Invited Review

Genomic dissection of inbreeding depression: a gate to new opportunities

Ino Curik¹*, Maja Ferenčaković¹, Johann Sölkner²

¹ University of Zagreb, Faculty of Agriculture, Department of Animal Science, Zagreb, Croatia.
² University of Natural Resources and Life Sciences Vienna, Department of Sustainable Agricultural Systems, Division of Livestock Sciences, Vienna, Austria.

ABSTRACT - Inbreeding depression, reduction in performance of quantitative traits, including reproduction and survival, caused by inbreeding, is a well-known phenomenon observed in almost all experimental, domesticated, and natural populations. In spite of its importance to the fate of a small population and numerous research performed in the last century, the genetic basis of inbreeding depression is still unclear. Recent fast development of molecular techniques has enabled estimation of a genomic inbreeding coefficient \( F_{ROH} \), which reflects realized autozygosity and can be further partitioned to chromosomes and chromosomal segments. In this review, we first describe classical approach used in the estimation of inbreeding in livestock populations, followed by early concepts of replacing pedigree inbreeding coefficient by individual heterozygosity. Then, we explain runs of homozygosity as key approach in estimating realized autozygosity. Furthermore, we present two different concepts of analysing regions that substantially contribute to the inbreeding depression. Thus, we describe how to identify or map mutations that result in the reduction of performance and, in terms of quantitative genetics, how to analyse the architecture of inbreeding depression. At the end, we discuss future perspectives in eliminating deleterious mutations from livestock populations.

Key Words: autozygosity, detrimental load, genomics, runs of homozygosity

Introduction

Any diploid individual is inbred if its chromosomal segments located on homologous chromosome pairs, one from each parent, are identical by descent. Inbred individuals arise as a consequence of inbreeding or, at its most restrictive and classical definition, mating of parents that are more closely related than a randomly sampled couple chosen from that population (Crow and Kimura, 1970; Lush, 1994). In populations with finite size, inbreeding is unavoidable and changes genotype frequencies by increasing homozygosity at the expense of heterozygosity, while leaving allele frequencies unaffected. For a quantitative trait, besides the redistribution of the genetic variations within and between populations (Fernandez et al., 1995), this change may lead to the inbreeding depression, a negative consequence of inbreeding that is threatening to the survival of genetically small populations. While harmful consequences of inbreeding had been known for millennia, Darwin (1868, 1876) was the first who provided detailed and recorded evidence. Although it can have various manifestations, such as lethal and detrimental malformations and abnormalities, inbreeding depression is defined as a reduction in performance of the fitness-related traits (Falconer and Mackay, 1996; Charlesworth and Willis, 2009). Inbreeding depression is a ubiquitous phenomenon observed in almost all experimental, domesticated, and natural populations, including humans (Wright, 1977; Pirchner, 1985; Falconer and Mackay, 1996).

In spite of its importance to the fate of a small population and numerous research performed in the last century, the genetic basis of inbreeding depression is still unclear. There are three theoretical hypotheses that provide explanation for the existence of inbreeding depression. The first is the partial dominance hypothesis presuming that, under directional dominance, a large number of recessive/partially recessive genes cause inbreeding depression (Davenport, 1908; Crow, 1952). The second is the overdominance hypothesis that presumes inbreeding depression as the consequence of the superiority of heterozygous genotypes
Curik et al. 2017

Curik et al., 2014). This opens the gate to the number of experimental possibilities that would further elucidate the genetic basis of inbreeding depression.

The main objective of this review is to stimulate animal breeders to perform research that would increase our understanding of inbreeding depression and reduce its negative effects on livestock populations. To achieve this objective, we started by describing the classical approach used in the estimation of inbreeding in livestock populations, followed by early concepts of surrogating the pedigree inbreeding coefficient (F_{PED}) by individual heterozygosity. Then, we explained runs of homozygosity (ROH) as the key approach in estimating realized autozygosity. Furthermore, we presented two different concepts of analysing regions that substantially contribute to the inbreeding depression. At the end, we discussed future perspectives in eliminating deleterious mutations from livestock populations.

