

Téma 2: Analýza stavu biodiverzity živočíšnych genetických zdrojov pomocou genomických údajov

Prednáška

V súvislosti s hodnotením genomických údajov je potrebné v prvom rade v skratke popísať ako takéto informácie môžeme získať. Na analýzu genómu, môžeme využiť rôzne nástroje, vrátane jednotlivých genetických markerov, genotypizačných čipov či celogenómového sekvenovania. Rozdiel medzi uvedenými metódami súvisí jednak so samotným laboratórnym postupom ich stanovenia a súčasne s množstvom údajov o genóme, ktoré prostredníctvom nich získame.

Čo môžeme chápať pod pojmom genetický marker. Jedná sa akýkoľvek charakteristický znak alebo prejav organizmu, ktorý môže byť využitý na určenie špecifického chromozómu, bunky alebo jedinca. Pojmom genetický marker môže byť označený gén, krátky úsek DNA, prípadne iné prejavy genotypu, chromozómov alebo karyotypu. Dôležité však je si zapamätať, že genetický marker je zvyčajne polymorfný variant, ktorý vykazuje mendelistickú dedičnosť a je vo vzťahu s variabilitou fenotypového znaku (vlastnosti), ktorá má význam napríklad z pohľadu šľachtenia.

Z pohľadu produkčných vlastností hospodárskych zvierat sa sledujú jednak kandidátske gény, pretože ich alely a genotypy ovplyvňujú utváranie kvantitatívnych znakov a súčasne aj lokusy pre kvantitatívne vlastnosti.

Výhodou DNA markerov je najmä fakt, že sú priamo detekovateľné v sekvenciách nukleotidov, vykazujú zvýšenú mieru polymorfizmu a dominantnú resp. kodominantnú dedičnosť. DNA markery sa vyskytujú v genóme pomerne často, pričom ich v súčasnosti vieme relatívne ľahko a rýchlo testovať s vysokou mierou reprodukovateľnosti.

Medzi v súčasnosti najčastejšie využívané genetické markery patria jednoznačne jednonukletidové polymorfizmy, označované skratkou SNP. SNP vznikajú zvyčajne bodovou mutáciou, napr. zámenou nukleotidu v DNA v určitom mieste. Oproti iným typom genetických markerov sa v genóme vyskytujú často a to každých 100 až 300 báзовých párov. Za SNP sa zvyčajne považujú mutácie, ktoré sa v danej populácii vyskytujú s frekvenciou vyššou ako 1%, t.j. minoritná alebo menej častá alela je prítomná v genotype minimálne jedného 1% jedincov patriacich do danej populácie. Jedná sa o bialelický marker, t.j. v rámci populácie rozpoznávame pre SNP iba dve alely a to dominantnú a recesívnu. Pojem dominantná vyjadruje, že sa jedná o prevládajúcu alelu u jedincov v danej populácii. Minoritná je naopak alela, ktorá sa v rámci populácie vyskytuje v genotypoch jedincov menej často. Avšak treba si uvedomiť, že aj keď je v jednej populácii alela dominantná v inej z rozdielnym genetickým pôvodom to tak byť nemusí. SNP markery majú naozaj široké využitie a to od hodnotenia biodiverzity až po genomickú selekciu.

Celogenómové sekvenovanie je pojem, ktorý sa využíva na označenie procesu stanovenia presného poradia nukleotidov v reťazci molekuly DNA, teda určenia je primárnej štruktúry. Za klasické metódy možno považovať Maxam-Gilbertovu a Sangerovu, od ktorej sa odvíjajú v súčasnosti využívané metódy sekvenovania novej generácie. V rámci NGS metód existuje viacero platforiem, ktoré sa síce líšia v technologickom prístupe ale prinášajú porovnateľné výstupy.

