Original article

Estimation of breed contributions to present and future genetic diversity of 44 North Eurasian cattle breeds using core set diversity measures

Jörn BENNEWITZ[*a], Juha KANTANEN[b], Ilma TAPIO[b],
Meng Hua LI[b], Ernst KALM[a], Johanna VILKKI[b],
Innokentyi AMMOSOV[c], Zoya IVANOVA[d], Tatyana KISELYOVA[e],
Ruslan POPOV[f], Theo H.E. MEUWISSEN[g]

[a] Institute of Animal Breeding and Husbandry, Christian-Albrechts-University,
24098 Kiel, Germany
[b] Biotechnology and Food Research, MTT Agrifood Research Finland, 31600 Jokioinen,
Finland
[c] Batagay-Alyta, 678580, Sakha, Russia
[d] Yakut State Agricultural Academy, Ul. Krasilnikova 15, Yakutsk, 677002, Sakha, Russia
[e] Department of Genetics and Biotechnology, All-Russian Research Institute for Farm Animal
Genetics and Breeding, Moskowskoye shosse–55a, 189620 St. Petersburg-Pushkin, Russia
[f] Department of Farm Animals and Breeding, Ministry of Agriculture and Resources of Sakha,
Yakutsk, 677007, Sakha, Russia
[g] Institute of Animal and Aquacultural Sciences, Agriculture University of Norway, Box 5052,
1432 Ås, Norway

Abstract – Extinction of breeds threatens genetic diversity of livestock species. The need to conserve genetic diversity is widely accepted but involves in general two questions: (i) is the expected loss of diversity in a set of breeds within a defined future time horizon large enough to establish a conservation plan, and if so (ii) which breeds should be prioritised for such a conservation plan? The present study uses a marker assisted methodology to address these questions. The methodology combines core set diversity measures with a stochastic method for the estimation of expected future diversity and breed marginal diversities. The latter is defined as the change in the total diversity of all breeds caused by a one unit decrease in extinction probability of a particular breed. The stochastic method was validated by means of simulations. A large field data set consisting of 44 North Eurasian cattle breeds was analysed using simplified determined extinction probabilities. The results show that the expected loss of diversity in this set within the next 20 to 50 years is between 1 and 3% of the actual diversity, provided that

---

[*] Corresponding author: jbennewitz@tierzucht.uni-kiel.de

the extinction probabilities which were used are approximately valid. If this loss is to be reduced, it is sufficient to include those three to five breeds with the highest marginal diversity in a conservation scheme.

diversity measure / marginal diversity / extinction probability / cattle breeds / genetic conservation

1. INTRODUCTION

Extinction of endangered farm animal breeds leads to an irreversible loss of genetic diversity. According to the FAO [8], around one third of the recorded livestock breeds are classified as having a high risk of extinction and around 1000 have vanished during the last 100 years. The need to conserve genetic diversity is widely accepted for biological, economic and cultural reasons [13]. A main reason is that an abundant resource of genetic diversity within each livestock species is the prerequisite of coping with putative future changes in livestock farming conditions. Because financial funds available for conservation of diversity are limited, it is in general only possible to conserve a subset of important breeds rather than all endangered breeds. However, for any investigation regarding genetic diversity within a set of breeds and subsequently for the assessment of importance of particular breeds for diversity, a suitable diversity measure has to be applied.

Weitzman [18, 20] described nice mathematical and biological properties of a suitable diversity measure (the so-called Weitzman criteria) and developed a diversity measure that fulfilled these criteria. However, the Weitzman diversity measure was developed to assess diversity across species but is inappropriate across breeds [4, 5]. Alternatively, Eding *et al.* [7] introduced a core set that is built by relative breed contributions in order to maximise genetic diversity within the core set. In their approach, diversity is defined as the genetic variance that can be found in putative offspring that are obtained from interbreeding of those breeds that contribute to the core set [7]. A similar approach was developed by Caballero and Toro [4]. A drawback of this approach might be that it gives no particular weight to the between breed variance, *i.e.* to the special allele and genotype combinations that are present within breeds. Therefore, an alternative core set was recently introduced by Bennewitz and Meuwissen [2]. Their core set algorithm estimates relative breed contributions in order to maximise total genetic variance that can be found within and between breeds. Both core sets agree with the Weitzman criteria for a proper diversity measure [2,7] and additionally they are less computationally demanding even if a large number of breeds is included in the experiment.

For quantification of expected future diversity and hence of the expected loss of diversity, extinction probabilities for a defined time horizon have to be taken into account. Given that extinction probabilities are known (in real life their estimation is not a trivial task, see [3, 12, 15]), Simianer *et al.* [17] presented a deterministic method for the simultaneous calculation of expected future diversity and of marginal diversities of the breeds. The latter one is defined as the change in total diversity of all breeds caused by a one unit decrease in extinction probability of a particular breed by a conservation effort [17]. However, the deterministic approach involves 2^N times the computation of the diversity algorithm, where N is the number of breeds included in the experiment. This exponential increase in computation effort limits the application of this algorithm to smaller data sets. This is an even greater problem when the Weitzman diversity measure is used because the Weitzman diversity algorithm is itself computationally very demanding if many breeds are included [18].

This study introduces a stochastic method for the simultaneous estimation of expected future diversity and marginal diversities that is tailored to large data sets. The method was validated by means of simulations and was applied to a large field data set consisting of 44 North Eurasian cattle breeds using the two core set diversity measures mentioned above. The results (i) demonstrated the usefulness of the stochastic method and (ii) of the core set genetic diversity measures for the marker assisted estimation of present and expected future genetic diversity and (iii) they help to identify the most important breeds for the conservation of diversity within this set of North Eurasian cattle breeds, provided that the assigned extinction probabilities are approximately valid.