**Classical approach of the inbreeding depression estimation**

For the various species and populations, depending on their reproductive biology and data available, various statistical approaches and experimental designs have been applied (Lynch, 1988). However, in livestock populations, regression of individual performance on F_{PED} of each animal is the most frequent procedure applied (Kristensen and Sorensen, 2005; Leroy, 2014). Although widely applied, the methodology has several critical points, some related to the statistical analysis and the other related to the estimation of inbreeding, which should be taken into account when interpreting results. Here, we will mention some statistical problems that might arise when estimating inbreeding depression. Firstly, homogeneity of variance across the inbreeding range is violated by definition, since inbreeding affects the variance of the quantitative trait (e.g. Abney et al., 2000). The low variance in inbreeding might lead to the low power of the regression analysis (Keller et al., 2011). Quite often, inbreeding range found in analyzed individuals is narrow, from 0.0 to 0.1 with several outliers having very high inbreeding >0.2, making it difficult to conclude what would happen at higher inbreeding levels. Analyses performed often require the use of mixed models (individual animal model) to account for the simultaneous rise in genetic trend and inbreeding level (sometimes resulting from more informative pedigrees) over multiple generations of breeding (Becker et al., 2015). Still, the major factor influencing the precision and bias of the inbreeding depression estimation is the error that arise in the estimation of true or realized inbreeding coefficient. Incomplete and unbalanced pedigree information is one source of errors (Cassell et al., 2003) in estimation of inbreeding depression. In experimental analyses, different strategies have been applied to account for the bias introduced by incomplete pedigrees (VanRaden, 1992; Lutanya et al., 1999; González-Recio et al., 2007 and Nagy et al., 2013). Pedigree errors are another source of errors that is present in livestock populations (Leroy and Baumann, 2011). However, a large error is introduced by theoretical properties of the F_{PED}, since it is an expectation that neglects stochastic variation of inbreeding and recombination, i.e., all individuals within a litter will have the same estimate of inbreeding. Furthermore, F_{PED} is based on infinitesimal model assuming evenly distributed autozygosity across the genome (Wray et al., 1990) and, thus, neglects the impact of selection on the regional autozygosity (Curik et al., 2001, 2002).

**Molecular measures as surrogates of the pedigree inbreeding coefficient**

*Individual heterozygosity and derivative coefficients*

In the absence of genealogical records, which is the case for almost all natural populations (for some exceptions, see Pemberton (2008)), the estimation of the negative consequences of inbreeding should rely on molecular information. Thus, it is not surprising that biologists were first to put considerable efforts in finding a surrogate coefficient of F_{PED}, that would enable inference of inbreeding depression (Mitton and Grant, 1984; David, 1998; Coltman and Slate, 2003). Inbreeding is functionally related to the individual decrease in heterozygosity, because it induces identity disequilibrium within a
Genomic dissection of inbreeding depression: a gate to new opportunities

