Open Access
Asian Australas. J. Anim. Sci.
Vol. 27, No. 5 : 616-627 May 2014
http://dx.doi.org/10.5713/ajas.2013.13365
www.ajas.info
pISSN 1011-2367 eISSN 1976-5517

Empirical Selection of Informative Microsatellite Markers within Co-ancestry Pig Populations Is Required for Improving the Individual Assignment Efficiency

Y. H. Li[a], H. P. Chu[1,a], Y. N. Jiang, C. Y. Lin[1], S. H. Li[2], K. T. Li[3], G. J. Weng[4], C. C. Cheng[5], D. J. Lu[6], and Y. T. Ju*
Department of Animal Science and Technology, National Taiwan University, Taipei, Taiwan

ABSTRACT: The Lanyu is a miniature pig breed indigenous to Lanyu Island, Taiwan. It is distantly related to Asian and European pig breeds. It has been inbred to generate two breeds and crossed with Landrace and Duroc to produce two hybrids for laboratory use. Selecting sets of informative genetic markers to track the genetic qualities of laboratory animals and stud stock is an important function of genetic databases. For more than two decades, Lanyu derived breeds of common ancestry and crossbreeds have been used to examine the effectiveness of genetic marker selection and optimal approaches for individual assignment. In this paper, these pigs and the following breeds: Berkshire, