

# Manuál pro program

# *PSP*

Populations Study - Pedigree

## Program pro stanovení genetické variability populací na základě rodokmenových dat

verze 0.95

2018

**Luboš Vostrý**

Česká zemědělská univerzita v Praze  
Kamýcká 129,  
165 00 Praha 6  
Česká republika

**Hana Vostrá-Vydrová, Emil Krupa**

Výzkumný ústav živočišné výroby (Institute of Animal Science)  
Přátelství 815  
104 00 Praha – Uhřetěves  
Česká republika

Vypracováno za podpory projektu Mze - Celostátní informační systém genetického hodnocení  
hospodářských zvířat (QJ1510141)

# Úvod

Analýza rodokmenových údajů je jednou z možností popisu genetické rozmanitosti a jejího vývoje během generací (Boichard et al., 1997). Metoda je založená na vztahu mezi přírůstkem hodnoty koeficientu příbuzenské plemenitby a poklesem heterozygotnosti určitého lokusu v dané populaci (Kadlečík, 2008). Analýza rodokmenu podobně jako molekulárně genetické analýzy na základě mikrosatelitů či SNP (Single Nucleotide Polymorphism) je možné použít ke studiu genetické proměnlivosti a genetické rozmanitosti uvnitř i mezi rozdílnými plemeny hospodářských zvířat. Program PSP je určen právě ke studiu genetické rozmanitosti na základě rodokmenových dat.

## Systémové prostředí

Program *PSP* je modulární program dostupný ve formě zdrojového kódu programovacího jazyku FORTRAN 90. Zdrojový kód je možné zkompileovat běžným kompiláterem pod licenci GNU (gfortran) pod systémy Linux a Windows (32 a 64 bit) viz kapitola instalace.

## Postup instalace

Po získání zdrojových kódů v balíku `psp.tar.gz` je nutné balík se zdrojovými kódy a ukázkovými soubory rozbalit. Ukázkový postup je uveden v prostředí LINUX (UNIX) a pomocí GNU Fortran kompilátoru GNU (gfortran) a jiné Fortran kompilery (např. ifort aj.), ale obdobný způsob je možný i pro prostředí WINDOWS .

Rozbalení se provede příkazem

```
tar -xvzf psp.tar.gz
```

Po rozbalení balíku je nutné program zkompileovat. Kompilaci provedeme příkazy:

```
cd ./Psp
```

```
make
```

Po zkompileování programu vznikne spustitelný program: **psp**. Tento program je již možné zkopírovat do adresáře s rodokmeny, či do adresáře `/usr/local/bin`.

Spuštění programu

Pokud byl soubor zkopírován do adresáře s rodokmeny, spuštění programu provedeme příkazem:

```
./psp
```

Zdrojové kódy souboru jsou uvedeny v adresáři *./psp/libs* .

Metodický postup

Program psp odhaduje následující parametry populace:

*Demografické parametry*

**Generační interval** je definován, jako průměrný věk rodičů při narození potomka, který byl zařazen do reprodukce. Generační interval zahrnuje čtyři reálné cesty genetického přenosu: od otce na syna, od matky na dceru, od matky na syna a od otce na dceru.

### **Hodnocení kompletnosti rodokmenu**

Kompletnost rodokmenu a počet rodičovských generací ovlivňuje odhady koeficientu příbuzenské plemenitby a koeficientu příbuznosti mezi jedinci. Úroveň kompletnosti rodokmenu s využitím indexu kompletnosti rodokmenu (*PCI*), který byl hodnocena jako podíl známých předků v každé generaci pro každého jedince (MacCluer et al., 1983). Celkový index kompletnosti rodokmenu je vyjádřen jako harmonický průměr rodičovských a mateřských indexů:

$$PIC_{jedince} = \frac{4C_{otec} C_{matka}}{C_{otec} + C_{matka}},$$

kde  $C_{otec}$  a  $C_{matka}$  jsou příspěvky z otcovské a mateřské strany:

$$C = \frac{1}{d} \sum_{i=1}^d a_i,$$

kde  $a_i$  je podíl známých předků v generaci  $i$  a  $d$  je počet generací předků, které jsou brány v úvahu. Ekvivalent kompletních generací předků ( $t$ ) v rodokmenu byl odhadnut pomocí vztahu  $\Sigma(1/2)^n$  (Maignel et al., 1996), kde  $n$  je počet generací mezi jedincem a jeho předkem.