775

The best concept that quantifies genomic inbreeding was proposed by McQuillan et al. (2008), as it estimates true or realized autozygosity directly. As introduced, F_{ROH} is a genomic measure of individual autozygosity defined as the proportion of the autosomal genome, in which autozygosity is derived from the assumption that very long stretches of homozygosity (ROH) can only result from inbreeding. When calculating F_{ROH}, McQuillan et al. (2008) excluded all the regions around centromeres, long genomic stretches devoid of single nucleotide polymorphisms (SNP), and considered only autosomal chromosome length covered by SNP. Subsequently, they calculated F_{ROH} according to the general formula: \( F_{ROH} = \Sigma L_{ROH}/L_{AUTOSOME} \), in which \( \Sigma L_{ROH} \) is the total length of all ROH according to a priory specified threshold of succeeding number of homozygotes SNP obtained from the chip arrays, while \( L_{AUTOSOME} \) is the specified length of the autosomal genome covered by SNP in chip (McQuillan et al., 2008; Ferencakovic et al., 2011). Very soon after it has been proposed, F_{ROH} has been empirically evaluated in comparison to F_{PED} and with respect to technical computations in a number of studies. All those studies were related to human (Kirin et al., 2010; Nothnagel et al., 2010; Pemberton et al., 2012) and livestock populations (Ferencakovic et al., 2011; Bosse et al., 2012; Purfield et al., 2012; Ferenčakovič et al., 2013a,b), since those are well known for systematic and informative pedigrees as well as in computer simulations (Howrigan et al., 2011). Currently, F_{ROH} is considered as standard procedure for quantifying autozygosity and its popularity is exponentially increasing to a number of studies in human (Pippucci et al., 2014; Ben Halim et al., 2015), cattle (Mészáros et al., 2015; Zavarez et al., 2015; Zhang et al., 2015a), pig (Gomez-Raya et al., 2015; Saura et al., 2015; Silió et al., 2015), poultry (Orazietti, 2015), horse (Metzger et al., 2015), and sheep and goat (Al-Mamun et al., 2015; Kim et al., 2016) populations as well as in wild (Iacolina et al., 2016), and captive populations (Nuijten et al. 2016). Not only it is a measure of true or realized inbreeding that is sensitive to the selection, the concept of F_{ROH} is easy to interpret and has several features that surpass F_{PED} in estimating negative consequences of inbreeding. Thus, F_{ROH} does allow its partition to the chromosomal level, specific chromosomal segments, and even SNP. For example, chromosomal
inbreeding can be calculated as \( F_{\text{ROH-CHROMOSOME}} = \frac{\sum L_{\text{ROH-CHROMOSOME}}}{L_{\text{CHROMOSOME}}} \) in which \( \sum L_{\text{ROH-CHROMOSOME}} \) is the total length of all ROH addressed to the chromosome in question, according to a priory specified threshold of succeeding number of homozygous SNP obtained from the chip arrays, while \( L_{\text{CHROMOSOME}} \) is the specified length of the related autosomal chromosome covered by SNP in chip. Furthermore, the length of the ROH used as a threshold in defining autozygosity is functionally related to the expected distance (in generations) to the parental common ancestor, which further can be used as proxy for the reference population. Thus, recombination events interrupt long chromosome segments and, over time, very long autozygous ROH are expected to originate from recent common ancestors. On the other side, most of short ROH are likely derived from more remote ancestors, while some short ROH might persist in a population for a very long time, much above defined base population, as a consequence of the lack of recombination or just by chance. This phenomenon is called background noise of ROH and has stronger magnitude in short ROH. In this way, ROH length can give insight into the age of inbreeding, as described in Howrigan et al. (2011) and Curik et al. (2014).

**Genomic analysis of the negative consequences of inbreeding**

In a classical quantitative genetics framework, negative consequences of inbreeding are separated to the increase in the incidence of recessively inherited disorders and to inbreeding depression. The main difference between the two phenomena is in the definition of the trait we are considering. Thus, we can look at the incidence of a single gene specific defect as a single gene trait inherited according to Mendelian rules. We can easily extend the concept of a single specific defect to a more complex variable defined as appearance of any single gene-inherited defect. If so, our model shifts to oligogenic inheritance. In contrast, inbreeding depression is originally defined for the quantitative trait with no sampling variation in the inbreeding level for individuals with the same \( F_{\text{PED}} \) (for example, all littermates have the equal \( F_{\text{PED}} \)) and with the rate of inbreeding at selected loci that is expected to be the same as at neutral loci. Thus, estimation of inbreeding depression based on \( F_{\text{PED}} \) (Kristensen and Sorensen, 2005; Casellas et al., 2009; Leroy, 2014) was perfectly adapted for the infinitesimal model (Fisher, 1918). Unfortunately, infinitesimal model (modification with dominance) does not correspond well to the inheritance of all traits, particularly traits controlled by a finite number of loci (Curik et al., 2001, 2002). Latest findings on genetic architecture of livestock species points to the mixed inheritance model, which states that a quantitative trait is controlled by a few genes with large effects and many genes with small effects (Kemper and Goddard, 2012; Kemper et al., 2012; Curik et al., 2013). Thus, to achieve better understanding of the negative effects of inbreeding, we need new methods and approaches that will respect the sampling nature of inbreeding and genetic architecture of a trait. Moreover, the complete challenge is to cover the whole spectrum of trait inheritance, starting from single gene-inherited defects to oligogenic models and across mixed inheritance models to pure polygenic models close to the infinitesimal model. We are quite much enthusiastic that the concept of \( F_{\text{ROH}} \) has all properties of an ideal inbreeding (autozygosity) coefficient needed to respond to all stated challenges. Here, we are introducing some ideas and approaches that can lead to the satisfactory solutions and development of well settled methodologies.

