MICROSATELLITE MARKERS FOR PEDIGREE VERIFICATION IN CATTLE

Nikola Adamov, Ljupče Mickov, Vladimir Petkov, Mihajlo Adamov

Faculty of Veterinary Medicine,
Lazar Pop Trajkov 5-7, 1000 Skopje, Republic of Macedonia
adamovn@gmail.com

In order to achieve efficient implementation of the breeding programs that include selective reproduction, correct pedigree information and relationships between the animals is the basic condition. Breeding strategies in domesticated livestock are based on animal model evaluations, which include all known genetic relationships between the animals in the calculations. However, as demonstrated in some studies, the main assumption of this model for achieving unbiased evaluations – that all pedigrees and relationships are correctly recorded, is not always fulfilled. Misidentified animals are expected to bias the estimation of genetic parameters and breeding values and can lead to loss in selection response and reduced genetic progress as they are a function of the number of misidentified progeny. Errors in pedigrees could also result in incorrect decisions about selective mating of the animals, which could increase the inbreeding depression. In the past, parentage/paternity testing in cattle has been carried out through the blood group and the protein polymorphism analysis, but because of some drawbacks, these tests have been replaced with new ones that are based on detection of certain "genetic markers". Most informative and most commonly used are the microsatellite markers (Short Tandem Repeats) which are highly polymorphic and are located on the noncoding intron regions of the bovine genome. The advantage of microsatellite based tests is that theoretically any sample containing nuclear DNA can be used for analysis, and when genotyping recommended set of markers, the accuracy of the test is much higher as the probability of detecting mistaken parentage is a direct function of the polymorphism of the markers used. The research of the cattle genome conducted in the past resulted in identification of several thousand microsatellite loci, among which nine most informative are recognized by the International Society for Animal Genetics (ISAG) as "international marker set" and are recommended as a minimal panel to be included in cattle pedigree verification in order to facilitate the record exchange between laboratories.

Key words: bovine microsatellite markers; parentage/paternity testing; pedigree verification

МИКРОСАТЕЛИТСКИ МАРКЕРИ ЗА ВЕРИФИКАЦИЈА НА ПЕДИГРЕТО КАЈ ГОВЕДАТА

За постигнување ефикасна имплементација на одгледувачките програми кои вклучуваат селективно размножување, неопходно е точно евидентирање на педигреата, односно на родинските врски меѓу животните. Програмите за одгледување говеда се засновани врз евалуација на приплодната вредност на животните преку моделот на индивидуа (animal-model), кој во своите кalkулации ги користи сите достапни информации за генетските сродства меѓу животните вклучени во евалуацијата. Меѓутоа, како што е потврдено и со неколку студии, главниот предсав за постигнување непристрасна евалуација со овој модел – педигреата на сите животни да бидат точно евидентирани, не секогаш е исполит. Грешките во педигреата можат да резултираат со необектива претпредизија на генетските параметри и на одгледувачките вредности, послаб селекциони одговор и забавување на генетскиот прогрес, чиишто интензитет ќе зависи правопропорционално од бројот на потешно евидентираниот гращ. Погрешноста евиденција на предци може исто така да доведе и до погрешни одлуки за селективното парење на животните, со што би се придонело за зголемување на инбридинг-депресијата. Во минатото, верификацијата на педигреето кај говедата се вршеше преку анализа на крвните групи и на протеинскиот полиморфизам, но поради некои недостатоци на овие тестови, тие денес се заменети со нови кои се базираат на детекција на одредени генетски маркери". Накорисни и најчесто употребувани се микросателитските маркери (Short Tandem Repeats) кои се високо полиморфни и кои се лоцирани на интрооизните некодирачки секвенции од геномот. Предноста на овие тестови е во тоа што теоретски секоја мостра која содржи нуклеарна ДНК може да биде употребена за анализа, а при генотипизирање на препорачан сет марке-
ri verodostojnosti na testot e znachenito visoka, bidejki verojatnost za deteksija na pogrešno zaveden roditelj e direktna funkcija od polimorfnosti na upotrebene markeri. Istragašuwa na taj vo izminatite godini bе vreshhen na genomot na govedata rezultira so definirowane na nekolku izjadi mikrosatelitski lokusi, od kon deveteto najinformativni se priznani od Međunarodnog združenje za animalna genetika (ISAG) kako „interнационални set на маркери“ i se preporčani kako minimalen panel vo testovite za verifikacija na pedigrea so cе polese spojeduvane na rezultatite dobveni od razlichen laboratorini.

