GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Effectiveness of microsatellite and single nucleotide polymorphism markers for parentage analysis in European domestic pigs

G.C. Yu*, Q.Z. Tang*, K.R. Long, T.D. Che, M.Z. Li and S.R. Shuai

Institute of Animal Genetics & Breeding,
College of Animal Science & Technology, Sichuan Agricultural University,
Ya'an, China

*These authors contributed equally to this study.
Corresponding author: S.R. Shuai
E-mail: shuaisr2656@sicau.edu.cn

Genet. Mol. Res. 14 (1): 1362-1370 (2015)
Received December 10, 2013
Accepted March 21, 2014
Published February 13, 2015
DOI http://dx.doi.org/10.4238/2015.February.13.15

ABSTRACT. Parentage analysis and individual identification are recent, promising methods that have been applied to evolutionary and ecological studies, as well as conservation management. Parental exclusion relying on polymorphic microsatellites has been used worldwide in parentage determination, while the low mutation rate and genotyping error rate of single nucleotide polymorphisms (SNPs) make them another important marker for pedigree tracing. Here, we compared the effectiveness of microsatellites and SNP markers in European pigs. We also measured and presented the minimum and optimal criteria for SNP markers to be used in paternity and identity analysis. Our findings may contribute to the development of techniques for future molecular evolution and conservation studies, as well as breeding programs.

Key words: Pig; Parentage analysis; Microsatellites; Exclusion power; Single nucleotide polymorphism

INTRODUCTION

Parentage analysis is a promising method for use in evolutionary, ecological, and conservation studies (He et al., 2008; Hill et al., 2008; Shen et al., 2009; Qian et al., 2010; Lu et al., 2011; Ogden et al., 2012), while parentage verification is crucial for optimal genetic management in livestock husbandry and breeding industries. Traditionally, kinship records have been the foundation of breeding programs. However, if such genealogies are incomplete or inaccurate, these data can lead to deviations and errors in analysis that could cause unexpected or stochastic production, leading to reduced economic profits. To achieve stable breeding in the desired direction, molecular markers have therefore been used to distinguish the ancestors of breeding stocks.

Short tandem repeats (STRs, microsatellites) have been the recent and internationally preferred molecular markers to trace studbook information from the species down to the individual level in all types of organisms ranging from lianas (Lacombe et al., 2013) to fish (Liu et al., 2012), insects (Wong et al., 2012), birds (Hadfield et al., 2006; Guerier et al., 2012), and mammals (Cutullé et al., 2010; Stevanovic et al., 2010; Souza et al., 2012). Microsatellites with a high level of heterozygosity are relatively more powerful in assigning relatedness per locus than those with a low level of heterozygosity. However, their high mutation rate (10^{-2}-10^{-5} per generation) (Agrafioti and Stumpf, 2007) and the presence of null alleles (Dakin and Avise, 2004) interfere with accurate pedigree reconstruction.

Single nucleotide polymorphisms (SNPs), which are found throughout the genome, have both a low mutation rate (~2.5 x 10^{-8} per generation) and low genotyping error rate, and they are easily transferable via high-throughput screening (Werner et al., 2004; Baruch and Weller, 2008; Honda et al., 2009). Furthermore, they can be used to investigate both coding and non-coding regions, thus achieving broader genome coverage than microsatellites. Therefore, SNPs offer a readily available method for rapid, large-scale, and cost-effective genotyping, providing an up-to-date and effective origin-tracing system.

In this paper, we applied microsatellites and SNPs to the parentage assignment of pigs and assessed their effectiveness and limitations in an unbiased manner.

MATERIAL AND METHODS

Ethics statement and DNA extraction

The study used 20 female and four male European pigs. The same five pigs were used for the STR and SNP projects, and the additional 19 pigs were used for the SNP-based analysis only. Pigs were allowed free access to food and water under normal conditions, and blood samples were drawn according to the guidelines for the care and use of experimental animals established by the Ministry of Agriculture of China. DNA was extracted from pig whole blood samples using the DNeasy blood and tissue kit (Qiagen, Duesseldorf, Germany) according to manufacturer protocols. The DNA quality was evaluated by 2.5% agarose gel electrophoresis followed by ethidium bromide staining.