*Pravděpodobnost originálních genů*

Genetická variabilita byla hodnocena z rodokmenových hodnot pomocí následujících parametrů. **Celkový počet zakladatelů** byl stanoven jako celkový počet přetků s neznámými rodiči. **Efektivní počet zakladatelů** byl definován jako počet zakladatelů, kteří vysvětlují shodnou úroveň genetické proměnlivosti jako je pozorována u referenční populace (Lacy, 1989). Tento parametr je možné také vysvětlit jako pravděpodobnost, že dva náhodně vybrané geny z analyzované populace pocházejí od samého zakladatele (James, 1972). Efektivní počet zakladatelů byl odhadnut pomocí vztahu:

$$f_e = \frac{1}{\sum_{i=1}^z q_i^2},$$

kde  $q_i$  je genetický příspěvek  $i$ -tého zakladatele do referenční populace.

**Efektivní počet předků** ( $f_a$ ) je definovaný jako minimální počet předků, né nezbytně zakladatelů, kteří vysvětlují stejnou genetickou diverzitu jako je v referenční populaci (Boichard *et al.*, 1997). Efektivní počet předků byl získán po stanovení marginálního genetického příspěvku každého předka. Parametr  $f_a$  představuje ztrátu genetické variability způsobené efektem hrdla láhve (bottleneck) způsobený nerovnoměrným příspěvkem mezi reprodukcujícími se jedinci do následujících generací. Efektivní počet předků byl odhadnut pomocí vztahu:

$$f_a = \frac{1}{\sum_{i=1}^{an} p_i^2} ,$$

kde  $p_i$  je marginální genetický příspěvek předka  $i$ , a  $an$  je celkový počet předků

**Ekvivalent počtu genotypů zakladatelů** ( $f_{ge}$ ) je definována jako efektivní počet předků s nenáhodnou ztrátou genů zakladatelů kteří vysvětlují shodnou genetickou variabilitu jako je v referenční populaci (Lacy, 1989). Ekvivalent genotypů zakladatelů představuje ztrátu genetické variability způsobenou jak nerovnoměrným příspěvkem zakladatelů a náhodnou ztrátou alel způsobenou genetickým driftem (Lacy, 1989). Ekvivalent genotypu předků byl odhadnut pomocí vztahu dle Caballero a Toro (2000):

$$f_{ge} = \frac{1}{2f_g} ,$$

kde  $f_g$  je průměrný koeficient příbuznosti v referenční populaci. Dle Caballero a Toro (2000) je možné odhadnout průměrný koeficient příbuznosti ze vztahu:

$$f_g = \frac{\sum_{i=1}^n \sum_{j=1}^n a_{ij}}{2n^2} ,$$

kde  $a_{ij}$  představuje prvky matice příbuznosti a  $n$  je počet jedinců v rodokmenu.

### **Koeficient inbreedingu ( $F_i$ ) a koeficient příbuznosti ( $AR$ )**

**Koeficient inbreedingu:** Koeficient inbreedingu každého jedince byl odhadnut s využitím tabelární metody (Falconer a Mackay, 1996) se zohledněním metody dle VanRadena (1992), která umožňuje zohlednit chybějící jedince v rodokmenu:

$$F_i = a_{ii} - 1$$

kde  $F_i$  je koeficient inbreedingu jedince  $i$ , a  $a_{ii}$  je aditivně genetická příbuznost mezi jedincem  $i$  sama na sebe.

Průměrný koeficient příbuznosti pro každého jedince ( $AR$ ) byl odhadnut jako průměrný koeficient zahrnující průměr z řádků matice příbuznosti  $A$  příslušící danému jedinci. Tento koeficient představuje pravděpodobnost, že náhodně vybraná alela se vyskytuje ve vybraném jedinci nebo uvnitř skupiny jedinců (Goyache *et al.*, 2003).

**Realizovaná efektivní velikost populace na základě nárůstu koeficientu příbuzenské plemenitby ( $N_{eF}$ )** představuje počet nepříbuzných jedinců, kteří mohou způsobit shodný nárůst hodnoty koeficientu inbreedingu jako je dosahován v referenční populaci. Realizovaná efektivní velikost populace ( $N_{eF}$ ) byla odhadnuta na základě nárůstu koeficientu inbreedingu jedince  $\Delta F_i$  (Gutierrez *et al.*, 2009), s ohledem na  $\Delta F_i = 1 - t_i^{-1} \sqrt{1 - F_i}$ , kde  $F_i$  je koeficient inbreedingu jedince  $i$  a  $t_i$  je ekvivalent kompletnosti generací předků jedince  $i$ . Realizovaná efektivní velikost populace byla odhadnuta s využitím vztahu (Cervantes *et al.*, 2008):

$$N_{eF} = \frac{1}{2\overline{\Delta F}},$$

kde  $\overline{\Delta F}$  je průměrný nárůst koeficientu inbreedingu v referenční populaci.