Ключни зборови: бовински микросателитски маркери; тест за родителство/татковство; верификација на pedigrea

1. INTRODUCTION

To achieve efficient implementation of the breeding programs that include selective reproduction it is necessary to have errorless pedigree information and correct relationships between the animals in the population. Contemporary cattle breeding programs are based on animal model evaluation techniques which account for all assumed genetic relationships between the animals in the calculation. However, as demonstrated in some studies, the main assumption for achieving unbiased evaluations with this model – that all of the pedigrees and relationships are correctly recorded, is not always fulfilled.

The proportion of errors in cattle pedigrees varies in different countries: 5 – 15% in Denmark (Christensen et al., 1982), 4 – 23% in Germany (Geldermann et al., 1986), 8 – 20% in Ireland (Beechinor and Kelly, 1987), 12% in Netherlands (Bovenhuis and Van Arendonk, 1991), 2,9 – 5,2% (Ron et al., 1996) or 11,7% (Weller et al., 2004) in Israel, 10% in dairy cattle in the United Kingdom (Visscher et al., 2002) and 10,7% in the Czech Republic (Rehout et al., 2006).

Christensen et al. (1982) summarize seven reasons for errors in paternity recording: 1) mistakes by AI institutes in semen labeling; 2) incorrect identification of semen straws by AI technicians; 3) insemination of cows already pregnant by a previous insemination; 4) incorrect entry of the bull’s herd book name or number into the insemination record; 5) use of natural – service bulls leading to pregnancies of previously inseminated cows which were assumed to be pregnant from the AI bull; 6) mistakes in sire identification when a cow enters a milking herd in schemes where pedigree information on milk recorded cow is obtained through the milk recording program, and 7) interchange of calves on the same farm (changeling). To these reasons Weller et al. (2004) added the factor 8) genotyping errors, mutations or occurrence of so called “null alleles” which can lead to rejection of otherwise correct pedigree.

The term “null allele” refers to those that fail to amplify during the PCR reaction and can arise as a result of: indel or point mutations in the primer annealing site, preferential amplification of short alleles (due to inconsistent DNA template quality or quantity), or slippage during the PCR amplification. In such cases although an animal is heterozygous for a specific locus the test will show it as homozygous. The progeny of a sire heterozygous for a null allele that receive the null allele will appear to be homozygous for their maternal allele. If this allele differs from the father’s detected allele, then incompatibility for this locus will be observed between the progeny and the sire and the paternity will be incorrectly rejected.

Pedigree errors are expected to bias the estimation of genetic parameters (Van Vleck, 1970), breeding values (Israel and Weller, 2000), selection response and genetic progress (Gelderman et al., 1986). The loss in response should be similar to the proportion of progeny misidentified (Gelderman et al., 1986). Israel and Weller (2000) presented 3-4% loss in selection response via the stochastic simulation study of a large dairy cattle population with 10% incorrect paternity. Also pedigree errors could lead to incorrect decisions about selective mating of the animals which would contribute to increasing the inbreeding depression in a cattle population.

2. GENETIC MARKERS FOR PARENTAGE TESTING IN CATTLE

In the past pedigree verification in dairy cattle has been carried out using blood group and protein polymorphism. These tests have shown their usefulness but their major drawback is the necessity of obtaining blood samples from animals. Moreover usefulness of blood groups in parentage
testing is somewhat questionable due to limited variability of some blood groups in some populations of cattle.