Microsatellite polymerase chain reaction (PCR) amplification

For microsatellite amplification, primer pair sequences were obtained from the MARC network (http://www.marc.usda.gov/genome/swine/swine.html), and the forward primers were

labeled with FAM (5- and 6-carboxy fluorescein) fluorescent dye at the 5'-end. PCRs were performed using a Thermo thermal cycler (Thermo Electron Corporation, Shanghai, China) in a 25-µL reaction mixture containing 1.5 mM $MgCl_2$, 200 µM dNTPs, 0.3 µM primers, 1X PCR buffer, 1 U Taq polymerase (AmpliTaq Gold, ABI) and 100 ng DNA template. The thermal cycling conditions included the following details: an initial denaturation step at 94°C for 10 min; 35 cycles of 94°C for 30 s, 57°-61°C for 60 s and 72°C for 60 s; and a final extension step at 72°C for 10 min.

Microsatellite genotyping and variation

A total of 15 microsatellites were selected for individual identification, and **Table S1** lists loci information including repeat motif, bacterial artificial chromosome (BAC) clone identification, BAC-end sequence accession number, forward and reverse primer sequences, number of alleles observed, PCR product sizes, position on the BAC fingerprint map, and alignment coordinates of the BAC-end sequence in the pig genome. PCR products were separated by capillary electrophoresis, and data were subsequently analyzed using the proprietary GeneMapper v4.0 software (Applied Biosystems, Foster City, CA, USA). All alleles were scored manually using the GENOTYPER version 2.5.2 program.

All microsatellite loci were tested for deviation from the Hardy-Weinberg equilibrium (HWE) in FSTAT version 2.9.3.2 using the Markov chain method. The Arlequin version 3.1 software was used to calculate the observed (H_O) and expected heterozygosity (H_E) at each locus. The presence of null alleles was then determined using the Micro-Checker version 2.2.3 software, while the CERVUS 3.0 software was used to calculate the polymorphism information content (PIC).

SNP genotyping and variation

The processing and genotyping of SNPs was carried out using the Illumina PorcineSNP60 BeadChip according to the manufacturer protocol (Infinium II Assay Multi-Sample). The GenomeStudio software (Illumina) was used to visualize, edit, and filter the genotyping data. Raw individual data have a high-genotyping quality (call rate >0.99). Filtered SNPs were further processed using the Plink software as described by Purcell et al. (2007). Briefly, SNPs were removed if they had a GenTrain Score lower than 0.85, Mendelian inconsistencies below 0.01, a minor allele frequency (MAF) under 0.35, were located in sex chromosomes, and were unmapped to the Sscrofa10 assembly or showed position errors in linkage mapping. Basic genetic parameters including HWE, H_O, H_E, and null allele frequency were computed using FASTAT, Alequin, or Micro-Checker as for microsatellites.

Parentage analysis

Parentage analysis of STR and SNP genetic data was consequently performed using CERVUS 3.0. Relationship analysis of microsatellites was carried out using the complete 15 and 12 loci with high PIC values. For SNPs, we sampled seven libraries including 15, 30, 60, 120, 240, 480, and 960 SNPs, with each library containing 10 different randomly sampled repeats, building a total of 70 subsets. CERVUS used a likelihood-based approach to generate likelihood ratios for parentage inference at a relaxed level (80%) and at a strict level (95%). CERVUS was also used to calculate the average non-exclusion probability for one candidate parent (NE-1P),

the average NE-1P given the genotype of a known parent of the opposite sex (NE-2P), and the average non-exclusion probability for a candidate parent pair (NE-PP). The power of exclusion for individuals and the average power of exclusion were computed by equation:

$$PEI = 1 - NEI \text{ and } PEA = PE_A = \sum_{r=1}^{10}(1 - NE_r)/10$$

where r = 1-10 and indicates SNP repeats, respectively. The simulated genotyping error rate was set at 1%, and the total number of simulated offspring was set at 1000 in maternity and paternity assignments.