**Whole genome estimation of inbreeding depression**

As discussed in this present review and in a number of studies (Curik et al., 2014; Marras et al., 2014; Zhang et al., 2015a), \( F_{\text{ROH}} \) is a better estimate of whole genome autozygosity of an individual than \( F_{\text{PED}} \); thus, it is expected to replace \( F_{\text{PED}} \) in the estimation of inbreeding depression. Moreover, the variance of \( F_{\text{ROH}} \) is higher than variance of \( F_{\text{PED}} \) leading to the higher power of the regression analyses. This has been demonstrated in a computer simulation by Keller et al. (2011), who showed that \( F_{\text{ROH}} \) outperforms \( F_{\text{PED}} \) in the estimation of genome-wide autozygosity and detection of inbreeding depression. Following the recommendation of Keller et al. (2011), or just logically replacing \( F_{\text{PED}} \) with \( F_{\text{ROH}} \) in a regression analysis, was the most obvious way, done by a number of researchers, on how to proceed in estimation of inbreeding depression (e.g., Keller et al., 2012). Thus, in dairy cattle, for increase of 1% of \( F_{\text{ROH}} \) Bjelland et al. (2013) observed reduction of total milk yield to 205 days postpartum of 20 kg, increases in days open of 1.72 day, increase in maternal calving difficulty of 0.03 on a 5-point scale, and decrease in some linear-type traits. Pryce et al. (2014) confirmed inbreeding depression for milk production with stronger unfavorable effects for \( F_{\text{ROH}} \) related to closer ancestors (longer ROH) in Holstein and Jersey populations. Ferenčáković (2015) and Ferenčáković et al. (2017) found inbreeding depression for the total number of spermatozoa per ejaculate (\( P<0.05; 1\% \) of the mean per 1% of \( F_{\text{ROH}-2Mb} \)), but not for the percent of live spermatozoa in Fleckvieh bulls. Saura et al. (2015) found
significant reduction of 0.23 for the number of piglets born alive (P<0.01) and total number of piglets born (P<0.05) for 10% increase in $F_{ROH>0.5Mb}$ in Guadyerbas pigs. Significant inbreeding depression was also observed for the close inbreeding ($F_{ROH>5Mb}$), while estimates were not significant for regression on remote inbreeding ($F_{ROH>0.5Mb-5Mb}$).

Genomic mapping of loci contributing to inbreeding depression

All methods in which inbreeding depression is analysed as a regression of phenotypic value on overall inbreeding, regardless if it was derived from pedigree or molecular information, assume that autozygosity is equal on the whole genome. Experimental evidence of the genome-wide autozygosity patterns clearly shows that this is quite an unrealistic assumption as there is quite large variation of autozygosity within a genome (Pemberton et al., 2012; Sölkner et al., 2014; Orazietti, 2015). Thus, one should be aware that when SNP/regional/chromosomal autozygosity is weakly correlated with the overall inbreeding coefficient, either $F_{ped}$, $F_{PLINK}$, or $F_{ROH}$ their contribution to inbreeding depression is not properly accounted. To illustrate this, we presented correlations between $F_{ped}$ and $F_{ROH>4Mb}$ with all 29 autosomal inbreeding coefficients $ROH>4Mb$ (from $F_{ROH=4Mb-CHROMOSOME1}$ to $F_{ROH=4Mb-CHROMOSOME29}$) (Figure 1) and showed the unequal distribution of SNP autozygosity in Brown Swiss cattle genome (Figure 2).

In the context of shifting from infinitesimal (with dominance) toward oligogenic and mixed inheritance models, new approaches and ideas on how to estimate contribution of inbreeding depression for small regions or individual genes are needed. Recently, three different approaches have been used. In the first approach, an extension of the whole genome inbreeding depression estimation, the whole chromosomes or chromosomal segments divided into pieces were modelled in a classical approach as covariates (Keller et al., 2012; Ferenčaković, 2015; Saura et al., 2015). In the second approach, an extension of a genome-wise association analysis, each SNP was jointly modelled for the ROH status (0/1) and gene substitution effect (Pryce et al., 2014; Ferenčaković, 2015; Ferenčaković et al., 2017). Howard et al. (2015) upgraded that approach by searching for interactions between ROH showing inbreeding depression, using sophisticated regression tree analysis generated by Gradient Boosted Machine algorithm.

Dissection of the inbreeding depression architecture

While identifying (mapping) regions that considerably contribute to inbreeding depression is very important for the conservation and breeding management of the studied population, it does not provide the full insight into genetic architecture of inbreeding depression. For complete understanding of the inbreeding depression mechanism, it is necessary to know the number of loci involved, their inheritance mode and magnitudes of estimated effects. An impressive approach to resolve the genetic architecture of inbreeding depression was presented by Ayroles et al. (2009), who analysed the variation of gene expression across the genome between inbred and outbred lines of Drosophila. The final conclusion of Ayroles et al. (2009) was that “a large proportion of the genome (i.e. large number of genes) is involved in the expression of inbreeding depression”. Here, we take opportunity to describe the concept that, following the standard quantitative genetics terminology, can be used in dissecting inbreeding depression.

Thus, according to the Falconer and Mackay (1996) notation, for a single gene bi-allelic locus, genotypic values and expected genotypic frequencies for the Hardy-Weinberg Equilibrium (HWE) and partially inbred population can be expressed (as in Table 1). Following
notation (Table 1), the mathematical expectation (mean) of the HWE population is in formula 1:

\[ \mu_{\text{HWE}} = a(p^2) + d(2pq) - a(q^2) = a(p - q) + 2pqd \]  

(1),

while the mathematical expectation of the partially inbred populations is in formula 2:

\[ \mu_{\text{F}} = a(p^2 + pqF) + d(2pq - 2pqF) - a(q^2 + pqF) = a(p - q) + 2pqd - 2pqF = \mu_{\text{HWE}} - 2pqdF \]  

(2).

From the above formula, it is obvious that in a single gene bi-allelic locus, inbreeding depression depends on a positive expression \( 2pqdF \) and its magnitude is proportionally determined by allele frequency (highest at intermediate frequencies), \( d \) value (no inbreeding depression in additive model as \( d = 0 \)), and inbreeding level (\( F \)). In the absence of epistasis, the single locus model can be extended to more loci by summation across all loci included in the analyses; thus, inbreeding depression can be expressed by positive expression \( 2 \sum pqdF \). The expression \( 2 \sum pqdF \) can only be positive in the presence of directional dominance, e.g. constant selection pressure that is not so effective with overdominance, dominance, and partial dominance gene action as it is with additive, negative dominance, negative partial dominance, and underdominance. That is why inbreeding depression is present for traits that are under constant selection pressure such as reproduction and survival. In domestic animal populations, long selection pressure can affect production traits and create inbreeding depression.

In the \( F_{\text{ROH}} \) concept, for each SNP, we are ascribing indicator variable 1 for SNP that are in ROH or 0 for SNP that are not in ROH. This means that for each SNP, inbreeding level (\( F \)) is just the incidence of being in a ROH. After calculating allele frequencies of each SNP (\( p \) and \( q \)), we only have to estimate dominance value (\( d \)) to know all unknowns in the single locus inbreeding depression formula \( (2 \sum pqdF) \). Estimation of dominance deviations in the context of genomic information has been described in details by Wellmann and Bennewitz (2011) and Vitezica et al. (2013). Thus, with the single locus inbreeding depression concept, we would be able to dissect inbreeding depression to chromosomes (Figure 3), chromosomal regions, and single SNP. Still, summing to the regional or chromosomal inbreeding depression is not a trivial task as summing inbreeding depression estimates of neighboring loci would introduce overestimation caused by summation of confounded effects. The simplest solution is to sum only SNP that are below certain linkage disequilibrium, i.e. that are pruned, but it is not clear where to set up the threshold. The problem behind is quite similar to the regional/chromosomal decomposition (partitioning) of variance components using genomic information. Obviously, more work has to be done to provide whole procedure. However, better understanding of inbreeding depression architecture can be achieved from the knowledge on distribution of single loci inbreeding depressions. Thus, for the male sperm quality in Brown Swiss bulls, results from our pilot study (Ferenčaković et al., 2014) showed considerable dominance polygenic component contributing to the inbreeding depression, while the largest single locus contribution to inbreeding depression \((2pqdF/2 \sum pqdF)\) was 0.06\%, 0.04\%, and 0.05\% for volume of ejaculate (mL), concentration of ejaculate \((10^9/\text{mL})\), and viable spermatozoa (%), respectively.

Table 1 - Single gene bi-allelic locus genotypic values and expected genotypic frequencies in Hardy-Weinberg Equilibrium (HWE) and partially inbred population (Falconer and Mackay, 1996; modified)

<table>
<thead>
<tr>
<th>Genotype</th>
<th>( A_1A_1 )</th>
<th>( A_1A_2 )</th>
<th>( A_2A_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotypic value</td>
<td>( a )</td>
<td>( d )</td>
<td>(-a)</td>
</tr>
<tr>
<td>HWE population</td>
<td>( p^2 )</td>
<td>( 2pq )</td>
<td>( q^2 )</td>
</tr>
<tr>
<td>Partially inbred population</td>
<td>( p^2 + pqF )</td>
<td>( 2pq - 2pqF )</td>
<td>( q^2 + pqF )</td>
</tr>
</tbody>
</table>

Note that \( p \) and \( q \) are allele frequencies of \( A_1 \) and \( A_2 \) alleles, respectively, while \( F \) is the inbreeding level in a population (proportion of autozygosity).
Future opportunities

It is clear that technological advances that enabled genotyping of a large number of markers scattered through the whole genome have brought dynamics into inbreeding depression research. As shown in this review, the gate is open and we are just in the beginning, starting to collect empirical evidence and trying to develop appropriate analytical methodologies. Still, very little has been done in finding and establishing an efficient strategy that will reduce or eliminate detrimental load from genetically small populations. At the same time, modern livestock breeds are genetically narrowing, while, as a consequence, the accumulation of detrimental load is a problem of growing concern. However, technological progress is rapidly evolving and very soon we will have whole genome sequences available for a large number of animals at affordable price. The idea applied in Szpiech et al. (2013), to analyse the pattern of detrimental variation from the exome sequence, offers quite ground-breaking opportunities. Thus, according to Szpiech et al. (2013), a significantly greater fraction of all genome-wide predicted damaging homozygotes falls in ROH, especially in long ROH, than would be expected from the corresponding fraction of non-damaging homozygotes in ROH. In livestock breeding and conservation, the power and potential benefit of this concept were recognised very soon by Bosse et al. (2015), Zhang et al. (2015b), and Nuijten et al. (2016). We believe that this approach, when combined with estimation of inbreeding depression and some other innovative technologies such as gene editing (gene editing or genome editing is the insertion, deletion, or replacement of DNA at a specific site in the genome; Kim and Kim, 2014), might open the door to the management of detrimental load in genetically small populations.

Acknowledgments

This research was supported by a bilateral collaborative project involving the Croatian Ministry of Science, Education and Sports and the Austrian Agency for International Cooperation in Education and Research (OeAD GmbH) (“Runs of homozygosity: detecting selection in livestock populations”). We are also grateful to the anonymous reviewer for valuable comments and suggestions that improved the quality of the original manuscript.

This paper was presented during the Symposium on Scientific Advances in Genetics and Animal Breeding at the First International Meeting of Advances in Animal Science held at Jaboticabal, São Paulo, Brazil, on June 8 to 10, 2016.

References


R. Bras. Zootec., 46(9):773-782, 2017