Aj keď sa cena celogenómového sekvenovania významne znížila v porovnaní s predchádzajúcim obdobím, v prípade ak chceme získať celopopulačné dáta je stále vysoká, hlavne aj sa jedná o sekvenovanie s vysokým pokrytím. V súčasnosti sú v rámci celopopulačných štúdií aj z tohto dôvodu využívané SNP genotypizačné čipy, ktoré umožňujú v rámci nižších nákladov získať informáciu o veľkom počte SNP markerov distribuovaných rovnomerne v genóme. SNP čipy umožňujú genotypovanie od niekoľko tisíc až po 700 000 SNP markerov. Sú dostupné pre väčšinu druhov nielen hospodárskych ale aj spoločenských druhov zvierat. Takto získané informácie možno použiť na rôznorodé účely, vrátane testovania

rodičovstva, stavu genomickej diverzity, celogenómové asociačné štúdie či odhad genomických plemenných hodnôt.

V rámci nasledujúcich snímkov budeme hovoriť o ukazovateľoch, ktoré sa používajú na odhad stavu biodiverzity živočíšnych genetických zdrojov na podklade analýzy genomických dát. Prvým z ukazovateľov je homozygotnosť genómu a genomický inbríding. V súvislosti s homozygotnosťou genómu sa v literatúre často stretnete s dvoma pojmami a to autozygotnosť a homozygotné úseky, označované skratou ROH. Autozygotnosť v podstate reflektuje všetky alely alebo chromozomálne úseky DNA, ktoré sú identické podľa pôvodu, t.j. pochádzajú od spoločného predka. Za homozygotné úseky sa považujú všetky genomické oblasti so špecifickým počtom za sebou nasledujúcich homozygotných genotypov resp. ak hovoríme o testovaní SNP markerov, všetkých homozygotných SNP markerov. Distribúcia, počet a dĺžka homozygotných úsekov závisí od rôznych faktorov, ktoré ovplyvňujú genóm hospodárskych zvierat. Za najvýznamnejšie z tohto pohľadu možno považovať zámernú selekciu a intenzitu príbuzenskej plemenitby. Samotná dĺžka ROH úsekov v genóme jedinca korešponduje so vzdialenosťou predkov v rodokmeni jedinca. V prípade ak majú rodičia jedinca spoločného predka, ich genóm bude v určitých oblastiach zdieľať rovnaké genetické varianty, čiže takíto rodičia budú identickí podľa pôvodu. Ak obidvaja rodičia prenesú na potomka rovnaký úsek, potom bude tento homozygotný pre dané genetické varianty, čím vznikne ROH úsek v jeho genóme. Práve tento predpoklad je základom prístupu pre odhad genomického koeficienta inbrídingu cez pokrytie genómu homozygotnými úsekmi.

Informáciu o výskyte a dĺžke ROH úsekov v genóme však možno využiť nielen na odhad genomického inbrídingu ale aj na testovanie vplyvu zámernej selekcie na špecifické oblasti v genóme či identifikáciu kauzálnych variant podieľajúcich sa na kontrole preferovaných znakov a vlastností.