2. MATERIALS AND METHODS

2.1. Expected future diversity and marginal diversities

Assume a set of N breeds with known extinction probabilities z for a defined time horizon t. Further assume that the genetic diversity D of this set is estimated and the applied diversity measure fulfils the following Weitzman criteria: Monotonicity in species (D should not increase when a population is removed) and twin property (addition of a breed that is a copy of a breed already present in the set should not change D). For the estimation of expected future diversity and of breed marginal diversities, the following sampling algorithm can be applied. The algorithm repeatedly generates a sample s from the breeds included in the set. It starts with the filling in of an indicator vector $\mathbf{k}$ of dimension N (N = number of breeds). Each element k_i in $\mathbf{k}$ is allocated for

one breed i, it is either set to zero with an extinction probability z_i (breed i is extinct at time t) or to one with a probability $1 - z_i$ (breed i is alive at time t). The breeds with $k_i = 0$ are removed from the current sample s and the diversity estimation algorithm is applied to this sample. The estimated diversity within the sample, D_s, is recorded. The algorithm is repeated S times (*i.e.* S different samples s). The expected diversity at the end of the defined time horizon t can be estimated as:

$$E(D_t) = \frac{1}{S} \sum_{s=1}^{S} D_s, \tag{1}$$

and the variance of the expected diversity as

$$\text{var}(D_t) = \sigma_{D_t}^2 = \frac{1}{S-1} \sum_{s=1}^{S} (D_s - E(D_t))^2. \tag{2}$$

The covariance structure of $\mathbf{k}$ and D_t is

$$\text{var} \begin{bmatrix} \mathbf{k} \\ \mathbf{D_t} \end{bmatrix} = \begin{bmatrix} \mathbf{Q} & \mathbf{g} \\ \mathbf{g'} & \sigma_{D_t}^2 \end{bmatrix}, \tag{3}$$

where $\mathbf{Q}$ is a matrix of dimension $N \times N$ and contains the variance of k_i (that is $z_i(1 - z_i)$) on the diagonal elements and zero elsewhere. $\sigma_{D_t}^2$ is a scalar and can be obtained using (2). The vector $\mathbf{g}$ (dimension N) contains the covariance between the k_i and D_t, and these can be obtained from the S samples. The marginal diversity of breed i, m_i, is then estimated using the following regression:

$$m_i = b_{D_t,k_i} \times k_i = \frac{\text{cov}(D_t, k_i)}{\text{var}(k_i)}. \tag{4}$$

Note that the obtained marginal diversities will be positive due to the regression on k_i in (4). It is expected that this method will yield accurate estimates for S being large. This stochastic approach was compared to the deterministic method of Simianer *et al.* [17] outlined in the following.

At the end of the time horizon t 2^N different combinations of k_i within $\mathbf{k}$ are possible, thus 2^N different vectors might exist, each with probability $P(\mathbf{k})$. For a certain vector $\mathbf{k}_j$ the probability can be estimated as follows:

$$P(\mathbf{k}_j) = \prod_{i=1}^{N} \left(k_i + (-1)^{k_i} z_i \right). \tag{5}$$

The mean and variance of D_t are

$$E(D_t) = \sum_{j=1}^{2^N} P(\mathbf{k}_j) D_j, \text{ and } \text{var}(D_t) = \sum_{j=1}^{2^N} P(\mathbf{k}_j) D_j^2 - [E(D_t)]^2, \tag{6}$$

where D_j is the diversity according to $\mathbf{k}_j$. The marginal diversity of breed i is calculated as the partial derivative of $E(D_t)$ with respect to z_i:

$$m_i = +\frac{\partial E(D_t)}{\partial z_i}. \tag{7}$$

The positive sign makes it directly comparable with the marginal diversities obtained from (4) (see [17] for computational details). This method will produce correct m_i estimates (ignoring errors in the diversity measure). However, it becomes obvious that these formulae require the calculation of 2^N times the diversity measure, which becomes computationally very difficult or even impossible for large N.

2.2. Core set diversity measures

Assume a set of N breeds with a known kinship matrix $\mathbf{M}$ of dimension $N \times N$. The maximum variance total (MVT) method forms a core set in which the total genetic variance of a hypothetical quantitative trait is maximised [2]. The relative contributions of the breeds to the MVT core set are estimated as [2]

$$\mathbf{c_{mvt}} = \frac{1}{4}\left(\mathbf{M}^{-1}\mathbf{F} - \frac{\mathbf{1}'_N\mathbf{M}^{-1}\mathbf{F} - 4}{\mathbf{1}'_N\mathbf{M}^{-1}\mathbf{1}_N} \cdot \mathbf{M}^{-1}\mathbf{1}_N\right), \tag{8}$$

where $\mathbf{c_{mvt}}$ is the relative contribution vector of dimension N containing the contributions, $\mathbf{F}$ is a vector of dimension N that contains the within breed kinship, $i.e.$ $\mathbf{F} = \mathrm{diag}(\mathbf{M})$, and $\mathbf{1}_N$ a vector of dimension N containing ones. The MVT diversity measure (D_{mvt}) within the core set is then calculated as [2]

$$D_{mvt}(\text{MVT core set}) = 1 + \mathbf{c}'_{\mathbf{mvt}}\mathbf{F} - 2\mathbf{c}'_{\mathbf{mvt}}\mathbf{Mc_{mvt}}. \tag{9}$$

The core set of Eding et $al.$ [7] is built by relative breed contributions in order to maximise the genetic variance in the potential offspring of a conserved population that is obtained by interbreeding the conserved breeds. It will be termed maximum variance offspring (MVO) core set in the following. The relative breed contributions to the MVO core set (stored in the vector $\mathbf{c_{mvo}}$) are estimated as [7]

$$\mathbf{c_{mvo}} = \frac{\mathbf{M}^{-1}\mathbf{1}_N}{\mathbf{1}'_N\mathbf{M}^{-1}\mathbf{1}_N}. \tag{10}$$

The MVO diversity measure (D_{mvo}) within the core set can be estimated as [7]

$$D_{mvo}(\text{MVO core set}) = 1 - \mathbf{c}'_{\mathbf{mvo}}\mathbf{Mc_{mvo}}. \tag{11}$$

Both contribution vectors, $\mathbf{c_{mvt}}$ and $\mathbf{c_{mvo}}$, are estimated under the restriction that the contributions are zero or positive and that they sum up to one. If the breeds showed negative contributions, the most negative contribution was set to zero and the contribution vector was recalculated without the corresponding breed. This is repeated until no further negative contribution estimates are observed.

In practice the average kinship matrix $\mathbf{M}$ is generally unknown, but can be estimated from molecular marker information [6], resulting in $\hat{\mathbf{M}}$. $\hat{\mathbf{M}}$ can then be used in the equations (8)–(11). However, more accurate contribution vectors are obtained if this method is extended with bootstrapping [2]. Briefly, a bootstrap sample b is generated by sampling the individuals within breed and the marker loci across breeds simultaneously with replacement. For each b, the kinship matrix is estimated by a log-linear model [6] and subsequently the corresponding contribution vectors ($\mathbf{c_{mvt}}_b$ and $\mathbf{c_{mvo}}_b$) are estimated using equations (8) and (10). Additionally the two diversity measures D_{mvt_b} and D_{mvo_b} are calculated for each b using (9) and (11). A total of B bootstrap samples are generated. The final bootstrap estimates of the contribution vectors are the following:

$$\mathbf{c_{mvt}} = \frac{1}{B} \sum_{b=1}^{B} \mathbf{c_{mvt}}_b, \text{ and } \mathbf{c_{mvo}} = \frac{1}{B} \sum_{b=1}^{B} \mathbf{c_{mvo}}_b. \tag{12}$$

The final bootstrap estimates for D_{mvt} and D_{mvo} are the following:

$$D_{mvt} = \frac{1}{B} \sum_{b=1}^{B} D_{mvt_b}, \text{ and } D_{mvo} = \frac{1}{B} \sum_{b=1}^{B} D_{mvo_b}. \tag{13}$$

2.3. Simulation

To test the performance of the proposed sampling approach for its ability to estimate accurate expected future diversity and marginal diversities, it was compared by means of simulations with the deterministic approach. N breeds ($N = 10, 20$, respectively) were simulated for each replicate, one base breed (consisted of 50 individuals) and $N-1$ breeds that were formed by fission from the base breed. The number of generations considered was 50. For each individual a number of 20 unlinked genetic marker loci were assumed. For each breed an extinction probability was sampled from the interval 0.1/0.9. Because some constellations were computationally very demanding to simulate and analyse, the number of replicates was restricted to 10. For details of the simulation protocol see [2].

The pedigree information was recorded during the simulation and was used to calculate the true average kinship matrix $\mathbf{M}$. It was used to calculate the true actual diversities using equations (8)–(11) and to calculate the true expected future diversities and the true marginal diversities by the deterministic formulae (Eqs. (5)–(7)). The genotypes of generation 50 were used to estimate the marker estimated kinship matrix $\hat{\mathbf{M}}$ by a weighted log-linear model [6]. The actual diversity was estimated by two different methods. First by the use of $\hat{\mathbf{M}}$ in equations (8)–(11) and second by the bootstrap approach (Eqs. (12) and (13)). The expected future diversity as well as the marginal diversities of the breeds were estimated using the following three approaches. First by the use of $\hat{\mathbf{M}}$ in equations (8)–(11) and the deterministic formulae (5)–(7), second by the bootstrap approach (Eqs. (12) and (13), $B = 100$) and the deterministic formulae (5)–(7), and third by the bootstrap approach (Eqs. (12) and (13), $B = 100$) and the sampling algorithm (Eq. (1)–(4)), breeds with $k_i = 0$ were removed from all bootstrap samples). For the last approach the number of samples was varied ($S = 10, 100, 1000, 10000$).

2.4. North Eurasian cattle breeds

A data set of 44 different native and commercial cattle breeds originating from a large geographic region (*i.e.* from the Scandinavian and the Baltic countries, Finland, Russia, Byelorussia, Ukraine and Poland) was examined. The Russian breeds included in the study were from the European part of the Russian Federation except the Yakutian cattle, which originate from Asia. The Yakutian cattle make the data set of particular interest because this breed is classified as a Turano-Mongolicus type of cattle [1, 9]. This cattle breed is an endangered native breed in the Sakha Republic (formerly the Yakutia Republic) in the northeast of Siberia in Russia. The data set includes both intensively selected commercial breeds and less selected landraces. Further information of the breeds can be found at http://neurocad.lva.lt/. The breed samples were genotyped at the following 20 microsatellite markers: BM1824, BM2113, ETH10, ETH225, ETH3, HEL5, ILSTS005, INRA023, INRA035, INRA005, BM1818, CSSM66, ETH152, HEL1, HEL13, HEL9, ILSTS006, INRA032, INRA037 and INRA063. A more detailed description of the breed genotype data set will be published elsewhere. It was analysed by the two core set algorithms using the bootstrap approach as described above. A total of 100 ($B = 100$) bootstrap samples were generated and these were stored for the marginal diversity estimation. The relative breed contribution vectors as well as the conserved diversity were estimated using the equations (8)–(13). Genetic

distances were obtained from the marker estimated kinships as described in [5] and they were visualised in a dendrogram using the PHYLIP software [10].

Expected future diversity as well as marginal diversities were estimated using the sampling algorithm (Eqs. (1)–(4)) applied to the stored 100 bootstrap samples and using the diversity measures obtained from equations (12) and (13). A total of twenty thousand samples were performed ($S = 20000$). The estimation of extinction probabilities needs a substantial amount of data [3, 15]. These were not available for the majority of the 44 breeds. Therefore, the breeds were classified into five different risk classes according to their number of breeding females. Simplified extinction probabilities of the breeds were then obtained by assigning extinction probabilities to the corresponding risk class. It was assumed that these are valid for a time horizon t between 20 and 50 years into the future. The five different risk classes and the assigned extinction probabilities z are the following: class one (less than 100 breeding females) $z = 0.8$; class two (between 100 and 1000 breeding females) $z = 0.6$; class three (between 1000 and 5000 breeding females) $z = 0.4$; class four (between 5000 and 10000 breeding females) $z = 0.2$; and class five (more than 10000 breeding females) $z = 0.02$. An extinction probability above zero was assigned to the five, because a completely safe breed is not valid [15]. For the risk class of the breeds in this study as well as for other breed information see the Appendix. For each breed, the conservation potential (CP) was estimated as $CP_i = z_i \times m_i$. The conservation potential quantifies how beneficial it would be in terms of diversity to make a breed completely safe.

