Int J Legal Med (2011) 125:111–119
DOI 10.1007/s00414-009-0353-8

SHORT COMMUNICATION

Population studies of 16 bovine STR loci for forensic purposes

L. H. P. van de Goor · M. T. Koskinen · W. A. van Haeringen

Received: 4 August 2008 / Accepted: 19 May 2009 / Published online: 9 June 2009

Abstract As a consequence of the close integration of cattle into the food chain of humans, forensically relevant cases involving cattle (*Bos taurus*) DNA analysis are common. However, scientific publications reporting the information content of the commonly used bovine short tandem repeat (STR) loci remains scarce. Population studies were performed for 16 polymorphic STR loci (BM1818, BM1824, BM2113, CSRM60, CSSM66, ETH3, ETH10, ETH225, HAUT27, ILSTS006, INRA023, SPS115, TGLA53, TGLA122, TGLA126, and TGLA227) including 4,162 randomly selected cattle representing 20 distinct breeds. The power of parental exclusion, expected and observed heterozygosity, probability of identity, and non-amplifying ("null") allele frequencies were calculated. Major differences existed in the information content between different cattle breeds. The selection of 16 STR loci, partially recommended by International Society for Animal Genetics as the minimum standard needed for bovine STR typing, was sufficient for forensic analysis. Furthermore, the efficacy of the loci was assessed in assigning unknown individuals to the correct breed based on genotype data. The individual assignment tests provided excellent success in several breeds.

Electronic supplementary material The online version of this article (doi:10.1007/s00414-009-0353-8) contains supplementary material, which is available to authorized users.

L. H. P. van de Goor (✉) · W. A. van Haeringen
Dr. Van Haeringen Laboratorium BV,
Agro Business Park 100,
6708 PW Wageningen, The Netherlands
e-mail: lgo@vhladmin.nl

M. T. Koskinen
Finnzymes Diagnostics,
Keilaranta 16A,
02150 Espoo, Finland

Keywords Bovine short tandem repeats · *Bos taurus* · STR population studies · Breed assignment · Forensic DNA analysis

Introduction

The recently sequenced bovine (*Bos taurus*) genome suggests a continued genetic exchange between wild cattle populations during their coexistence over a wide geographical range [1]. At present, since the domestication event of cattle several 1,000 years ago, more than 50 distinct breeds are recognized. As a consequence of the close integration of cattle into the food chain of humans, forensically relevant cases involving cattle, such as identity forgery or cattle theft, are relatively common. Some of these examples have been published recently, demonstrating the relevance of this issue in forensic casework [2]. However, population studies detailing the information content of the majority of widely used cattle short tandem repeat (STR) loci remains limited [3, 4], making the application of these loci difficult, especially in a forensic setting.

Bovine STR loci are extensively used for parentage verification by the animal breeding industry, and the first description of cattle microsatellites in the 1990s has eventually led to international recommendations for these loci in 1998 by the International Society for Animal Genetics (ISAG); ISAG currently recognizes and has validated in numerous interlaboratory comparison tests 12 microsatellite loci for routine use in bovine kinship analysis (http://www.isag.org.uk/ISAG/all/ISAG2008_CattleParentage.pdf). New challenges are now emerging for genotyping laboratories because official organizations have initiated recommendations and minimum requirements for identity and kinship analysis. The International Committee for

Animal Recording has recently instituted a working group on guidelines of accreditation of DNA paternity testing in cattle (www.icar.org). Recently, Budowle et al. [5] described the recommendations that need to be implemented by different laboratories for animal DNA forensic and identity testing. As a first step in the improvements in bovine genotyping, recently a proposal for a repeat-based nomenclature of 16 microsatellite markers (BM1818, BM1824, BM2113, CSRM60, CSSM66, ETH3, ETH10, ETH225, HAUT27, ILSTS006, INRA023, SPS115, TGLA53, TGLA122, TGLA126, and TGLA227) was published [6]. A second step in this process is the reporting of large-scale population data, similar to the standard operating procedure in humans and recently described in other nonhuman species relevant in forensic casework, such as cats [7, 8] and dogs [9–11]. Measures including power of exclusion (PE) and the observed heterozygosity (HO) have been used in bovine parentage verification to assess the quality and power of routine analysis [12–14]. In this study, based on a data set of 4,162 animals representing 20 breeds, the efficacy of a set of 16 bovine STRs (BM1818, BM1824, BM2113, CSRM60, CSSM66, ETH3, ETH10, ETH225, HAUT27, ILSTS006, INRA023, SPS115, TGLA53, TGLA122, TGLA126, and TGLA227), all recommended by the Food and Agriculture Organization of the UN (www.fao.org), is reported for forensic investigations. This set of 16 STRs, which includes the 12 loci recommended by ISAG for bovine parentage testing, is included in a commonly used commercial kit (Finnzymes Diagnostics, Espoo, Finland) and routinely employed by bovine genotyping providers. Recently, all 16 STR loci, originally described as dinucleotide repeat sequences, were shown to exhibit a simple or compound variable repeat structure [6], and a proposal was presented for the allele nomenclature of the 16 STR loci based on the number of repeat units and adopted from the recommendations of the International Society of Forensic Genetics (ISFG) for the nomenclature of human STRs. Furthermore, Van de Goor et al. [6] provide information on details such as locus name, chromosomal location, repeat structure and sequence, original reference, primer sequences, and size range of the amplicon length. It is a natural addition to present population data on the performance of the STR loci, as presented in this publication. The genotype data of the STR loci were used to assess marker statistics (PE, expected heterozygosity (HE), HO, probability of identity (P_{ID}), and null allele frequencies). Furthermore, the improved power of the 16 loci, compared to the 12 loci recognized by ISAG as a minimum standard for identity and kinship analysis, has been investigated. Finally, the reliability of individual assignment tests in identifying the correct breeds of origin of unknown samples, based on genotype data, was explored.