RESULTS AND DISCUSSION

Microsatellite genotyping

The FASTAT software showed that all populations were likely to be in HWE, and Micro-Checker analysis revealed no evidence of null alleles in any set. The mean proportion of individual pigs in which 15 microsatellites were genotyped was 0.9556. The heterozygosities ranged from 0.500 to 0.889, while the PIC ranged from 0.440 to 0.827. The number of alleles per locus ranged from 4 to 9. The selection of loci with high PIC values (≥0.500) removed loci S0155, S0167, and S0225, leaving a total of 12 loci, as shown in Table 1.

Table 1. Genetic statistics of the 12 microsatellites used in this study.

Locus name	Number of alleles	Allele size (bp)		H_O	H_E	PIC
		Minimum	Maximum			
SW240	7	93	114	0.600	0.800	0.73023
SW72	6	101	115	0.600	0.747	0.67103
S0301	6	250	262	0.700	0.726	0.64313
S0005	7	203	243	0.600	0.858	0.79129
S0101	5	196	224	0.600	0.737	0.65553
S0070	5	261	293	0.700	0.600	0.51863
SW951	4	121	136	0.500	0.711	0.61159
S0090	4	240	253	0.700	0.737	0.64540
SW398	5	166	192	0.600	0.753	0.67709
SW857	9	145	159	0.700	0.889	0.82721
SW886	5	142	174	0.600	0.753	0.67709
S0355	5	245	271	0.500	0.784	0.70379

H_O = observed heterozygosity; H_E = expected heterozygosity; PIC = polymorphism information content.

SNP genotyping

Illumina PorcineSNP60 V2 Genotyping BeadChip scanning detected a total of 61,563 SNP loci. These were filtered using the following criteria: 1) distance between the nearest two loci ≤1 M, 2) call rate ≠100%, and 3) MAF ≤ 0.35. This left a total of 1460 SNPs, and we selected 988 polymorphic loci among all of the samples. Seven libraries were then sampled, including 15, 30, 60, 120, 240, 480, and 960 SNPs, with 10 repeats sampled per library. Procedures similar to those used for microsatellites were applied to all 70 subsets, and all populations were in HWE. As bi-allelic markers, their frequencies were shown to vary enormously. As shown in **Table S2**, the H_E ranged from 0.150 to 0.950, and the PIC ranged from 0.372 to 0.513.

Parentage analysis

Paternity and maternity exclusion are commonly used methods for parentage determination. Parentage analysis has been introduced to livestock culture and breeding industries where it is necessary to carry out infallible and error-free identification of relationships, particularly of half-sibs or full-sibs, because visual identification of relatives is usually impossible or fallible.

In the field of DNA marker-based parentage analysis, microsatellites have traditionally been the marker of choice. For instance, Li et al. (2010) employed 10 polymorphic microsatellite loci to identify paternity in a plateau pika (*Ochotona curzoniae*) population, Wang et al. (2012) successfully traced all offspring back to sole parent-pairs in the *Portunus trituberculatus* breeding project by screening 30 published microsatellites, and Guerier et al. (2012) also acquired accurate genetic relationship information using microsatellite genotyping.

To assess the parentage assignment accuracy, both unconstrained (all possible parents) and constrained (according to pedigree information) conditions were used for microsatellites. The results shown in Table 2 were based on known constraints because more unreliable geneology data would be obtained if the impossible parents were not removed (data not shown).

Table 2. Comparison of the probability of exclusion (PE) among the five shared individuals.