3. RESULTS

3.1. Results from the simulations

The results from the expected future diversity estimation are presented in Table I. It seems that it is slightly easier to estimate the expected future diversity if the number of breeds included is low. No substantial differences between the results obtained from the different methods were observed. Even the sampling approach with a low number of samples produced reliable future diversity estimates. The correlation between the estimated variances of the expected future diversities were on a similarly high level (not shown) indicating that the second moment can also be estimated accurately by the sampling approach.

The average correlation between true and estimated marginal diversities is shown in Table II. The deterministic approach produced more accurate estimates when applied to the bootstrap marker estimated kinship matrices.

Table I. Average correlation between estimated and true expected future diversity for the different methods and number of breeds (N), results from the simulations.

Method[a]	MVT core set		MVO core set	
	$N = 10$	$N = 20$	$N = 10$	$N = 20$
1	1	1	1	1
2	0.918	0.794	0.923	0.903
3	0.930	—[b]	0.921	—[b]
4 ($S = 10$)	0.888	0.561	0.763	0.938
4 ($S = 100$)	0.932	0.911	0.921	0.889
4 ($S = 1000$)	0.924	0.887	0.910	0.899
4 ($S = 10000$)	0.931	0.901	0.915	0.903

[a] Method 1: Use of **M** matrix in equations (8)–(11) and deterministic approach (Eqs. (5)–(7)), true scenario.
Method 2: Use of $\hat{\mathbf{M}}$ matrix in equations (8)–(11) and deterministic approach (Eqs. (5)–(7)).
Method 3: Use of bootstrap $\hat{\mathbf{M}}$ matrices and deterministic approach (Eqs. (5)–(7)).
Method 4: Use of bootstrap $\hat{\mathbf{M}}$ matrices and stochastic approach (Eqs. (1)–(4)), S denotes the number of samples.
[b] Computationally too demanding for estimation.

Table II. Average correlation between estimated and true marginal diversity for the different methods and number of breeds (N), results from the simulations.

Method[a]	MVT core set		MVO core set	
	$N = 10$	$N = 20$	$N = 10$	$N = 20$
1	1	1	1	1
2	0.871	0.812	0.856	0.821
3	0.921	—[b]	0.843	—[b]
4 ($S = 10$)	0.489	0.493	0.410	0.236
4 ($S = 100$)	0.833	0.688	0.807	0.736
4 ($S = 1000$)	0.908	0.853	0.874	0.825
4 ($S = 10000$)	0.916	0.858	0.891	0.843

[a,b] See Table I.

Furthermore, the deterministic approach always produced more accurate estimates than the stochastic approach (Tab. II). Hence, it is advisable to apply the deterministic approach if possible (small/moderate N) and to apply the bootstrap strategy. Otherwise, if the deterministic approach is replaced by the stochastic sampling algorithm, the reduction in accuracy is only small if a reasonably high number of samples are performed. In general, for a given S it is easier to obtain accurate estimates for a set with small N. Some of the estimates of the marginal diversities were negative due to estimation error. These estimates were set to zero.

3.2. Results from the North Eurasian cattle breeds

The results from the field data analysis are shown in Table III. The breeds with the highest relative contribution to the MVT core set were Yakutian cattle, Danish Jersey, Bohus Poll, Ringmala cattle and Väne cattle. Two thirds of the MVT core set was built by these five breeds. The genetic diversity D_{mvt} in this set was 0.989. One half of the MVO core set was built by the five top relative contributors Ringmala cattle, Bohus Poll, Doela cattle, Danish Jersey and Yakutian cattle. The genetic diversity D_{mvo} in this set was 0.878. It may be repeated at this point that D_{mvt} and D_{mvo} are not comparable due to different definitions. Not all breeds contributed to the core sets, the number of breeds with relative contribution estimates to the MVT (MVO) core set below 0.01 was 30 (22). The main contributions to the MVT core set was allocated mainly to a small number of breeds. This was less remarkable for the MVO core set.

Figure 1 shows the Neighbour-Joining dendrogram obtained from the kinship genetic distances. The breeds seemed to group into nine different clusters. Nearly each cluster was represented by one or two breeds from the top 10 relative contributors for both core sets, except cluster four and nine. In general, the MVT core set gave higher contributions to breeds with a higher within breed kinship and thus to breeds that show a longer branch length. In contrast, the MVO core set gave higher contributions to breeds that were closer to a hypothetical founder breed. For example, the Yakutian cattle (longer branch in the dendrogram) belongs to both top 10 contribution lists but shows a substantial higher contribution to the MVT core set than to the MVO core set. The opposite is true for the Doela cattle, which shows a shorter branch (Fig. 1, Tab. III).

In the MVT core set, the expected diversity D_{mvt} at time t was 0.958 and its standard deviation was 0.014. The highest marginal diversity within this core set was obtained by the Yakutian cattle, followed by the Danish Jersey, Bohus Poll, Icelandic cattle and Ringamala cattle (Tab. III). Because the Yakutian cattle and the Bohus Poll also showed a high extinction probability, their conservation potentials were also very high (Tab. III). In the MVO core, the expected D_{mvo} diversity at time t was 0.871 with a standard deviation of 0.003. The highest marginal diversity was obtained by the Danish Jersey, Ringamala cattle, Bohus Poll, Doela cattle and Latvian Blue. The highest conservation potential was obtained by the Ringamala cattle followed by the Bohus Poll. In general, the distribution of the marginal diversities in the MVO core set was much smoother than that of the MVT core set. Many breeds showed a zero marginal diversity for both diversity measures, and therefore also a zero conservation potential. This occurred even more often for the MVT core set than for the MVO core set.