Materials and methods

Samples and DNA extraction

The samples used in this study were collected between February 2006 and February 2007. In total 4,162 samples from 20 breeds (Table 1) were available, of which 75% were hair root samples, 23% blood samples, and 2% semen straws and other sample types. A limited number of breeds originate from the Netherlands, i.e., Brandrood Cattle, Dutch Friesian, Groningen Whiteheaded, Dutch Belted, Verbeterd Roodbont, and Maas Rijn IJssel), whereas all other breeds exist and are common throughout Europe and North America. Genomic DNA was isolated using routine procedures. For blood samples, 10 μl of blood was washed three times in 150 μl Tris–HCL-based buffer. The cell pellet was lysed with proteinase K (0.5 U for 45 min at 56°C followed by heat inactivation). For hair root samples, approximately eight hair follicles were placed into a polymerase chain reaction (PCR) tube and lysed with proteinase K (6 U overnight at 56°C followed by heat inactivation). For the DNA extraction from semen, a DNA isolation kit was used according to the manufacturer's instructions (Puregene, Gentra Systems, Minneapolis, MN, USA).

Table 1 Cattle breeds and sample numbers (*N*) included in this study

Breed	Abbreviation	Number
Blonde D'Aquitaine	BA	165
Belgian Blue	BBL	51
Brandrood Cattle	BRR	41
Charolais	CHL	28
Dexter	DEX	428
Dutch Friesian	FH	42
Groningen Whiteheaded	G	24
Galloway	GAL	88
Heck Cattle	HEC	39
Hereford	HER	62
Holstein Friesian	HF	2,507
Holstein	HOL	254
Limousin	LIM	126
Dutch Belted	LV	24
Marchigiana	MAR	17
Maas Rijn IJssel	MRY	41
Scottisch Highlander	SH	118
Verbeterd Roodbont	VRB	42
Wagyu	WAG	20
Waldviertler Blondvieh	WAL	45

STR loci and PCR

Details about the 16 loci (BM1818, BM1824, BM2113, CSRM60, CSSM66, ETH3, ETH10, ETH225, HAUT27, ILSTS006, INRA023, SPS115, TGLA53, TGLA122, TGLA126, and TGLA227) have been described by Van de Goor et al. [6] (an electropherogram is presented in Fig. 1 of the Electronic Supplementary Material). The loci were genotyped using the Bovine Genotypes™ Panels 1.1 and 2.1 (Finnzymes Diagnostics). The commercial kit was used according to the manufacturer's instructions.

Fragment analysis and allele calling

The products of the PCR reactions were separated following the manufacturers' instructions on an ABI 3100 sequencer with the capability to identify five fluorescent dyes. The data were reviewed, and alleles were called using GeneScan and Genotyper programs (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions.

Statistical analysis of population data

The programs CERVUS 2.0 [15] and Park Microsatellite Tool Kit [16] were used to calculate the average number of alleles, PE in the absence of genetic information from the other parent (PE1), PE if the DNA type of one known parent is available (PE2), HO, and HE for every population. The table with the original data as well as the frequency information is available in the Electronic Supplementary Material.

The sibling probability of identity (sib P_{ID}) and the Hardy–Weinberg probability of identity (HW P_{ID}) were calculated using the program GENECAP [17].

Based on data from routing parentage verifications, it was known that null alleles exist in the loci INRA023 and HAUT27 in a limited number of (local) breeds. In cases for which one parent was excluded only for locus INRA023 or locus HAUT27 and this parent and the suspected offspring were both homozygous for the locus, additional loci were genotyped. If, after typing a total of 22 loci, the parent could still be excluded only for that one locus, the presence of a null allele was assumed. The presence of those null alleles was recorded in a database, and their frequencies were calculated.

Table 2 Statistical parameters of 20 cattle breeds based on 12 ISAG-recognized loci

Breed	PE1	PE2	HE	HE (sd)	HO	HO (sd)	# Alleles	# Alleles (sd)	HW P_{ID}	sib P_{ID}
BA	0.9961	0.9999	0.7394	0.0286	0.7394	0.0099	7.00	2.45	5.8042E−13	1.9054E−05
BBL	0.9907	0.9997	0.6908	0.0447	0.6977	0.0186	6.58	1.93	1.7344E−11	5.2269E−05
BRR	0.9801	0.9992	0.6661	0.0362	0.6585	0.0214	5.75	1.86	1.5085E−10	8.9079E−05
CHL	0.9834	0.9994	0.6444	0.0538	0.6488	0.0260	6.25	2.45	5.5851E−11	8.8733E−05
DEX	0.9897	0.9997	0.7087	0.0282	0.6910	0.0064	6.92	2.07	1.0713E−11	3.5946E−05
FH	0.9957	0.9999	0.7191	0.0428	0.7004	0.0204	7.25	2.67	1.2738E−12	2.9015E−05
G	0.9135	0.9935	0.5688	0.0400	0.5833	0.0291	4.75	0.75	2.4532E−08	4.6570E−04
GAL	0.9730	0.9987	0.6175	0.0471	0.5464	0.0153	6.42	2.43	2.6957E−10	1.2387E−04
HEC	0.9503	0.9965	0.5939	0.0517	0.5769	0.0228	4.50	1.62	5.7119E−09	2.8087E−04
HER	0.9865	0.9995	0.6999	0.0262	0.7030	0.0168	5.67	1.61	3.2483E−11	4.7772E−05
HF	0.9945	0.9999	0.7207	0.0279	0.7216	0.0026	9.33	3.75	2.2500E−12	2.7805E−05
HOL	0.9931	0.9998	0.7161	0.0241	0.7208	0.0081	7.75	2.53	4.4335E−12	3.1122E−05
LIM	0.9956	0.9999	0.7452	0.0205	0.7454	0.0112	7.50	2.32	8.3602E−13	1.9575E−05
LV	0.9876	0.9996	0.6791	0.0591	0.7014	0.0270	5.25	1.82	4.9785E−11	7.0738E−05
MAR	0.9622	0.9977	0.6481	0.0321	0.5490	0.0348	4.75	1.29	8.4960E−10	1.3585E−04
MRY	0.9892	0.9997	0.7086	0.0281	0.7500	0.0195	5.83	1.85	1.4346E−11	4.0911E−05
SH	0.9571	0.9972	0.6239	0.0372	0.6095	0.0130	4.33	0.89	2.3307E−09	1.7690E−04
VRB	0.9947	0.9998	0.7233	0.0330	0.7262	0.0199	7.17	3.21	1.3893 E−12	2.6739E−05
WAG	0.9423	0.9956	0.6136	0.0344	0.6667	0.0304	4.58	1.00	2.7596E−09	1.8737E−04
WAL	0.9781	0.9991	0.6470	0.0412	0.6593	0.0204	5.83	1.95	1.4694E−11	4.4248E−05

PE1 power of exclusion in the absence of genetic information from the other parent, *PE2* power of exclusion if the DNA type of one known parent is available, *HE* expected heterozygosity, *HE (sd)* standard deviation HE, *HO* observed heterozygosity, *HO (sd)* standard deviation HO, # *Alleles* average number of alleles, # *Alleles (sd)* standard deviation # alleles, *HW* P_{ID} Hardy–Weinberg probability of identity, *sib* P_{ID} sibling probability of identity

The accuracy of assigning individuals to the breed of origin based on genotype data was studied using individual assignment tests, implemented in the program GeneClass v.2.0 g [18]. The program includes several assignment methods, but only the Bayesian statistical approach was applied due to its known efficacy [19]. Assignment success was assessed using at least 15 individuals per breed, all having a complete 16-locus genotype.

Results and discussion

Assuming one known parent, the 12 loci recognized as a minimum standard by ISAG had a PE1 for all breeds combined of 0.9777. Assuming two known parents, the 12 loci exhibited a PE2 of 0.9987. When using 16 loci, PE1 increased to 0.9936 and PE2 to 0.9998. Based on the 12 ISAG loci, the major breed of this study, i.e., Holstein Friesian, exhibited a PE1 of 0.9945 and a PE2 of 0.9999 (Table 2). Using 16 loci, Holstein Friesians showed a PE1 of 0.9989 and a PE2 of 0.9999 (Table 3). Furthermore, based on the 12 ISAG-recommended loci, the breed with the lowest PE1 (the local Dutch breed Groningen Whiteheaded) had a PE1 of only 0.9135 and PE2 of 0.9935 (Table 2). With the additional four loci, PE1 in this breed increased to 0.9818 and PE2 to 0.9996 (Table 3). Notably, when using 12 loci, only four breeds reached a PE2 of 0.9999 (Table 2). However, when using 16 loci, only five of the 20 breeds exhibited a PE2 lower than 0.9999 (Table 3).

The HE for 12 loci was lowest in the Groningen Whiteheaded (0.57, Table 2), and the HE for 16 loci was lowest in the Heck Cattle breed (0.59, Table 3). The HO for 12 and for 16 loci was lowest in Galloway Cattle (0.55 and 0.56, respectively; Tables 2 and 3). The average number of alleles calculated for 12 loci was lowest in the Scottisch Highlander (4.33, Table 2) and for 16 loci the average number of alleles was lowest in the Heck Cattle breed (4.25, Table 3).