**Realizovaná efektivní populace založená na nárůstu původového koeficientu ( $N_{eC}$ ):**

$$N_{eC} = \frac{1}{2\overline{\Delta C_{ij}}}, \text{ kde } \Delta C_{ij} = 1 - \frac{t_i + t_j}{2} \sqrt{1 - C_{ij}},$$

kde  $t_i$  a  $t_j$  jsou ekvivalenty kompletnosti generací (Maignel *et al.*, 1996) jedince  $i$  a  $j$ ,  $\Delta C_{jk}$  je nárůst původového koeficientu mezi dvěma jedinci  $j$  a  $k$  a  $C_{jk}$  je hodnota koeficientu příbuzenské plemenitby potomka jedinců  $j$  a  $k$  (Cervantes *et al.*, 2010).

**Ztráta genetické diverzity** v referenční populaci v důsledku genetického driftu, nebo nerovnoměrného příspěvku předků byl odvozen z efektivního počtu zakladatelů ( $f_e$ ) a ekvivalentu genotypu zakladatelů ( $f_{ge}$ ). Celková ztráta genetické diverzity způsobená genetickým driftem nebo efektem hrdla láhve byl odhadnut pomocí vztahu (Lacy, 1995):

$$1 - GD$$

$$\text{kde } GD = 1 - \frac{1}{2f_{ge}}.$$

Ztráta genetické diverzity způsobená nerovnoměrným příspěvkem zakladatelů způsobený genetickým driftem byl odhadnut pomocí vztahu:

$$1 - GD^*,$$

$$\text{kde } GD^* = 1 - \frac{1}{2f_e},$$

přičemž rozdíl mezi  $GD$  a  $GD^*$  představuje ztrátu genetické diverzity způsobené genetickým driftem v generacích jedinců, které nejsou považováni za zakladatele a odpovídají převrácené hodnotě vztahu  $2N_{enf}$ :

$$GD - GD^* = \frac{1}{2N_{enf}} .$$

## Datové soubory

Program pracuje s jedním vstupním datovým souborem obsahující rodokmenové záznamy a jedním parametrovým souborem:

### Rodokmenový soubor

V tomto souboru jsou uvedeny rodokmenové záznamy studované populace.

Pro tento soubor jsou nutná následující data.

Sloupec identifikující jedince,

Sloupec identifikující otce jedince

Sloupec identifikující matku jedince

Sloupec identifikující generaci jedince

Sloupec identifikující rok narození jedince

Sloupec indentifikující pohlví jedince

Sloupec identifiující (sub)populaci – např. plemeno)

Sloupec identifiující referenční populace – jedinci, kteří se mohou podílet na tvorbě následné generace.

Pro správný běh programu je nutné aby jedinci v rodokmenu byly uváděny v numerické hodnotě a sestupně seříděny a překódovány sestupně od nejstaršího jedince, který nese kód 1, po jedince nejmladšího (n). Neznámí jedinci jsou označeni kódem záporného čísla (záporná hodnota představuje číslo generace v které se jedinec zahrnut do rodokmenu ( napřů -3). Rodokmenový soubor je požadován ve volném formátu (oddělovač mezera) se vstupními hodnotami integer.

### Parametrový soubor

V tomto souboru jsou uloženy všechny vstupní parametry potřebné k správnému běhu programu:

Ukázka prametrového souboru (vše za # je komentář):