Type of marker	Number of loci	Family line I			Family line II		
		PE-1P	PE-2P	PE-PP	PE-1P	PE-2P	PE-PP
STR	12	**0.9538**	**0.9370**	**0.9454**	**0.8865**	**0.8804**	**0.8835**
SNP	15	0.7547	0.7945	0.7746	0.6969	0.6763	0.6866
	30	**0.9604**	**0.9764**	**0.9684**	**0.8920**	**0.8817**	**0.8868**
	60	0.9985	0.9994	0.9990	0.9906	0.9869	0.9888
	120	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999
	240	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	480	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	960	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

PE-1P = exclusion probability for one candidate parent; PE-2P = exclusion probability for one candidate parent of the opposite sex; PE-PP = exclusion probability for a candidate parent pair; STR = short tandem repeat; SNP = single nucleotide polymorphism.

Both informative (high PIC loci) and complete microsatellite data were analyzed. First, individual identification was carried out using a full set of 15 loci for five test samples including one male and four females. This detected only one family and one case of paternity at 80% confidence. Second, 12 informative loci were selected from the 15 loci, and a similar analysis was performed. As shown in Figure 1, two family lines sharing the same origin of paternity (five specimens in total) were detected in the analysis output, which is in consensus with the historic record of 91.4% reliability.

Nevertheless, parentage identification based on microsatellites can prove to be almost impossible in species with a low heterozygosity unless a large number of polymorphic microsatellite loci can be supplied (Schopen et al., 2008; Tokarska et al., 2009). In addition, microsatellite-based discrimination is significantly weakened if there is a high prevalence of genetic variation and null alleles.

In recent years, SNPs have instead proven to be more sensitive and accurate than microsatellites, and they have become the more favorable alternative marker worldwide (Pakstis et al., 2010; Helyar et al., 2011; Kayser and de Knijff, 2011; Kidd et al., 2011). In 2011, Hauser et al. verified that SNPs contributed to more successful discrimination than microsatellites.

Moreover, Phillips et al. (2012) used small-scale SNPs to improve the statistical power of kinship assignment compared with that of microsatellite data; more recently, Børsting et al. (2013) demonstrated that SNPs increase tracing strengths.

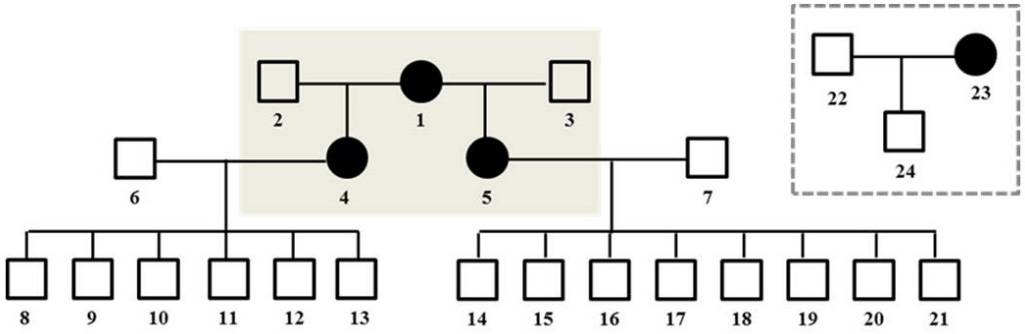

Figure 1. Genetic relationship of the sampled 24 European pigs. The dotted rectangle encloses the pedigree of the five samples shared by microsatellites and SNP projects. The rectangle with gray backgroud stands for a single two-generation family line. Numbers denote individual pigs. Solid blue lines denote candidate father-child relationships, while dotted red lines represent mother-child relationships.

SNP data were further applied to assign parentage for the five individuals whose parentage had already been determined by microsatellites; this identified the same genetic relationship as before. The PE-PPs of the two families were compared based on STRs and SNPs, as shown in Table 2, which demonstrated that the 12 polymorphic STRs and 30 SNPs were similarly effective in individual identification and that the 120 SNPs provided a more effective identity power. This suggests that SNP-based analysis is more effective than STR analysis.

Besides the five validated individuals, the SNP project also included 19 pigs of uncertain origin. Through a series of analyses, two three-generation and one two-generation blood family pedigrees were identified as shown in Figure 1. The SNP analysis was highly consistent with the hypothesis (provided genealogy), and all offspring were successfully assigned to their biological father and mother.

In this article, we explored the minimum and optimal criteria for SNPs in estimating pedigrees with a PE_A value above 90%. As hypothesized, an increase in the number of selected loci resulted in an initial rapid rise in the confidence index followed by slower relative growth until the maximum (100%) was reached (Figure 2). For the 15 loci, the PE_I ranged from 0.593 to 0.861, and the PE_A was 69.4%. Compared with 15 loci, the exclusion power for 30 loci experienced visibly rapid growth, with a PE_I ranging from 0.825 to 0.993, and the PE_A for our set of 30 SNP loci (minimum criteria) was over 90.0%, which was comparable to that of 12 polymorphic STR loci. Moreover, the PE_A grew remarkably after the addition of extra loci: 120 SNP loci could ensure a PE_A of almost 99.9% and accurately revealed genetic relationships (optimal criteria), thus proving to be a reasonable point to trace parental origins.

In conclusion, SNP marker-based individual identification has become an important tool in biochemical and physiological research (Seeb et al., 2011; Goedbloed et al., 2013a,b).

As shown previously and confirmed in the present study, this is because SNPs offer several advantages over microsatellites in parentage analysis (Herráeza et al., 2005; Schopen et

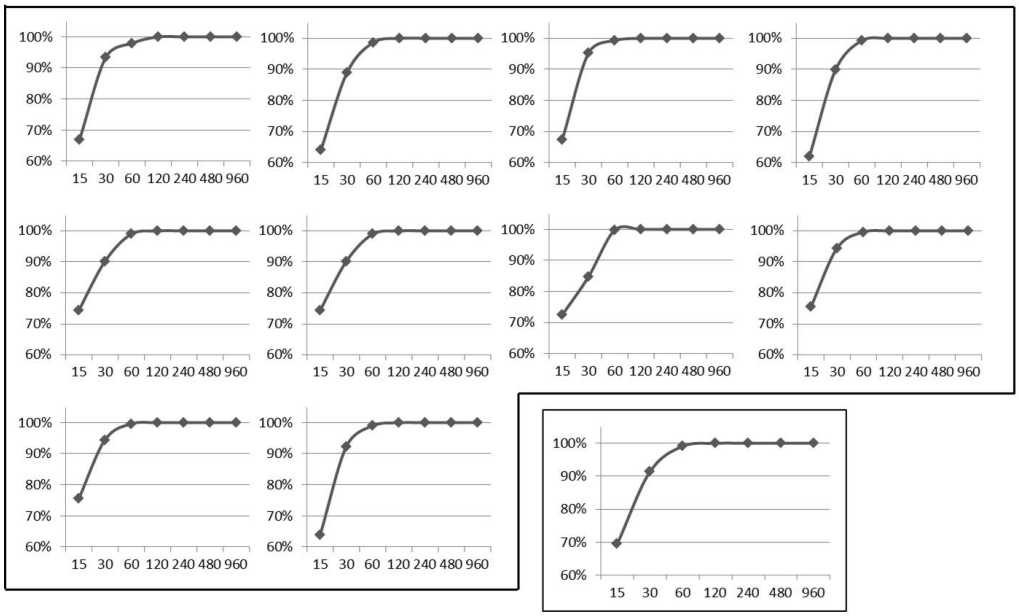

Figure 2. Probabilities of exclusion (PE) for 10 replicates of seven SNP libraries. Ten repeats of 7 libraries each including 15, 30, 60, 120, 240, 480, and 960 SNPs were sampled. Average probabilities of exclusion (PE_A) values for the 10 replicates are shown in the 10 figures respectively within the box; PE_A values of the 10 replicates are boxed separately in the lower right corner. PE values of 17 descendant individuals are shown on the vertical axis, while numbers of SNPs are presented on the horizontal axis.

al., 2008; Gärke et al., 2012). Our research has constructed a robust and accurate tracing system of 120 polymorphic SNPs, which may be used in future conservation and molecular evolution studies.

ACKNOWLEDGMENTS

Research supported by grants from the Fund for Distinguished Young Scientists of Sichuan Province (#2013JQ0013), the Program for Changjiang Scholars and Innovative Research Team in University (#IRT13083), the Specialized Research Fund of the Ministry of Agriculture of China (#NYCYTX-009), and the Project of Provincial Twelfth Five Years' Animal Breeding of Sichuan Province (#2011YZGG15).

Supplementary material

REFERENCES

Agrafioti I and Stumpf MPH (2007). SNPSTR: a database of compound microsatellite-SNP markers. *Nucleic Acids Res.* 35: D71-D75.

Baruch E and Weller JI (2008). Estimation of the number of SNP genetic markers required for parentage verification. *Anim. Genet.* 39: 474-479.

Børsting C, Mogensen HS and Morling N (2013). Forensic genetic SNP typing of low-template DNA and highly degraded

DNA from crime case samples. *Forensic Sci. Int. Genet.* 7: 345-352.
Cutullé C, Jonsson NN and Seddon JM (2010). Multiple paternity in *Rhipicephalus* (*Boophilus*) *microplus* confirmed by microsatellite analysis. *Exp. Appl. Acarol.* 50: 51-58.
Dakin EE and Avise JC (2004). Microsatellite null alleles in parentage analysis. *Heredity* 93: 504-509.
Gärke C, Ytournel F, Bed'hom B, Gut I, et al. (2012). Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. *Anim. Genet.* 43: 419-428.
Goedbloed DJ, Megens H, van Hooft P, Herrero-Medrano JM, et al. (2013a). Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. *Mol. Ecol.* 22: 856-866.
Goedbloed DJ, van Hooft P, Megens HJ, Langenbeck K, et al. (2013b). Reintroductions and genetic introgression from domestic pigs have shaped the genetic population structure of Northwest European wild boar. *BMC Genet.* 14: 43.
Guerier AS, Bishop JM, Crawford SJ, Schmidt-Küntzel A, et al. (2012). Parentage analysis in a managed free ranging population of southern white rhinoceros: genetic diversity, pedigrees and management. *Conserv. Genet.* 13: 811-822.
Hadfield JD, Richardson DS and Burke T (2006). Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Mol. Ecol.* 15: 3715-3730.
Hauser L, Baird M, Hilborn R, Seeb LW, et al. (2011). An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Mol. Ecol. Resour.* 11: 150-161.
He ZX, Zhang JC, Qu KX, Huang BZ, et al. (2008). Application of frozen sperm of Mithun in superovulation of Zebu and parentage identification of crossbred F1 (Mithun x Brahman). *Sci. Agr. Sin.* 41: 2436-2441.
Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, et al. (2011). Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Mol. Ecol. Resour.* 11: 123-136.
Herráeza DL, Schäfer H, Mosner J, Fries HR, et al. (2005). Comparison of microsatellite and single nucleotide polymorphism markers for the genetic analysis of a Galloway cattle population. *Z. Naturforsch. C* 60: 637-643.
Hill WG, Salisbury BA and Webb AJ (2008). Parentage identification using single nucleotide polymorphism genotypes: application to product tracing. *J. Anim. Sci.* 86: 2508-2517.
Honda T, Katsuta T and Mukai F (2009). Simulation study on parentage analysis with SNPs in the Japanese Black Cattle population. *Asian Austral. J. Anim.* 22: 1351-1358.
Kayser M and de Knijff P (2011). Improving human forensics through advances in genetics, genomics and molecular biology. *Nat. Rev. Genet.* 12: 179-192.
Kidd JR, Friedlaender FR, Speed WC, Pakstis AJ, et al. (2011). Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples. *Investig. Genet*. 2: 1.
Lacombe T, Boursiquot JM, Laucou V, de Vecchi-Staraz M, et al. (2013). Large-scale parentage analysis in an extended set of grapevine cultivars (*Vitis vinifera* L.). *Theor. Appl. Genet.* 126: 401-414.
Li K, Geng J, Qu J, Zhang Y, et al. (2010). Effectiveness of 10 polymorphic microsatellite markers for parentage and pedigree analysis in plateau pika (*Ochotona curzoniae*). *BMC Genet*. 11: 101.
Liu P, Xia JH, Lin G, Sun F, et al. (2012). Molecular parentage analysis is essential in breeding Asian seabass. *PloS One* 7: e51142.
Lu X, Wang H, Liu B and Xiang J (2011). An effective method for parentage determination of the clam (*Meretrix meretrix*) based on SSR and COI markers. *Aquaculture* 318: 223-228.
Ogden R, Mellanby RJ, Clements D, Gow AG, et al. (2012). Genetic data from 15 STR loci for forensic individual identification and parentage analyses in UK domestic dogs (*Canis lupus familiaris*). *Forensic Sci. Int. Genet.* 6: e63-e65.
Pakstis AJ, Speed WC, Fang R, Hyland FC, et al. (2010). SNPs for a universal individual identification panel. *Hum. Genet.* 127: 315-324.
Phillips C, García Magariños M, Salas A, Carracedo Á, et al. (2012). SNPs as supplements in simple kinship analysis or as core markers in distant pairwise relationship tests: when do SNPs add value or replace well-established and powerful STR tests? *Transfus. Med. Hemother.* 39: 202-210.
Purcell S, Neale B, Todd-Brown K, Thomas L, et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81: 559-575.
Qian LD, Zhang ZF, Tian YH, Yuan F, et al. (2010). Parentage identification of yellow cattle based on microsatellite DNA markers. *J. Yunnan Agric. Univ.* 1: 69-74.
Schopen GC, Bovenhuis H, Visker M and van Arendonk JA (2008). Comparison of information content for microsatellites and SNPs in poultry and cattle. *Anim. Genet.* 39: 451-453.
Seeb LW, Templin WD, Sato S, Abe S, et al. (2011). Single nucleotide polymorphisms across a species' range: implications for conservation studies of Pacific salmon. *Mol. Ecol. Resour.* 11: 195-217.
Shen T, Lu X and Liu X (2009). Identification of the parentage of corn variety using SSR markers. *Agric. Sci. Technol*.

