliteti (prag je podešen na 90 % pouzdanosti) i po individuama (barkod).

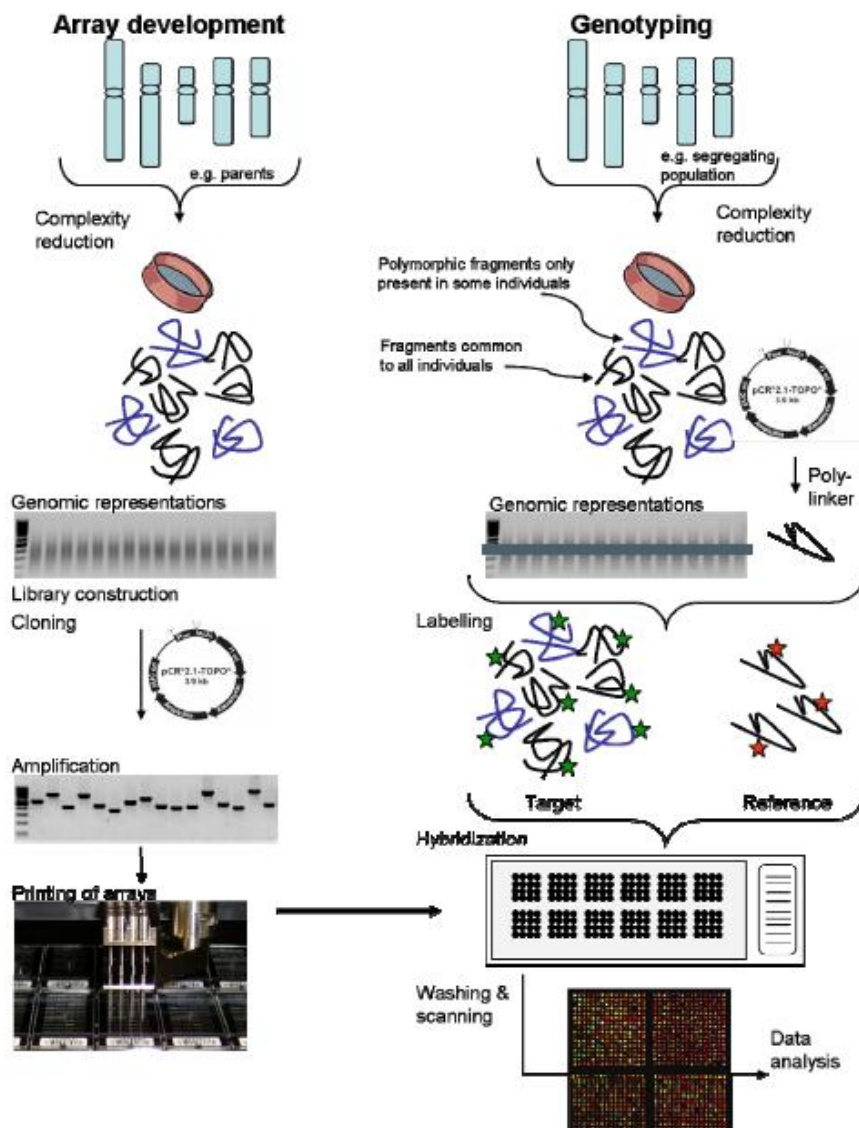


Figure 1. Kratki pregled DArT postupka. U prvom koraku konstruira se genomska biblioteka na način da se ukupna genomska DNA izrezuje u kratke DNA fragmente koja se zatim označavaju fluorescencijski te na kraju utiskuje na stakalca (microarray construction). U drugom koraku DNA jedinki koje se žele genotipizirati se fragmentira, fluorescencijski označi i hibridizira s fragmentima pričvršćenim na mikromatrice. Na kraju postupka poseban računalni program identificira pojedine klonove DNA-fragmenata kao prisutne (1) ili odsutne (0) ovisno o tome jesu li komplementarni DNA slijedovi prisutni u genomskoj biblioteci na stakalcu (preuređeno iz Wittenberg 2007).

U konačnici ako je genom analizirane vrste u potpunosti sekvenciran moguće je dobivene sekvence povezati sa referentnim genomom u bazi kakva je npr. Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>) ili Blast (<https://blast.ncbi.nlm.nih.gov>). Budući da je genom graha sekvencioniran, dobivene DArTseq sekvence bilo je moguće povezati s referentnim genomom.



#### 4. LITERATURA

Akibode S, Maredia M, 2011. Global and Regional Trends in Production, Trade and Consumption of Food Legume Crops.