# Jméno rodokmenového souboru

PEDFILE

rod\_psp.txt

# Pozice sloupce ve kterém se nachází jedinci

AID

1

# Pozice sloupce ve kterém se nachází otec

SID

2

# Pozice sloupce ve kterém se nachází matka

DID

3

# Pozice sloupce ve kterém se nachází generace jedince

GEN

4

# Pozice sloupce ve kterém se nachází rok narození jedince

YOB

5

# Pozice sloupce ve kterém se nachází kód plemene

BREED

6

# Pozice sloupce ve kterém se nachází kód pohlaví (samec=1, samice =2)

SEX

7

# Position of the reference population declaration column in the data set

STAT

8

#

# Počet jedinců v rodokmenu

N\_AID

29749

# počet iterací – požadováno pro odhad koeficientu příbuzenské plemenitby

N\_ITER

10

# Kód plemene, pro který má být studie uskutečněna

SUB\_CODES

86

**Ukázkové soubory jsou** uvedeny v adresáři ./PedOrder/example :

Rodokmenový soubor: rod.txt

Parametrový soubor: Parametry.txt

### **Popis programu *psp***

Po spuštění programu jsou vygenerovány následující soubory:

Soubor inbreeding.txt – obsahuje hodnoty koeficientu příbuzenské plemenitby pro jednotlivé jedince.

Soubor average\_Fx.txt – obsahuje průměrné hodnoty koeficientu příbuzenské plemenitby pro jedince dle roku narození.

Soubor GD\_loss.txt – obsahuje kvantifikaci ztráty genetické diverzity.

Soubor ngen.res – obsahuje hodnoty ekvivalentu kompletnosti generací a hodnoty Indexu kompletnosti rodokmenu.

Soubor result1.lst – obsahuje jedince, kteří vysvětlují největší podíl genetické proměnlivosti dané populace.

Soubor results.lst – obsahuje základní charakteristiky populace:

- generační interval: - otec – syn, otec – dcera, matky – syn, matka – dcera, celkový,
- průměrný inbreeding za referenční populaci,
- průměrný původový koeficient za referenční populaci,
- průměrný koeficient příbuznosti za referenční populaci,
- směrodatnou odchylku koeficientu příbuznosti za referenční populaci,



- realizovanou efektivní velikost populace na základě nárůstu koef. příbuzenské plemenitby z generace na generaci,
- realizovanou efektivní velikost populace na základě nárůstu koef. příbuznosti z generace na generaci,
- efektivní velikost populace založený na poměru pohlaví,
- průměrný nárůst koeficientu příbuzenské plemenitby z generace na generaci,
- počet samců v referenční populaci,
- počet samic v referenční populaci,
- počet zakladatelů,
- efektivní počet zakladatelů,
- efektivní počet předků.

### **Použitá literatura**

- Boichard, D., Maignel, L., Verrier, E. 1997. The value using probabilities of gene origin to measure genetic variability in a populations. *Genetic Selection Evolution*. 29. 5-23.
- Caballero, A., Toro, M. A. 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetical Research*. 75. 331-343.
- Cervantes, I., Goyache, F., Molina, A., Valera, M., Gutiérrez, J. P. 2008. Application of individual increase in inbreeding to estimate effective size from real pedigrees. *Journal of Animal Breeding and Genetics*. 125. 301–310.
- Falconer, D. S., Mackay, T. F. C. 1996. *Introduction into quantitative genetics*. Longman House, Harlow Essex., p. 464. ISBN: 978-0582243026.
- Goyache, F., Gutiérrez, J. P., Fernández, I., Gómez, E., Álvarez, I., Díez, J., Royo, L. J. 2003. Using pedigree information to monitor genetic variability of endangered populations: the Xalda sheep breed of Asturias as an example. *Journal of Animal Breeding and Genetics*. 120. 95-103.
- Gutierrez J.P., Cervantes I., Molina A., Valera M., Goyache F. (2008): Individual increase in inbreeding allows estimating effective size from pedigrees. *Genetics Selection Evolution*, 40,

359–378.

James J.W. (1972): Computation of genetic contribution from Pedigrees. *Theoretical and Applied Genetics*, 42, 272-273.

Lacy, R. C. 1989. Analysis of founder representation in pedigrees: Founder Equivalent and Founder Genome Equivalents. *Zoo Biology*. 8. 111-123.

Lacy, R. C. 1995. Clarification of genetic terms and their use in the management of captive populations. *Zoo Biology*. 14. 565-578.

MacCluer J., Boyce B., Dyke L., Weitzkamp D., Pfenning A., Parsons C. (1983): Inbreeding and pedigree structure in Standardbred horses. *Journal of Heredity*, 74, 394–399.

Maignel, L., Boichard, D., Verrier, E. 1996. Genetic variability of French dairy breeds estimated from pedigree information. *Interbull Bull*, 14. 49-54.

VanRaden, P. M. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. *Journal of Dairy Science*. 75. 3136-3144.