Contemporary tests for parentage verification are based on detection of certain “genetic markers”. Any attribute that can be easily detected and its inheritance traced can serve as a marker. The advantage of these tests is that theoretically any sample containing diploid nuclear DNA can be used for analysis which renders sampling significantly easier, faster and non-invasive, as well as higher reliability than the traditional blood testing. For example Ron et al. (2003) developed and tested a method to sample cows using vaginal swabs as a DNA source for paternity testing.

These tests are based on two main principles:

- on detection of genetically inherited markers that remain the same through the animal’s life (if mutations are excluded), and
- on the fact that one animal can possess only two alleles of every locus (or marker), one of which was inherited from the sire and the other from the dam.

The most useful and most widely accepted are the microsatellite markers mainly because of their ease of use and analysis, and their high informative value provided by the large number of alleles per locus (Baumung et al., 2004), even though recently the SNP (Single Nucleotide Polymorphism) markers have gained high popularity because of their advantages such as high throughput automated analysis and genetic stability in mammals. Recently, two different SNP marker sets were reported for animal identification and parentage testing in American beef cattle (Heaton et al., 2002) and European dairy breeds (Werner et al., 2004). Having into consideration that SNP markers are only bi-allelic, several papers (reviewed by Morin et al., 2004) predict that at least two to six times more SNP markers will be necessary to achieve the same resolution as microsatellites when used for animal identification and parentage assessment.

Microsatellite markers also called STRs (Short Tandem Repeats) or SSRs (Simple Sequence Repeats) are short segments of DNA that contain 1–6 bp repeat nucleotide motifs such as (CAG)n, and they tend to occur in non-coding regions of the mammalian genome. These regions in a population of animals are highly polymorphic, i.e. they are present in several different alleles, and this is the basis for their usefulness in animal identification and parentage testing. The accuracy of these tests is high because the probability of detecting mistaken parentage is a function of the polymorphism of the genotyped loci (Ron et al., 1996).

Not every microsatellite is informative enough to be used in parentage investigations. To determine the level of usefulness of a microsatellite for parentage testing, the number and frequency of its alleles in the investigated population must be evaluated. With this information it is possible to calculate the heterozygosis and the degree of polymorphism which in turn enables the determination of exclusion probability of each individual marker as well as of the entire panel of markers as a group.

In order to use a microsatellite marker in parentage investigation it has to be checked for consistency with the Hardy-Weinberg expectations by comparing the observed and expected genotype frequencies. The observed heterozygosity is defined as the number of heterozygotes divided by the sample size and the

1) **Expected heterozygosity** is the frequency of the heterozygous genotypes assuming the population is in HW equilibrium. It is calculated from the allele frequencies by the following formula:

\[
H_{\text{exp}} = 1 - \sum_{i=1}^{n} p_i^2
\]

(Nei, 1978)

where \( p_i \) is the frequency of \( i \)-th allele, and \( n \) is the total number of alleles on certain locus.

2) **Polymorphism information content – PIC** expresses the informative value of a marker as a result of its polymorphism:

\[
PIC = H_{\text{exp}} - 2 \sum_{i=1}^{n-1} p_i^2 \times \sum_{j=i+1}^{n} p_j^2
\]

(Botstein et al., 1980)

with \( p_i \) and \( p_j \) being the frequencies of alleles \( i \) and \( j \), and total number of alleles \( n \).

In parentage testing the usefulness of any codominant marker is defined by its 3) **Exclusion probability – EP** (Jamieson & Taylor, 1997). Exclusion probability indicates the possibility of excluding as progenitor an animal assigned as such incorrectly, i.e. it expresses the probability that two random (unrelated) animals do not share any
identical allele. It is calculated as powers of allele frequencies for every marker separately, as well as for the entire panel of markers together. The three most common situations are:

3.1. Given two parents and one offspring; exclude a parent. An example of this is a familiar paternity case but it can be applied for maternity testing as well:

\[ EP = 1 - 2\sum_{i=1}^{n} p_i^2 + \sum_{i=1}^{n} p_i^3 + 2\sum_{i=1}^{n} p_i^4 - 3\sum_{i=1}^{n} p_i^5 - 2(\sum_{i=1}^{n} p_i^2)^2 + 3\sum_{i=1}^{n} p_i^2 \times \sum_{i=1}^{n} p_i^3 \]  (Jamieson, 1994)

3.2. Given two parents and one offspring; exclude both parents. An example of this is changing or substituted offspring:

\[ EP = 1 + 4\sum_{i=1}^{n} p_i^4 - 4\sum_{i=1}^{n} p_i^5 - 3\sum_{i=1}^{n} p_i^6 - 8(\sum_{i=1}^{n} p_i^2)^2 + 8\sum_{i=1}^{n} p_i^2 \times \sum_{i=1}^{n} p_i^3 + 2(\sum_{i=1}^{n} p_i^3)^2 \]  (Jamieson & Taylor, 1997)

3.3. Given one parent and one offspring; exclude their relationship. An example of this is when other parental genotype is for some reason unavailable:

\[ EP = 1 - 4\sum_{i=1}^{n} p_i^2 + 2(\sum_{i=1}^{n} p_i^2)^2 + 4\sum_{i=1}^{n} p_i^3 - 3\sum_{i=1}^{n} p_i^4 \]  (Jamieson & Taylor, 1997)

where \( p_i \) is the frequency of allele \( i \), and \( n \) is the total number of alleles of the locus.

In any of the above situations a number (\( n \)) of loci may be required to achieve high exclusion rates. Their Combined exclusion probability is calculated as:

\[ EP_c = 1 - (1 - EP_1) \times (1 - EP_2) \times (1 - EP_3) \ldots (1 - EP_n) \]

The values of the exclusion probabilities depend directly on the allelic frequency distribution of each marker within the studied population. Microsatellite based parentage testing works by the method of elimination or rejection since no test can assign a parent to an offspring with 100% accuracy. In other words with these tests we can state with certain degree of certainty that some animal is not a biological parent to an offspring and reject the relatedness, or assign the parentage to an animal for which we observe no discrepancies with the offspring’s genotype. When using a recommended set of 10–12 markers a combined exclusion probability of more than 0.998 is achieved. The accuracy of the test is directly dependent on the number and the polymorphism of the analyzed loci.

In a population of animals the variability of the microsatellite loci arises as a consequence of slippage mutations in these regions which enables formation of new alleles that differ in the number of repetitions. Because of this it is recommended that a parentage should be rejected if differences between the putative parent and the offspring are found on at least two unrelated loci. The discrepancy on one locus could be a consequence of mutation, a genotyping mistake or presence of a null allele. The observed frequency of mutation of short tandem repeats was 0.001/locus/gamete/generation (Weber and Wong, 1993). Thus, to reject the possibility of mutation, sufficient genetic markers must be used until exclusion is confirmed by two independent loci (Ron et al., 1996). According to Visscher et al. (2002) when the offspring genotype is compared to a non-related one, the probability of finding differences on only one locus out of ten analyzed is less than 10%, so it is most likely due to mutation or genotyping error.

For easier comparison of the results obtained from different laboratories, the International Society for Animal Genetics – ISAG, recommends the use of “international microsatellite panel” for cattle parentage testing. Their nomenclature, location and primer sequences are listed in Table 1.

The nine markers listed above are recommended to be shared among laboratories for purpose of record exchange. Also other markers should be added to this panel to increase efficacy in parentage testing. It is recommended that 12–14 should be used routinely. The additional 3–5 markers may vary among laboratories depending on possibility of multiplexing. Some of them most frequently used are: ETH3, TGLA53, BM1818, INRA005, INRA035 and ETH185. Also ISAG recommends exclusion probability (two parents and one parent) of each individual marker as well as of the whole panel of markers to be calculated and declared. The type of population and the number of animals used for calculations are also to be declared. ICAR recommends using Holstein as
reference where possible. To ensure sufficient experience within the lab, a number of 500 animals analyzed per year is generally set as minimum requirement for accreditation, but lower numbers may be considered acceptable in special cases.