Bellucci E, et al. 2014. Genomics of Origin, Domestication and Evolution of *Phaseolus vulgaris*. U: *Genomics of Plant Genetic Resources*. Ur. R. Tuberosa et al., Springer, Dordrecht, pp. 483-502. Doi: 10.1007/978-94-007-7572-5

Carović-Stanko K, et al. 2017. Genetic Diversity of Croatian Common Bean Landraces. *Frontiers in Plant Science*, 8, doi: 10.3389/fpls.2017.00604

Delgado-Salinas A, Bibler R, Lavin M, 2006. Phylogeny of the Genus *Phaseolus* (Leguminosae): A Recent Diversification in an Ancient Landscape. *Systematic Botany*, 31(4), 779-791.

Gaut B S, 2014. The complex domestication history of the common bean. *Nature Genetics*, 46(7), 663-664

Greguraš D, 2013. Genetička raznolikost i struktura populacija ljekovite kadulje. Doktorski rad. Zagreb, Sveučilište u Zagrebu.

Jaccoud D, et al. 2001. Diversity Arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acid Research* 29(4).

Kwok P-Y, Duan S, 2003. SNP Discovery by Direct DNA Sequencing. U *Single Nucleotide Polymorphisms Methods and Protocol - Methods in Molecular Biology* 212, Ur. P-Y Kwok, Humana Press Inc. Totowa, New Jersey, pp. 71-85.

LaFramboise T, 2009. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Research* 37(13), 4181-4193. doi: 10.1093/nar/gkp552.

Livak KJ, 2003. SNP Genotyping by the 5'-Nuclease Reaction. U *Single Nucleotide Polymorphisms Methods and Protocol - Methods in Molecular Biology* 212, Ur. P-Y Kwok, Humana Press Inc. Totowa, New Jersey, pp. 129-148.

Mate A, 2009. Genetička raznolikost velebitske degenije. Doktorska disertacija. Zagreb, Sveučilište u Zagrebu.



Moxon R., Wills C. 1999. DNA Microsatellites: Agents of Evolution? Scientific American, 94-99.

Song Q, et al. 2015. SNP Assay Development for Linkage Map Construction, Anchoring Whole-Genome Sequence, and Other Genetic and Genomic Applications in Common Bean. Genes, Genoms, Genetics 5, 2285-2290.

Šatović Z, 1999. Genetski biljezi i njihova uporaba u biljnoj genetici, oplemenjivanju i sjemenarstvu. Sjemenarstvo 16(99)1-2, 73-95.

Vos P, et al. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23(21), 4407-4414.

Wang X, 2016. Next-Generation Sequencing Data Analysis. CRC Press, Taylor & Francis Group, Boca Raton.

<http://beanqual.agr.hr>

<http://data.kew.org/cvalues>

<https://phytozome.jgi.doe.gov/pz/portal.html>

[http://www3.appliedbiosystems.com/cms/groups/mcb\\_marketing/documents/generaldocuments/cms\\_040597.pdf](http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_040597.pdf)

<http://www.diversityarrays.com>

<https://www.illumina.com/systems/sequencing-platforms.html>

<https://www.youtube.com/watch?v=9YxExTSwgPM>

## **5. SAŽETAK**

Cilj ovog završnog seminara je bio prikazati osnovne principa metoda genotipizacije, razmatranih i primjenjivanih prilikom genotipizacije hrvatskih kultivara graha unutar projekta: „Genetska osnova količine bioaktivnih hranivih tvari hrvatskih tradicijskih kultivara graha“, financiranog od strane Hrvatske zaklade za znanost. U seminarskom radu su opisane tradicionalne metode genotipizacije kao što su AFLP, mikrosateliti i SNP biljezi. Budući da se DArTseq metoda pokazale najpogodnijim za dostizanje glavnih ciljeva projekta, osobita pažnja je posvećena metodama temeljenim na sekvenciranju nove generacije (eng. Next Generation Sequencing, skraćena NGS).

## **6. SUMMARY**

The aim of this final seminar was to present the basic principles of the methods of genotyping that were considered and applied during the genotyping of Croatian bean cultivars within the project: "Genetic basis of bioactive nutrient content in Croatian common bean landraces", funded by the Croatian Science Foundation. The traditional methods of genotyping such as AFLP, microsatellite and SNP markers were described. Since the DArTseq method has proved to be the most suitable for reaching the main project objectives, special attention was paid to the Next Generation Sequencing (NGS) methods.