Table III. Assigned extinction probability (z), relative contribution (c), marginal diversity (m) and conservation potential (CP) for the breeds and the corresponding core set.

Breed	z_i	c_i	m_i^a	CP_i^a	c_i	m_i^a	CP_i^a
		MVT core set			MVO core set		
Byelorussian Red	0.02	0	0	0	0.007	0	0
Danish Jersey	0.02	0.126	158.20	3.16	0.100	40.48	0.81
Estonian Red	0.02	0.001	0	0	0.004	1.97	0.04
Finnish Ayrshire	0.02	0.024	6.89	0.14	0.043	13.96	0.28
Finnish Holstein-Friesian	0.02	0	0	0	0.001	0	0
Icelandic cattle	0.02	0.066	102.62	2.05	0.038	12.27	0.25
Kholmogory	0.02	0.001	3.36	0.07	0.001	0	0
Latvian Brown	0.02	0.005	0	0	0.014	0	0
Lithuanian Black and White	0.02	0	32.69	0.65	0.009	11.33	0.23
Lithuanian Red	0.02	0.001	5.93	0.12	0.001	1.15	0.02
Norwegian Dairy cattle	0.02	0	0	0	0.002	0	0
Polish Black and White	0.02	0	9.14	0.18	0	2.13	0.04
Swedish Holstein-Friesian	0.02	0	0.95	0.02	0.001	0.66	0.01
Swedish Red and White	0.02	0	0	0	0	0	0
Yaroslavl	0.02	0	0	0	0.001	0	0
Istoben	0.2	0.007	9.11	1.82	0.024	6.34	1.27
Suksun	0.2	0.004	0	0	0.013	1.05	0.21
Blacksided Troender	0.4	0.002	0	0	0.004	0	0
Estonian Native	0.4	0.001	0	0	0.009	0	0
Swedish Mountain cattle	0.4	0.023	17.77	7.11	0.024	21.05	8.42
Telemark cattle	0.4	0.008	2.37	0.95	0.023	11.90	4.76
Ukrainian Whiteheaded	0.4	0.001	0	0	0.008	0	0
Western Finncattle	0.4	0	0	0	0.002	0	0
Doela cattle	0.6	0.064	25.80	15.48	0.101	22.73	13.64
Eastern Finncattle	0.6	0.001	0.52	0.31	0.015	7.47	4.48
Eastern Red Polled	0.6	0.024	24.58	14.75	0.025	1.65	0.99
Jutland breed	0.6	0.002	4.65	2.79	0.007	0.75	0.45
Latvian Blue	0.6	0.017	24.58	14.75	0.028	22.10	13.26
Latvian Danish Red	0.6	0.010	0	0	0.028	0.72	0.43
Lithuanian Light Grey	0.6	0	2.46	1.48	0	0	0
Lithuanian White Backed	0.6	0.001	0	0	0.009	0	0
North Finncattle	0.6	0.001	0	0	0	1.50	0.90
Pechora	0.6	0	0	0	0	0	0
Red Danish 1970	0.6	0.046	71.95	43.17	0.039	4.96	2.98
Swedish Red Polled	0.6	0	0	0	0.002	0	0
Ukrainian Grey	0.6	0.006	0	0	0.011	0	0
Väne cattle	0.6	0.096	50.54	30.32	0.046	11.16	6.70
Western Fjord cattle	0.6	0.001	0	0	0.014	0	0
Western Red Polled	0.6	0.017	10.14	6.08	0.029	3.60	2.16
Yakutian cattle	0.6	0.200	218.24	130.94	0.070	8.96	5.38
Bohus Poll	0.8	0.123	125.86	100.69	0.113	25.81	20.65
Danish Black-Pied 1965	0.8	0.002	0.57	0.46	0.006	1.17	0.94
Fjällnära cattle	0.8	0.006	0	0	0.015	0	0
Ringamala cattle	0.8	0.115	91.23	72.98	0.114	29.83	23.86

[a] Multiplied by 10^4.
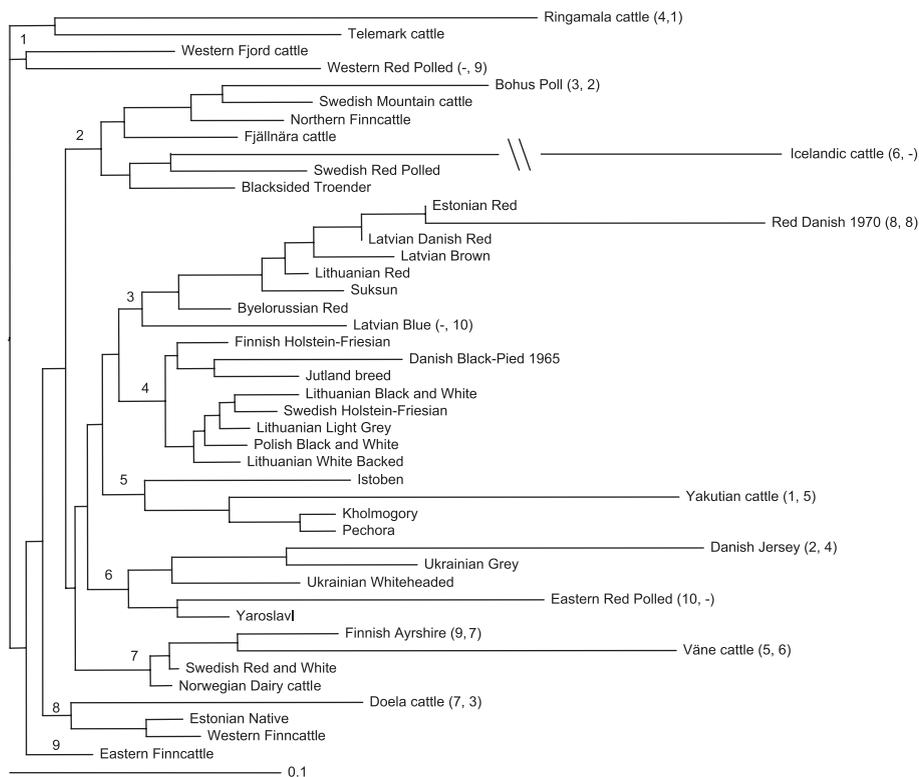
Figure 1. Neighbour-Joining dendrogram of the 44 cattle breeds obtained from kinship genetic distances. The cluster number assigned by the authors is given at the beginning of the common branch of the cluster. In parentheses the ranking number of the top ten core set contributors, the first figure is the ranking number to the MVT core set and the second to the MVO core set. The branch length of the Icelandic cattle is approximately 4 units.