Both the overall HW P_{ID} and sib P_{ID} across the 12 ISAG-recognized loci were lowest for Blonde D'Aquitaine (5.80E−13 and 1.91E−05, respectively; Table 2) and highest for Groningen Whiteheaded (2.45E−08 and 4.66E−04, respectively; Table 2). Both the overall HW P_{ID} and sib P_{ID} based on the 16 loci were lowest for Blonde D'aquitaine (3.61E−17 and 4.64E−07, respectively; Table 3) and highest for Heck Cattle (1.69E−11 and 2.12E−05, respectively; Table 3). The results for the two P_{ID} calculation methods hence had a close association, but the HW P_{ID}-based values

Table 3 Statistical parameters of 20 cattle breeds based on 16 microsatellite loci

Breed	PE1	PE2	HE	HE (sd)	HO	HO (sd)	# Alleles	# Alleles (sd)	HW P_{ID}	sib P_{ID}
BA	0.9994	0.9999	0.7445	0.0225	0.7492	0.0084	7.38	2.39	3.6135E−17	4.6444E−07
BBL	0.9988	0.9999	0.7112	0.0353	0.7169	0.0158	6.88	1.86	9.3268E−16	1.2303E−06
BRR	0.9950	0.9999	0.6675	0.0333	0.6677	0.0184	5.81	1.76	6.2547E−14	3.7793E−06
CHL	0.9980	0.9999	0.6819	0.0437	0.6920	0.0218	6.44	2.16	1.8725E−15	1.8179E−06
DEX	0.9982	0.9999	0.7198	0.0218	0.6985	0.0055	7.06	1.88	9.7739E−16	9.0543E−07
FH	0.9994	0.9999	0.7346	0.0328	0.6830	0.0179	7.13	2.33	7.2897E−17	6.5354E−07
G	0.9818	0.9996	0.6082	0.0357	0.6198	0.0248	4.94	0.85	4.3745E−12	1.4586E−05
GAL	0.9928	0.9999	0.6280	0.0385	0.5589	0.0132	6.63	2.19	9.1643E−14	4.9362E−06
HEC	0.9790	0.9993	0.5875	0.0454	0.5673	0.0198	4.25	1.57	1.6876E−11	2.1152E−05
HER	0.9979	0.9999	0.7143	0.0217	0.7137	0.0144	5.88	1.50	2.2681E−15	1.1730E−06
HF	0.9989	0.9999	0.7221	0.0217	0.7223	0.0022	9.19	3.27	3.4939E−16	8.1949E−07
HOL	0.9983	0.9999	0.7156	0.0195	0.7175	0.0071	7.69	2.33	9.8730E−16	1.0128E−06
LIM	0.9991	0.9999	0.7391	0.0176	0.7297	0.0099	7.50	2.07	1.4851E−16	6.3615E−07
LV	0.9965	0.9999	0.6825	0.0447	0.6849	0.0237	5.19	1.60	2.6860E−14	2.8644E−06
MAR	0.9902	0.9998	0.6554	0.0277	0.5882	0.0298	5.06	1.29	2.8239E−13	5.5714E−06
MRY	0.9982	0.9999	0.7235	0.0220	0.7637	0.0166	6.00	1.71	1.1834E−15	9.8974E−07
SH	0.9820	0.9995	0.6101	0.0348	0.5900	0.0113	4.38	0.81	6.0064E−12	1.3015E−05
VRB	0.9992	0.9999	0.7344	0.0257	0.7247	0.0172	7.19	2.81	8.7031E−17	6.3130E−07
WAG	0.9745	0.9991	0.6095	0.0271	0.6469	0.0267	4.38	0.96	7.6389E−12	1.3666E−05
WAL	0.9947	0.9999	0.6500	0.0371	0.6708	0.0175	5.94	1.84	2.8006E−15	1.5663E−06

PE1 power of exclusion in the absence of genetic information from the other parent, *PE2* power of exclusion if the DNA type of one known parent is available, *HE* expected heterozygosity, *HE (sd)* standard deviation HE, *HO* observed heterozygosity, *HO (sd)* standard deviation HO, # *Alleles* average number of alleles, # *Alleles (sd)* standard deviation # alleles, *HW* P_{ID} Hardy–Weinberg probability of identity, *sib* P_{ID} sibling probability of identity

Table 4 Power of exclusion in presence of the DNA type of one known parent (PE2) of each microsatellite locus for 20 cattle breeds

	BM	BM	BM	CSRM	CSSM	ETH	ETH	ETH	HAUT	ILSTS	INRA	SPS	TGLA	TGLA	TGLA	TGLA
Breed	1818	1824	2113	60	66	10	225	3	27	006	023	115	122	126	227	53
BA	0.428	0.477	0.692	0.482	0.703	0.482	0.540	0.446	0.498	0.531	0.579	0.313	0.660	0.387	0.637	0.712
BBL	0.286	0.405	0.561	0.520	0.704	0.401	0.504	0.650	0.593	0.443	0.628	0.156	0.491	0.352	0.584	0.664
BRR	0.297	0.369	0.505	0.221	0.533	0.394	0.327	0.230	0.547	0.516	0.434	0.323	0.448	0.549	0.643	0.649
CHL	0.431	0.351	0.600	0.549	0.490	0.121	0.350	0.472	0.667	0.615	0.497	0.279	0.447	0.212	0.629	0.742
DEX	0.551	0.460	0.465	0.467	0.542	0.503	0.454	0.485	0.600	0.523	0.590	0.582	0.251	0.377	0.558	0.541
FH	0.469	0.420	0.610	0.609	0.533	0.565	0.446	0.456	0.606	0.507	0.660	0.204	0.592	0.312	0.660	0.778
G	0.341	0.326	0.399	0.430	0.626	0.423	0.330	0.435	0.503	0.419	0.326	0.102	0.300	0.352	0.456	0.257
GAL	0.209	0.486	0.306	0.295	0.568	0.220	0.540	0.207	0.540	0.385	0.599	0.504	0.280	0.280	0.676	0.515
HEC	0.249	0.133	0.327	0.436	0.411	0.365	0.347	0.413	0.106	0.358	0.403	0.106	0.593	0.324	0.458	0.581
HER	0.347	0.382	0.615	0.461	0.684	0.429	0.575	0.208	0.533	0.483	0.331	0.557	0.486	0.474	0.542	0.548
HF	0.306	0.524	0.535	0.448	0.561	0.428	0.474	0.402	0.584	0.418	0.540	0.373	0.625	0.398	0.708	0.749
HOL	0.341	0.484	0.489	0.416	0.551	0.467	0.478	0.411	0.571	0.396	0.512	0.386	0.662	0.366	0.653	0.688
LIM	0.440	0.379	0.609	0.452	0.676	0.561	0.451	0.444	0.432	0.449	0.588	0.481	0.639	0.418	0.644	0.686
LV	0.554	0.474	0.546	0.372	0.349	0.584	0.402	0.208	0.466	0.539	0.519	0.055	0.611	0.466	0.477	0.592
MAR	0.475	0.188	0.459	0.661	0.387	0.300	0.458	0.295	0.347	0.375	0.513	0.240	0.277	0.546	0.466	0.440
MRY	0.392	0.402	0.613	0.554	0.569	0.392	0.366	0.551	0.510	0.527	0.350	0.285	0.568	0.453	0.622	0.667
SH	0.504	0.143	0.513	0.163	0.427	0.337	0.525	0.356	0.413	0.371	0.284	0.343	0.356	0.285	0.519	0.372
VRB	0.379	0.384	0.664	0.583	0.620	0.313	0.525	0.468	0.545	0.467	0.609	0.291	0.636	0.433	0.553	0.778
WAG	0.218	0.320	0.444	0.296	0.312	0.200	0.311	0.379	0.434	0.300	0.215	0.532	0.400	0.268	0.466	0.492
WAL	0.343	0.483	0.375	0.207	0.610	0.141	0.405	0.360	0.438	0.555	0.466	0.423	0.382	0.311	0.649	0.708
Average	0.378	0.380	0.516	0.431	0.543	0.381	0.440	0.394	0.497	0.459	0.482	0.327	0.485	0.378	0.580	0.608

Table 5 Observed heterozygosity of each microsatellite locus for 20 cattle breeds

	BM	BM	BM	CSRM	CSSM	ETH	ETH	ETH	HAUT	ILSTS	INRA	SPS	TGLA	TGLA	TGLA	TGLA
Locus	1818	1824	2113	60	66	10	225	3	27	006	023	115	122	126	227	53
BA	0.61	0.68	0.83	0.67	0.83	0.68	0.72	0.64	0.68	0.71	0.75	0.49	0.81	0.59	0.79	0.84
BBL	0.45	0.61	0.74	0.70	0.83	0.60	0.70	0.80	0.76	0.63	0.78	0.29	0.67	0.55	0.75	0.81
BRR	0.50	0.57	0.69	0.37	0.71	0.60	0.50	0.38	0.73	0.70	0.62	0.53	0.64	0.73	0.80	0.80
CHL	0.63	0.55	0.76	0.73	0.67	0.22	0.55	0.66	0.81	0.78	0.68	0.45	0.63	0.37	0.78	0.86
DEX	0.73	0.66	0.66	0.67	0.72	0.70	0.66	0.66	0.77	0.71	0.75	0.75	0.40	0.59	0.73	0.72
FH	0.66	0.63	0.77	0.77	0.71	0.74	0.63	0.64	0.77	0.69	0.81	0.35	0.76	0.49	0.81	0.88
G	0.53	0.50	0.60	0.62	0.79	0.61	0.53	0.64	0.69	0.60	0.50	0.19	0.49	0.54	0.65	0.42
GAL	0.39	0.69	0.47	0.46	0.74	0.38	0.72	0.35	0.72	0.57	0.76	0.69	0.44	0.45	0.82	0.69
HEC	0.44	0.24	0.53	0.64	0.61	0.59	0.55	0.61	0.21	0.58	0.61	0.21	0.76	0.50	0.64	0.75
HER	0.55	0.59	0.77	0.65	0.82	0.63	0.75	0.40	0.71	0.67	0.52	0.74	0.67	0.67	0.72	0.73
HF	0.50	0.71	0.72	0.63	0.73	0.60	0.67	0.59	0.76	0.62	0.72	0.57	0.78	0.59	0.84	0.86
HOL	0.54	0.68	0.68	0.60	0.73	0.65	0.67	0.61	0.75	0.60	0.70	0.57	0.81	0.56	0.80	0.82
LIM	0.62	0.59	0.77	0.64	0.81	0.74	0.65	0.63	0.62	0.64	0.76	0.67	0.79	0.61	0.79	0.82
LV	0.74	0.68	0.73	0.58	0.54	0.75	0.60	0.37	0.66	0.72	0.70	0.11	0.77	0.66	0.66	0.76
MAR	0.66	0.38	0.65	0.81	0.59	0.48	0.64	0.48	0.53	0.55	0.71	0.41	0.47	0.73	0.67	0.62
MRY	0.59	0.61	0.78	0.74	0.74	0.59	0.55	0.73	0.70	0.71	0.55	0.48	0.74	0.65	0.78	0.81
SH	0.69	0.26	0.70	0.28	0.63	0.55	0.71	0.55	0.61	0.57	0.48	0.55	0.54	0.48	0.71	0.58
VRB	0.58	0.60	0.81	0.76	0.78	0.50	0.71	0.65	0.72	0.65	0.77	0.46	0.79	0.63	0.72	0.88
WAG	0.41	0.52	0.64	0.47	0.50	0.35	0.51	0.57	0.64	0.50	0.38	0.72	0.61	0.45	0.65	0.68
WAL	0.52	0.68	0.55	0.36	0.76	0.27	0.59	0.56	0.62	0.74	0.66	0.62	0.59	0.52	0.80	0.84
Average	0.57	0.57	0.69	0.61	0.71	0.56	0.63	0.58	0.67	0.65	0.66	0.49	0.66	0.57	0.74	0.76

Table 6 Hardy–Weinberg probability of identity of each microsatellite locus for 20 cattle breeds

	BM	BM	BM	CSRM	CSSM	ETH	ETH	ETH	HAUT	ILSTS	INRA	SPS	TGLA	TGLA	TGLA	TGLA
Breed	1818	1824	2113	60	66	10	225	3	27	006	023	115	122	126	227	53
BA	0.156	0.123	0.042	0.126	0.040	0.124	0.094	0.138	0.121	0.103	0.079	0.247	0.053	0.192	0.057	0.040
BBL	0.301	0.176	0.083	0.113	0.040	0.176	0.110	0.054	0.078	0.154	0.063	0.534	0.130	0.211	0.079	0.052
BRR	0.251	0.190	0.111	0.397	0.110	0.168	0.292	0.397	0.096	0.099	0.153	0.238	0.142	0.098	0.056	0.052
CHL	0.170	0.183	0.065	0.088	0.116	0.528	0.186	0.138	0.058	0.057	0.109	0.260	0.134	0.422	0.045	0.029
DEX	0.098	0.134	0.132	0.126	0.099	0.113	0.132	0.132	0.068	0.108	0.072	0.074	0.375	0.182	0.091	0.095
FH	0.127	0.161	0.069	0.072	0.095	0.083	0.149	0.145	0.069	0.122	0.053	0.406	0.078	0.259	0.050	0.023
G	0.226	0.196	0.203	0.137	0.066	0.184	0.254	0.141	0.111	0.175	0.320	0.603	0.261	0.212	0.125	0.310
GAL	0.356	0.121	0.234	0.256	0.073	0.329	0.092	0.398	0.108	0.168	0.063	0.110	0.256	0.264	0.044	0.107
HEC	0.310	0.570	0.229	0.150	0.166	0.190	0.218	0.165	0.603	0.196	0.170	0.603	0.073	0.247	0.141	0.079
HER	0.213	0.178	0.065	0.135	0.045	0.147	0.081	0.349	0.094	0.122	0.238	0.083	0.124	0.130	0.100	0.100
HF	0.253	0.101	0.099	0.145	0.087	0.166	0.131	0.179	0.077	0.160	0.095	0.198	0.063	0.175	0.038	0.028
HOL	0.219	0.112	0.112	0.166	0.088	0.136	0.139	0.180	0.089	0.173	0.100	0.200	0.052	0.204	0.054	0.041
LIM	0.165	0.190	0.066	0.153	0.047	0.081	0.137	0.152	0.168	0.147	0.079	0.130	0.056	0.175	0.056	0.042
LV	0.088	0.130	0.096	0.191	0.220	0.082	0.187	0.397	0.133	0.097	0.111	0.838	0.070	0.132	0.130	0.067
MAR	0.105	0.376	0.114	0.051	0.182	0.292	0.112	0.250	0.205	0.174	0.108	0.340	0.290	0.100	0.125	0.172
MRY	0.187	0.164	0.066	0.091	0.083	0.188	0.191	0.089	0.113	0.097	0.213	0.273	0.079	0.138	0.072	0.048
SH	0.103	0.608	0.118	0.502	0.155	0.203	0.110	0.201	0.171	0.194	0.280	0.202	0.196	0.270	0.105	0.221
VRB	0.196	0.174	0.051	0.079	0.069	0.261	0.096	0.121	0.092	0.124	0.063	0.264	0.066	0.138	0.088	0.020
WAG	0.330	0.228	0.118	0.241	0.207	0.327	0.230	0.207	0.149	0.372	0.318	0.092	0.166	0.244	0.132	0.127
WAL	0.219	0.133	0.105	0.337	0.063	0.280	0.149	0.151	0.097	0.093	0.073	0.124	0.161	0.231	0.049	0.047
Average	0.204	0.213	0.109	0.178	0.103	0.203	0.155	0.199	0.135	0.147	0.138	0.291	0.141	0.201	0.082	0.085