Na tejto snímke môžete vidieť v prvej časti vzorec pre odhad genomického inbrídingu označovaného ako F_{ROH} , pričom čitateľ vyjadruje celkovú dĺžku homozygotných úsekov v genóme jedinca a menovateľ celkovú dĺžku genómu odvodenú od fyzickej pozície testovaných markerov. F_{ROH} umožňuje stanoviť trend vývoja inbrídingu pričom úseky dlhšie ako 4 Mb reflektujú autozygotné segmenty pochádzajúce od predkov vzdialených približne 12 generácií, úseky dlhšie ako 8 Mb pochádzajú od predkov vzdialených 6 generácií a úseky dlhšie ako 16 Mb zodpovedajú podielu autozygotnosti zdedeného od predkov z posledných 3 generácií. Podobne ako pri rodokmeňovom inbrídingu sa rozpätie hodnôt genomického inbrídingu pohybuje od 0 po 1 resp. v percentuálnom vyjadrení od 0 po 100%. Informácia o prírastku inbrídingu za generáciu a celkovo koeficiente príbuzenskej plemenitby je dôležitá jednak z pohľadu výskytu inbrédnej depresie a súčasne prežiteľnosti populácie z dlhodobého hľadiska. Jedným z dôvodov je fakt, že kumulácia inbrídingu naprieč generáciami vedie k zníženiu genetickej diverzity. Vo všeobecnosti sa uvádza, že prírastok inbrídingu za generáciu by v prípade málopočetných populácií nemal presiahnuť 1% a vo veľkých populáciách 4%. Medzi najčastejšie využívané programy na stanovenie genomického koeficienta inbrídingu patrí detectRUNS, plink či cgaTOH. Na obrázku môžete vidieť výsledky z komparatívnej analýzy koeficienta inbrídingu u 15 plemien dobytka na základe úsekov dlhších ako 4 a 8 Mbp.

Ďalším ukazovateľom biodiverzity, ktorý vieme stanoviť na základe testovania genomických ukazovateľov je väzbová nerovnováha medzi SNP markermi v genóme a následne na jej podklade efektívna veľkosť populácie. Pojem väzbová nerovnováha vyjadruje v podstate nenáhodný vzťah alebo asociáciu medzi alelami rôznych SNP markerov v genóme hodnotenej populácie, ktorá vznikla pravdepodobne v dôsledku selekcie, systémov pripravovania či rekombinácie alebo genetického driftu. Vo výsledku to znamená, že takéto genetické varianty môžu v populácii vytvárať špecifické kombinácie genotypov, ktoré sa nazývajú aj haplotypy. Informáciu o úrovni väzbovej nerovnováhy môžeme využiť na zhodnotenie evolučného utvárania populácií, odhad efektívnej veľkosti, či podobne ako v prípade ROH úsekov na

testovanie výskytu špecifických genetických variant, ktoré podliehali intenzívnemu vplyvu zámernej alebo prírodnej selekcie.

Na snímke je uvedený najčastejšie využívaný vzorec na výpočet väzbovej nerovnováhy medzi SNP markermi. Okrem tohto vzorca navrhnutého Hillom a Robertsonom však existujú aj iné jeho modifikácie, ktoré zohľadňujú napríklad rýchlosť mutácií či povahu testovaných genetických markerov (bialelické alebo viac multialelické). Rozpätie hodnôt sa v prípade väzbovej nerovnováhy pohybuje od 0 po 1, kedy 0 vyjadruje väzbovú rovnováhu medzi markermi a 1 naopak kompletnú väzbovú nerovnováhu.

Efektívna veľkosť populácie v podstate vyjadruje počet jedincov, ktorí sú v danej populácii reprodukčne aktívni, t.j. môžu poskytovať jedincov do ďalšej generácie. Odhad tohto parametra je v prípade genomických dát najčastejšie založený na jeho vzťahu k miere väzbovej nerovnováhy v genóme, pričom je možné testovať nie len súčasnú efektívnu veľkosť aj trend jej vývoja v minulosti.

Efektívnu veľkosť populácie, označovanú skratkou N_e , možno stanoviť napríklad na základe vzorca navrhnutého Corbinom et al., ktorý vidíte na tejto snímke. Tento vzorec zohľadňuje model dedivosti, fyzickú vzdialenosť medzi SNP markermi či intenzitu mutácií. Historická efektívna veľkosť je v tomto prípade odhadovaná ako funkcia času a fyzickej vzdialenosti medzi dvoma markermi za predpokladu konštantného lineárneho rastu N_e s časom vyjadreným minulými generáciami. Na obrázku vpravo vidíte reprezentatívne výsledky analýzy trendu vývoja efektívnej veľkosti u dvoch plemien hovädzieho dobytku, slovenského strakatého a slovenského pinzgauského. Podobne ako pri rodokmeňových informáciách môže efektívna veľkosť populácie nadobúdať hodnoty od 0 po n . Všeobecne je akceptované, že efektívna veľkosť populácie by nemala byť nižšia ako 50 jedincov v prípade málopočetných populácií resp. 100 jedincov v prípade veľkých populácií. Z pohľadu dlhodobej udržateľnosti by mala byť efektívna veľkosť populácie minimálne 500 jedincov. V prípade genomických dát môžeme na výpočet použiť programy ako *SneP* alebo *GONE*.

U živočíšnych genetických zdrojov sú často vyhodnocované aj ukazovatele popisujúce populačnú štruktúru, t.j. ich rôznorodosť na vnútro aj medzipopulačnej úrovni. Z tohto pohľadu sú najčastejšie vyhodnocované genetické vzdialenosti, ktoré vyjadrujú stupeň génovej rozdielnosti medzi jedincami, populáciami alebo druhmi. Najčastejšie sa v literatúre stretávame s Neiovými genetickými vzdialenosťami, Wrightovým fixačným indexom F_{ST} , analýzou hlavných komponentov, či metódami kvantifikujúcimi stupeň genetického premiešania a toku génov medzi populáciami.

Teória Neiovej genetickej vzdialenosti predpokladá, že ak sú dve populácie vykazujúce nízke genetické vzdialenosti podobné, tak s vysokou mierou spoľahlivosti majú spoločných predkov. Z tohto dôvodu možno tento ukazovateľ považovať aj za molekulárny ekvivalent koeficienta príbuznosti počítaného na základe rodokmeňových informácií. Vzorec pre výpočet štandardnej Neiovej genetickej vzdialenosti máte uvedený na snímke. Minimálna hodnota, ktorá môže Neiova genetická vzdialenosť nadobúdať je 0. Táto hodnota znamená, že jedince alebo populácie majú v genóme rovnaké varianty (alely alebo genotypy), t.j. sú geneticky identické. Maximálna hodnota, ktorú môžu nadobúdať Neiova genetická vzdialenosť je 1. Táto hodnota vyjadruje fakt, že v dôsledku úplne rozdielných genetických variant sú jedince alebo populácie geneticky rozdielne, môžeme povedať aj nepríbuzné. Na výpočet Neiových genetických vzdialeností môžeme využiť balíku programu *R StAMPP*, *Poppr* alebo iné programy.

V porovnaní s Neiovou genetickou vzdialenosťou umožňuje Wrightov fixačný index F_{ST} stanoviť mieru rôznorodosti iba na populačnej úrovni. Tento index je v podstate indikátorom intenzity fragmentácie populácie vyjadrenej poklesom heterozygotnosti v subpopuláciách v dôsledku pôsobenia genetického driftu. Z tohto vyplýva, že ak chceme tento index stanoviť musíme mať k dispozícii informáciu o očakávanej heterozygotnosti v rámci metapopulácie a priemernej heterozygotnosti v rámci subpopulácií ako môžete vidieť vo vzorci na snímke.

Wrightov fixačný index F_{ST} nadobúda hodnoty od 0 po 1, pričom interpretácia hodnôt je podobná ako v prípade Neiových genetických vzdialeností. To znamená, že ak je hodnota indexu rovná 0 populácie sú geneticky identické a naopak ak je hodnota rovná 1 populácie sú geneticky rozdielne. V reálnych populáciách hospodárskych zvierat sa hodnota tohto indexu pohybuje zvyčajne od 0 po 0,5, samozrejme ak testujeme jeden druh. Za geneticky dostatočne diferencované alebo inak povedané odlišiteľné považujeme populácie s hodnotou F_{ST} vyššou 0,25.

Medzi ďalšie často používané prístupy na vyhodnotenie populačnej štruktúry a genetických vzťahov medzi populáciami patria analýza hlavných komponentov a Bayesiánska analýza genetického premiešania alebo inak povedané admixie. Analýza hlavných komponentov je populárna viacrozmerná štatistická metóda, ktorá našla využitie v rôznych vedných odvetviach, vrátane populačnej genetiky. Zjednodušene možno povedať, že táto analýza slúži na znázornenie nadrozmerných dát, napr. genomických informácií o jedincoch alebo populáciách, v menšom počte dimenzií. Bayesiánska štatistika je rovnako metódou, ktorá sa využíva aj v iných vedných odboroch. Táto štatistika pracuje s podmienenou pravdepodobnosťou a umožňuje spresniť pravdepodobnosť východiskovej hypotézy v slede ako sa objavujú ďalšie relevantné skutočnosti.

Na tejto snímke sú znázornené reprezentatívne výsledky testovania podielu genetického premiešania, analýzy hlavných komponentov a toku génov. V prípade prvého obrázku sa jedná o Bayesiankú analýzu admixie v rámci 15 plemien hovädzieho dobytku, pričom podiel premiešania v rámci plemien je znázornený rozdielnymi farbami čiar. Na druhom obrázku vľavo sú zasa uvedené reprezentatívne výsledky analýzy hlavných komponentov, pričom miera premiešania je najlepšie vidieť v časti D prostredníctvom prekrývajúcich sa peak-ov rozdielných farieb. Na treťom obrázku vpravo vydáme opätovne výsledky analýzy genetického premiešania 4 plemien a nižšie číselné vyjadrenie toku génov medzi genofondami populácií.

Ako som už spomínala pri ROH úsekoch a väzbovej nerovnováhe, okrem štandardných ukazovateľov ako N_e a F , vieme hodnotiť aj vplyv selekcie na štruktúru resp. konkrétne genetické varianty v špecifických oblastiach v genóme populácií. Tento prístup na rozdiel od celogenómových asociačných štúdií nevyžaduje prístup k fenotypovým informáciám o jedincoch. Jedná sa v podstate o identifikáciu tzv. selekčných signálov, ktorých výskyt závisí od rôznych faktorov, u hospodárskych zvierat hlavne zámernej selekcie. Na tento účel sa využívajú v podstate dve skupiny metód a to metódy testujúce rozdiely medzi populáciami resp. plemenami a metódy analyzujúce vnútro populačné rozdiely.

Z pohľadu medzipopulačných rozdielov sa na identifikáciu selekčných signálov najčastejšie využíva celogenómový skríning fixačného indexu F_{ST} , analýza variability vo väzbovej nerovnováhe či výpočet integrovaného skóre haplotypov. Na tento účel existuje množstvo programov, ako napríklad PLINK, varLD či balík programu R rehh. Na obrázku vpravo je výsledok analýzy testovania selekčných signálov, ktoré reflektujú rozdiely medzi slovenským strakatým a slovenským pinzgauským dobytkom. Najsilnejšie selekčné signály boli zistené v rodine kazeínových génov a génov KIT a KDR zodpovedných za strakatosť.

V rámci vnútro populačných rozdielov sa zvyčajne selekčné signály stanovujú na základe distribúcie homozygotných úsekov či variability vo väzbovej nerovnováhe. Na výpočet možno použiť rovnaké programy ako v prípade predchádzajúcich metód. Na obrázku vpravo sú znázornené výsledky testovania homozygotných úsekov v genóme plemien slovenského straktého a slovenského pinzgauského dobytku. Z výsledkov je zrejmé, že podobne ako pri predchádzajúcom prístupe sú selekčné signály najsilnejšie v oblasti génov kazeínovej rodiny.

V prípade ďalších otázok k prezentácii ma prosím kontaktujte na emailovej adrese uvedenej na poslednej snímke dole. Informácie o projekte vrátane prístupu k ďalším prezentáciám nájdete po naskenovaní čiarového kódu vľavo na rovnakej snímke. Ďakujem za pozornosť.