4. DISCUSSION

The first section of the discussion focuses on the stochastic method for the expected future diversity and the marginal diversity estimation. In the second section the applied core set diversity measures are compared using the results from the analysis of the 44 North Eurasian cattle breeds. Finally, the assessment of the present and expected future diversity within this set of North Eurasian cattle breeds and the importance of particular breeds for the diversity are discussed.

4.1. Sampling algorithm

The numerical results of the simulations showed that the introduced sampling algorithm was suitable to accurately estimate the expected future diversity and hence, the expected loss of diversity even with a low number of samples. Furthermore, the sampling algorithm was suitable to obtain marginal diversity estimates with a reasonably high level of accuracy, if a sufficient number of samples was chosen. The required number of samples is a function of the number of breeds included in the set and of the distribution of the breed extinction probabilities. If many breeds exhibit an intermediate extinction probability (around 0.5), a higher number of samples is required because the probability of the breed constellations at time t (breeds extinct/alive combinations) is more equally distributed, *i.e.* rare constellations receive a higher probability. A practical solution to determine the appropriate number of samples is as follows: during the sampling process the temporary results of the algorithm are frequently divided into two parts and temporary marginal diversities are estimated from both parts separately. If the correlation between the temporary estimates obtained from the two parts reaches a defined threshold level (0.99, for instance), the sampling algorithm can be stopped and the final marginal diversities can be obtained using all samples. This strategy was used to determine the required number of samples in the field data analysis.

A further outcome of the simulation is that the bootstrap strategy as proposed by [2] is not only appropriate for the estimation of relative breed core set contributions but also for the estimation of marginal diversities. Hence the combination of the bootstrap method and the sampling algorithm is a powerful and computationally feasible tool to estimate marginal diversities for the two core set diversity measures.

4.2. Present and expected future diversity of the North Eurasian cattle breeds

Assuming that the determined extinction probabilities are approximately valid, the expected loss of genetic diversity within the time horizon t compared to the actual diversity seems to be low for both core set diversity measures (around 3% for the MVT and only around 1% for the MVO diversity measure). One reason is that many breeds are not or only slightly endangered and hence, show only a small extinction probability. A second reason is due to the nature of both core set algorithms. If a breed is extinct and thus removed from the core set, related breeds obtain higher contributions. Hence, a loss of

a breed can be compensated to some extent by a re-adjustment of the breed contribution vector, which reduces the loss of diversity.

The compensating mechanism is more efficient in the MVO core set than in the MVT core set, as indicated by the three times higher loss of the MVT diversity than the MVO diversity at time t. This agrees with the fact that the MVT core set distributes the relative breed contributions across fewer breeds than the MVO core set does. Additionally, as already mentioned, the distribution of the MVO core set marginal diversities is much more equal compared to the MVT core set (Tab. III). The reason is that the MVT core set values the particular combination of alleles and genotypes in a certain breed and the MVO core set the number of rare alleles in a breed. It seems that the distribution of these combinations over the breeds is more unequal compared to the distribution of distinct alleles. Following this, a breed with a high contribution to the MVT core set is in general more difficult to replace with other breeds compared to a breed with a high contribution to the MVO core set. A good example is the Yakutian cattle, which receives the highest relative contribution and also the highest marginal diversity to the MVT core set suggesting that the breed has specific allelic combinations. A loss of this breed is more difficult to compensate for by the MVT core set than by the MVO core set.

The attractiveness of the MVT core set diversity measure arises from the fact that it attempts to conserve breeds with a large difference in the respective population mean of a hypothetical quantitative trait, which enables a strong selection among the conserved breeds for a desirable trait. However, the MVO core set strives to conserve as many alleles as possible and hence, maximises the possible selection directions. The question of which diversity measure is appropriate depends on the assumed future scenario. If a putative future condition forces the breeders to create a new synthetic breed through interbreeding breeds, the MVO diversity measure is appropriate. In contrast, if the putative future scenario still allows the use of a commercial breed that is upgraded by genetics from conserved breed(s), the MVT diversity measure is superior in that more extreme breeds are conserved. It is difficult to decide which scenario is more likely to occur but it can be argued that putative changes in the production environment might occur slowly rather than overnight, hence time might be available to adapt the commercial breed by introducing genetics from a conserved breed. This favours the MVT diversity measure. Recently, Toro and Caballero [19] raised the question of the relative importance of the between breed diversity versus the within breed diversity and suggested finding a compromise between them. With regards to this, Piyasatian and Kinghorn [14] suggested giving five times greater weight to the between breed diversity because

Table IV. Correlation matrix of the assigned extinction probability (z), the relative breed contribution (c), the marginal diversity (m) and the conservation potential (CP), results from the field data analysis. c, m and CP are grouped by the two core sets.

		MVT core set			MVO core set		
		c_i	m_i	CP_i	c_i	m_i	CP_i
	z_i	0.243	0.129	0.389^a	0.307	0.110	0.452^a
MVT core set	c_i		0.950^b	0.854^b	0.852^b	0.682^b	0.589^b
	m_i			0.806^b	0.754^b	0.644^b	0.452^a
	CP_i				0.687^b	0.430^a	0.663^b
MVO core set	c_i					0.852^b	0.777^b
	m_i						0.723^b

[a] Significant ($P_{approximate} < 0.01$), approximation due to the large number of zero elements and taking the error probability values from standard tables.
[b] Highly significant ($P_{approximate} < 0.001$).

this is more accessible compared to the within breed diversity. The MVT diversity values between breed variation over and above the MVO diversity (as advocated by Toro and Caballero [19]) and the allelic variation criterion of Piyasatian and Kinghorn [14]. The latter one only values the contribution of between breed variance to the variance of the putative offspring of the conserved breeds, instead of also valuing the variance contained in the genotypes of the conserved breeds. It must be noted at this point that despite the differences between the two core set diversity measures mentioned above, they are based on similar concepts and they produce to a large extent comparable results. This becomes obvious when looking at the correlation between the relative contributions of the breeds to the two core sets as well as at the corresponding marginal diversities and conservation potentials (Tab. IV).

The correlations between the relative breed contributions to the core sets and the corresponding marginal diversities are high in this study (around 0.9, see Tab. IV). However, the relative contributions are valid only for time $t = 0$, *i.e.* all breeds are alive. They do not help to quantify the expected future diversity within the set of the breeds nor the effect of the reduction of an extinction probability of a certain breed because they ignore the extinction probabilities.

The approach for the estimation of expected future diversity and breed marginal diversities assumes that the breed diversity contribution remains constant over time if the breed is alive or drops to zero if the breed is extinct. A putative change in the effective population size over time is not considered, but it can reasonably be assumed that it might affect the breed contribution. Further research is needed in this field.

As already mentioned, the estimation of extinction probabilities is a diffi-
cult task [3, 12, 15]. Therefore, in this study they were determined by simply
assigning probabilities to the defined five risk classes for endangerment. In or-
der to test the sensitivity of the somewhat arbitrary values, two different sets
of assigned extinction probabilities were used. The first set was as described
above and the extinction probabilities of the second set were exactly the half
from those of the first set. Consequently, two marginal diversity estimates for
each breed were estimated. A linear model was applied that included the breed
and the set of extinction probabilities (either set one or set two) as fixed ef-
fects. The null hypothesis was that both marginal diversities within a breed
were the same, the alternative hypothesis was that at least for one breed the
marginal diversities were not the same. The results of this model suggested to
reject the null hypothesis ($P < 0.01$ for both diversity measures). In general,
the marginal diversities were somewhat higher for the higher extinction prob-
abilities, however, without changing the ranking order of the breed marginal
diversities (not shown). The expected loss of diversity was around 50% lower
for the set of lower extinction probabilities. Based on these results, it is ben-
eficial to have more accurate extinction probability estimates because more
precise conclusions could be drawn from the results obtained. An alternative
to the applied combination of extinction probabilities and diversity measures
is the so-called 'safe set safe set+1' approach as used in [7]. By using this
approach, the ranking of endangered breeds for conservation priority is done
according to their contribution to the diversity of a safe (*i.e.* not endangered)
set of breeds. The advantage is that no extinction probabilities are needed, it
only has to be decided which breeds form the safe set.

4.3. Conservation of the North Eurasian cattle breed genetic diversity

As mentioned above, even without any conservation effort the expected loss
of diversity within this set of breeds is low, regardless of the applied diversity
measure. If, however, even this small loss is to be reduced, it is not helpful
to reduce the extinction probabilities of the most endangered breeds without
considering the marginal diversities because there is virtually no relationship
between the extinction probability on the one hand and the relative breed
contribution and marginal diversity on the other hand as shown in Table IV.
Similar results found in a different data set were reported by [17].

Assume a conservation scheme in which the cost to make a breed safe (*i.e.*
bringing its extinction probability close to zero) is independent from its extinc-
tion probability and more or less equal for all breeds. Under these conditions,

the breeds with the highest conservation potential would receive the highest priority for the inclusion in the conservation programme. In the present study the five breeds with the highest conservation potential for the MVT diversity measure are Yakutian cattle, Bohus Poll, Ringamala cattle, Red Danish 1970 and Väne cattle. For the MVO, these breeds are the Ringamala cattle, Bohus Poll, Doela cattle, Latvian Blue and Swedish Mountain cattle. If by including them in a conservation plan the extinction probability of these breeds would be close to zero, there would be almost no expected loss of diversity at the end of the time horizon t (not shown). However, these assumptions might only be valid in *ex-situ* conservation schemes (*e.g.* transferring a deposit of genetic material from endangered breeds to a genebank), but not in *in-situ* conservation schemes, where the breeds are conserved within the production system. For the latter situation, Simianer *et al.* [17] proposed a more sophisticated framework to identify the most efficient conservation plan. The current study provides the prerequisite to apply the methods of [17], given that the unknowns in the method can be replaced by reliable estimates. However, it is reasonable to assume that those three to five breeds with high conservation potential will also be recommended for a conservation plan by the algorithms of [17].

Throughout this study the focus was based exclusively on genetic diversity as a criterion for the conservation of a breed. Other conservation criteria such as adaptation to a specific environment, special traits of economic interest or historical or cultural value are discussed by *e.g.* [11, 13, 16].

5. CONCLUSION

It was shown that the sampling algorithm in combination with the two core set genetic diversity measures provides a suitable statistical tool for the marker assisted estimation of present and expected future diversity and of breed marginal diversities, given that extinction probabilities of the breeds are known. The analysis of the North Eurasian cattle breeds revealed that without any conservation efforts the expected loss of diversity during the next 20 to 50 years is between 1 and 3% from actual diversity, provided that the simplified determined extinction probabilities are approximately valid. If this loss was to be reduced or even stopped by a limited conservation fund, it seems to be sufficient to invest the available money in the reduction of the extinction probability of those three to five breeds with the highest marginal diversity and the highest conservation potential. These are not necessarily the most endangered breeds.

REFERENCES

[1] Bannikova L.V., Zubareva L.A., Genetic structure of some native and commercial breeds of cattle (*Bos taurus*) from Eurasia, Russian J. Genet. 31 (1995) 597–607.