Table 1

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chromosome</th>
<th>Primer sequence (5'–3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM1824</td>
<td>1</td>
<td>GAG CAA GGT GTT TTT CCA ATC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>forward CAT TCT CCA ACT GCT TCC TTG</td>
</tr>
<tr>
<td>BM2113</td>
<td>2</td>
<td>GCT GCC TTC TAC CAA ATA CCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reverse CTT CCT GAG AGA AGC AAC ACC</td>
</tr>
<tr>
<td>INRA023</td>
<td>3</td>
<td>reverse TAA CTA CAG GGT GTT AGA TGA ACT C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>forward GAG TAG AGC TAC AAG ATA AAC TTC</td>
</tr>
<tr>
<td>SPS115</td>
<td>15</td>
<td>reverse AAA GTG ACA CAA CAG CTT CTC CAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>forward AAC GAG TGT CCT AGT TTG GCT GTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>forward CCC TCC TCC AGG TAA ATC AGC</td>
</tr>
<tr>
<td>TGLA122</td>
<td>21</td>
<td>reverse (1) AAT CAC ATG GCA AAT AAG TAC ATA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reverse (2)* AAT CAC ATG GCA AAT AAG TAC ATA</td>
</tr>
<tr>
<td>TGLA126</td>
<td>20</td>
<td>reverse CTA ATT TAG AAT GAG AGG TTC T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>forward TTG GTC TCT ATT CTC TGA ATA TTC C</td>
</tr>
<tr>
<td>TGLA227</td>
<td>18</td>
<td>reverse CTA ATT TAG AAT GAG AGG TTC T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>forward CCA ATT CCA AAT CTG TTA ATT TG C</td>
</tr>
<tr>
<td>ETH10</td>
<td>5</td>
<td>reverse ACA GAC AGA AAC TCA ATG AAA GCA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>forward GTT CAG GAC TGG CCC TGG TAA CA</td>
</tr>
<tr>
<td>ETH225</td>
<td>9</td>
<td>reverse CCT CCA GCC CAC TTT CTC TTC T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>forward GAT CAC TTT GCC ACT ATT TCC T</td>
</tr>
</tbody>
</table>

*Corrects the null allele problem

3. CONCLUSION

Microsatellite markers are a powerful tool for animal identification and pedigree verification in cattle. To determine the level of usefulness of a microsatellite for parentage testing in a certain population, the number and frequency of its alleles must be evaluated. With this information it is possible to calculate its heterozygosis and the degree of polymorphism which in turn enables to determine the exclusion probability of each individual marker as well as of the entire panel of markers as a group.

The accuracy of the microsatellite based tests depends directly on the number and on the polymorphism of the analyzed loci.

In one population of animals different alleles on one locus arise as a direct consequence of genetic mutations which if occur in gamete’s DNA will be transferred to the progeny. Because of this fact as well as because of the fact of possible occurrence of “null alleles” it is recommended that in order to exclude a relationship one has to find incompatibility on at least two unrelated loci with high exclusion probability between the progeny and the putative parent.

The International Society for Animal Genetics (ISAG) has defined an “international marker set” of nine loci with high exclusion probability namely: BM1824, BM2113, INRA023, SPS115, TGLA122, TGLA126, TGLA227, ETH10, and ETH225. These markers are recommended as a minimal panel in bovine parentage testing for purpose of facilitating the comparison of the results obtained from different laboratories. Also for achieving higher accuracy it is recommended to genotype routinely 12–14 markers which means
that the additional 3–5 markers may vary among laboratories depending on multiplexing suitability.

In the Republic of Macedonia verification of cattle pedigrees through microsatellite markers has not yet been conducted on larger scale. In order to achieve efficient implementation of a cattle breeding program, correct pedigree information is a must. Thus it is necessary to conduct fundamental research in this area in order to check the efficacy of the recommended marker panel as well as to expand it with additional loci with high exclusion power. This would enable an establishment of a marker panel which can be used with high reliability in parentage testing in our cattle population.

REFERENCES


N. Adamov, Lj. Mickov, V. Petkov, M. Adamov