were much lower than the more conservative sib P_{ID}. Furthermore, the P_{ID} based on 16 loci was much more powerful than that based on only 12 loci (Tables 2 and 3).

Regarding the individual loci, the average PE2 and HO, combined for all breeds, was lowest for the locus SPS115 (0.33 and 0.49, respectively; Tables 4 and 5), and the HW P_{ID} was highest for this locus (0.29; Table 6). The PE2 and HO were highest for the locus TGLA53 (0.61 and 0.76, respectively; Tables 4 and 5), whereas the HW P_{ID} was lowest for TGLA227 (0.08; Table 6).

Alleles were identified and grouped according to their estimated fragment length using fixed allelic bins encompassing one base pair. In order to improve forensic standards in cattle, the population data presented in this study provide extensive information on a selection of loci used by many laboratories, comparing the power of the 12 ISAG-recognized loci with a set containing 16 loci. The 12 loci recognized by ISAG are, in several cases, insufficient for a reliable analysis. Therefore, it is recommended that the number of recognized loci by ISAG should increase. On average, the 16-locus system used in this study is sufficiently powerful for resolving forensic parentage and kinship analysis, as the PE2 value reached a minimum of 0.9993 in all breeds. Selection of the 12 loci recommended as a minimum set by the ISAG was a long process, involving several rounds of interlaboratory comparison tests, which started during the 1990s by selecting the first nine followed by the selection of three additional markers in 2008. Following the eventual recommendations of the ISAG, the loci have been extensively used, and large databases have been generated. Hence, while it could be argued that some of the 12 loci display low polymorphism in some important breeds (e.g., TGLA122 in Dexter Cattle or SPS115 in Belgian Blue), it is not a realistic nor a good recommendation to replace any of the 12 loci. Instead, it is best to increase the number of loci to attain statistical values suitable for forensic applications, as has been done in this study.

Based on data from parentage verification cases, null alleles were identified in the Blonde D'Aquitaine breed for the locus INRA023 (frequency 0.036). For the locus HAUT27, null alleles were found in Blonde D'Aquitaine (0.006), Charolais (0.014), Dutch Friesian (0.007), Holstein Friesian (0.0002), and in the Limousin breeds (0.006). The presence of null alleles in loci (INRA023 and HAUT27) is undesirable for legal case work. Because INRA023 is one of ISAG-recognized loci, it is not recommendable to replace this locus, but instead, the null alleles should be sequenced and primers modified to prevent the problem. Based on the results of this study, replacement of the locus HAUT27 could be considered.

The ISAG loci were not initially selected based on large-scale scientific studies dealing with their informativeness. Therefore, reporting such data for all relevant breeds is important. Only after reviewing such data can the statistical power of the analysis be objectively assessed by the forensic community. While we provide here results for 20 cattle breeds, some important breeds are still missing from our data set (e.g., Angler, Ayrshire, Brown Swiss, Guernsey, Jersey, Montbeliarde, and Simmental), and their inclusion in later studies is warranted.

Statistical methods that estimate the probability of sampling identical genotypes using theoretical equations generally assume random associations between alleles within and among loci. These calculations may be subject to some inaccuracy for some animal populations due to population substructure. In any case, the probability of identity estimates calculated in this study confirm that the 16 loci should provide sufficient information for forensic analyses involving the identity of individual samples.