[2] Bennewitz J., Meuwissen T.H.E., A novel method for the estimation of the relative importance of breeds in order to conserve the total genetic variance, Genet. Sel. Evol. 37 (2005) 315–337.

[3] Bennewitz J., Meuwissen T.H.E., Estimation of extinction probabilities of five German cattle breeds by population viability analysis, J. Dairy Sci. 88 (2005) 2949–2961.

[4] Caballero A., Toro M.A., Analysis of genetic diversity for the management of conserved subdivided populations, Conserv. Genet. 3 (2002) 289–299.

[5] Eding J.H., Meuwissen T.H.E., Marker based estimates of between and within population kinships for the conservation of genetic diversity, J. Anim. Breed. Genet. 118 (2001) 141–159.

[6] Eding J.H., Meuwissen T.H.E., Linear methods to estimate kinships from genetic marker data for the construction of core sets in genetic conservation schemes, J. Anim. Breed. Genet. 120 (2003) 289–302.

[7] Eding J.H., Crooijmans R.P.M.A., Groenen M.A.M., Meuwissen T.H.E., Assessing the contribution of breeds to genetic diversity in conservation schemes, Genet. Sel. Evol. 34 (2002) 613–633.

[8] FAO, World watch list for domestic animal diversity, 3nd ed., Rome, 2000, http://www.fao.org/dad-is/.

[9] Felius M., Cattle breeds. An Encyclopedia, Misset uitgeverij bv, Doetinchem, NL, 1995.

[10] Felsenstein J., PHYLIP (Phylogeny Inference Package) Version 3.5c, University of Washington, 1995.

[11] Gandini G., Villa E., Analysis of the cultural value of local livestock breeds: a methodology, J. Anim. Breed. Genet. 120 (2003) 1–11.

[12] Gandini G., Ollivier L., Danell B., Distl O., Georgoudis A., Groeneveld E., Martyniuk E., van Arendonk J.A.M., Wolliams J.A., Criteria to assess the degree of endangerment of livestock breeds in Europe, Livest. Prod. Sci. 91 (2004) 173–182.

[13] Oldenbroek J.K., Genebanks and the conservation of farm animal genetic resources. DLO, Lelystad, 1999.

[14] Piyasatian N., Kinghorn B.P., Balancing genetic diversity, genetic merit and population viability in conservation programmes, J. Anim. Breed. Genet. 120 (2003) 137–149.

[15] Reist-Marti S.B., Simianer H., Gibson J., Hanotte O., Rege J.E.O., Weitzman's approach and conservation of breed diversity: an application to African breeds, Conserv. Biol. 17 (2003) 1299–1311.

[16] Ruane J., A critical review of the value of genetic distances in conservation of animal genetic resources, J. Anim. Breed. Genet. 116 (1999) 317–323.

[17] Simianer H., Marti S.B., Gibson J., Hanotte O., Rege J.E.O., An approach to the optimal allocation of conservation funds to minimize loss of genetic diversity between livestock breeds, Ecol. Econ. 45 (2003) 377–392.

[18] Thaon d'Arnoldi C., Foulley J.-L., Ollivier L., An overview of the Weitzman approach to diversity, Genet. Sel. Evol. 30 (1998) 149–161.

[19] Toro M.A., Caballero A., Characterisation and conservation of genetic diversity between breeds. Session G3.4. 55th Annual Meeting of the European Association for Animal Production, 5–8 September, 2004, Bled, Slovenia.

[20] Weitzman M.L., On diversity, Quart. J. Econ. 107 (1992) 363–405.

APPENDIX

Information about the breeds included in the field data set.

Breed	Sample size	Sample origin	Risk class[a]
Byelorussian Red	23	Byelorussia	Five
Danish Jersey	41	Denmark	Five
Estonian Red	40	Estonia	Five
Finnish Ayrshire	46	Finland	Five
Finnish Holstein-Friesian	43	Finland	Five
Icelandic cattle	44	Iceland	Five
Kholmogory	42	Russia	Five
Latvian Brown	40	Latvia	Five
Lithuanian Black and White	41	Lithuania	Five
Lithuanian Red	40	Lithuania	Five
Norwegian Dairy cattle	38	Norway	Five
Polish Black and White	30	Poland	Five
Swedish Holstein-Friesian	44	Sweden	Five
Swedish Red and White	39	Sweden	Five
Yaroslavl	44	Russia	Five
Istoben	49	Russia	Four
Suksun	40	Russia	Four
Blacksided Troender	34	Norway	Three
Estonian Native	40	Estonia	Three
Swedish Mountain cattle	41	Sweden	Three
Telemark cattle	46	Norway	Three
Ukrainian Whiteheaded	11	Ukraine	Three
Western Finncattle	41	Finland	Three
Doela cattle	35	Norway	Two
Eastern Finncattle	31	Finland	Two
Eastern Red Polled	11	Norway	Two
Jutland breed	49	Denmark	Two
Latvian Blue	40	Latvia	Two
Latvian Danish Red	40	Latvia	Two
Lithuanian Light Grey	41	Lithuania	Two
Lithuanian White Backed	40	Lithuania	Two
North Finncattle	26	Finland	Two
Pechora	33	Russia	Two
Red Danish 1970	39	Denmark	Two
Swedish Red Polled	34	Sweden	Two
Ukrainian Grey	30	Ukraine	Two
Väne cattle	18	Sweden	Two
Western Fjord cattle	41	Norway	Two
Western Red Polled	36	Norway	Two
Yakutian cattle	54	Russia	Two
Bohus Poll	14	Sweden	One
Danish Black-Pied 1965	27	Denmark	One
Fjällnära cattle	15	Sweden	One
Ringamala cattle	20	Sweden	One

[a] Classification done according to the number of breeding females as follows: class one (less than 100 females), class two (between 100 and 1000 females), class three (between 1000 and 5000 females), class four (between 5000 and 10000 females), class five (more than 10000 females). From this, class one is critically endangered and class five not endangered.