10: 55-60.
Souza CA, Paiva SR, McManus CM, Azevedo H, et al. (2012). Genetic diversity and assessment of 23 microsatellite markers for parentage testing of Santa Inês hair sheep in Brazil. *Genet. Mol. Res.* 11: 1217-1229.
Stevanovic J, Stanimirovic Z, Dimitrijevic V and Maletic M (2010). Evaluation of 11 microsatellite loci for their use in paternity testing in Yugoslav Pied cattle (YU Simmental cattle). *Czech J. Anim. Sci.* 55: 221-226.
Tokarska M, Marshall T, Kowalczyk R, Wójcik J, et al. (2009). Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. *Heredity* 103: 326-332.
Wang H, Cui Z, Wu D, Guo E, et al. (2012). Application of microsatellite DNA parentage markers in the swimming crab *Portunus trituberculatus*. *Aquac. Int.* 20: 649-656.
Werner FA, Durstewitz V, Habermann FA, Thaller G, et al. (2004). Detection and characterization of SNPs useful for identity control and parentage testing in major European dairy breeds. *Anim. Genet.* 35: 44-49.
Wong J, Chu YY, Stoddard ST, Lee Y, et al. (2012). Microsatellite-based parentage analysis of *Aedes aegypti* (Diptera: Culicidae) using nonlethal DNA sampling. *J. Med. Entomol.* 49: 85-93.