When using a data set consisting of complete 16-locus genotypes, the overall proportion of individuals correctly assigned to a breed was 88.2%. The individual assignment tests provided excellent assignment success in some of the breeds, and a total of 11 cattle breeds had >90% success in assignment of individuals into their correct reference populations, of which four breeds (Dexter, Heck Cattle, Hereford, and Wagyu breeds) showed 100% assignment success (Table 7). If similar power was also a reality in all other breeds, individual assignment tests could prove very

Table 7 Individual assignment success reached in the cattle breeds using 16 loci

Breed	Assignment success (%)
BA	98.2
BBL	96.1
BRR	77.3
CHL	85.7
DEX	100
FH	73.8
G	95.8
GAL	83
HEC	100
HER	100
HF	54.9
HOL	67.7
LIM	97.6
LV	95.8
MAR	82.4
MRY	87.8
SH	97.5
VRB	76.2
WAG	100
WAL	93.3

useful for numerous forensic applications. For example, they could be applied for determining the breed of origin of a meat or dairy product, as has been attempted in some earlier studies [20, 21]. However, based on our large-scale data, assignment success is too low for most breeds for such applications when using 16 polymorphic loci. In concordance with simulation studies [19], assignment success in cattle is clearly improved when increasing the number of loci from 12 to 16 (data not shown).

The routine set of 16 STR loci of our study may be the choice for routine forensic casework, as they are easy to type and display short fragment lengths, a fact that is important for the analysis of genomic DNA. The forensic usefulness of the cattle STR loci was further advanced by the recent publication of a repeat-based nomenclature system for the 16 bovine STR loci [6] as advocated by the ISFG recommendations. Finally, it is necessary to generate a community-wide allelic ladder for the 16 STR loci; this can be accomplished based on the allele sequence data and the repeat-based nomenclature.

References

1. Adelson DL (2008) Insights and applications from sequencing the bovine genome. Reprod Fertil Dev 20:54–60
2. van de Goor LHP, van Haeringen WA (2007) Identification of stolen cattle using 22 microsatellite markers. Nonhuman DNA typing, theory and casework applications. CRC, Boca Raton, FL, pp 122–123
3. Bredbacka P, Koskinen MT (1999) Microsatellite panels suggested for parentage testing in cattle: informativeness revealed in Finnish Ayrshire and Holstein Friesian populations. Agri Food Sci Finland 8:233–237
4. Sherman GB, Kachman SD, Hungerford LL et al (2004) Impact of candidate sire number and sire relatedness on DNA polymorphism-based measures of exclusion probability and probability of unambiguous parentage. Anim Genet 35:220–226
5. Budowle B, Garofano P, Hellman A et al (2005) Recommendations for animal DNA forensic and identity testing. Int J Legal Med 119:295–302
6. Goor van de LHP, Panneman H, Haeringen van WA (2009) A proposal for standardization in forensic bovine DNA typing: allele nomenclature of 16 cattle-specific STR loci. Anim Genet (in press)
7. Menotti-Raymond MA, David VA, Wachter LL et al (2005) An STR forensic typing system for genetic individualization of domestic cat (*Felis catus*) samples. J Forensic Sci 50:1061–1070
8. Coomber N, David VA, O'Brien SJ et al (2007) Validation of a short tandem repeat multiplex typing system for genetic individualization of domestic cat samples. Croat Med J 48:547–555
9. Eichmann C, Berger B, Parson W (2004) A proposed nomenclature for 15 canine-specific polymorphic STR loci for forensic purposes. Int J Legal Med 118:249–266
10. Hellmann AP, Rohleder U, Eichmann C et al (2006) A proposal for standardization in forensic canine DNA typing: allele nomenclature of six canine-specific STR loci. J Forensic Sci 51:274–281
11. van Asch B, Alves C, Gusmão L et al (2009) A new autosomal STR nineplex for canine identification and parentage testing. Electrophoresis 30:417–423
12. Glowatzki-Mullis ML, Gaillard C, Wigger G et al (1995) Microsatellite-based parentage control in cattle. Anim Genet 26:7–12
13. Peelman LJ, Mortiaux F, Van Zeveren A et al (1998) Evaluation of the genetic variability of 23 bovine microsatellite markers in four Belgian cattle breeds. Anim Genet 29:161–167
14. Schnabel RD, Ward TJ, Der JN (2000) Validation of 15 microsatellites for parentage testing in North American bison, *Bison bison* and domestic cattle. Anim Genet 31:360–366
15. Marshall TC, Slate J, Kruuk L et al (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655
16. Park SDE (2001) Trypanotolerance in west African cattle and the population genetic effects of selection. PhD thesis, University of Dublin
17. Wilberg MJ, Dreher BP (2005) GENECAP software, department of fisheries and wildlife, Michigan State University
18. Piry S, Alapetite A, Cornuet JM et al (2004) GeneClass2: a software for genetic assignment and first-generation migrant detection. J Hered 95:536–539
19. Cornuet JM, Piry S, Luikart G et al (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 153:1989–2000
20. Ciampolini R, Cetica V, Ciani E et al (2006) Statistical analysis of individual assignment tests among four cattle breeds using fifteen STR loci. J Anim Sci 84:11–19
21. Moioli N, Napolitano F, Catillo G (2004) Genetic diversity between piedmontese, maremmana and Podolica cattle creeds. J Hered 95:250–256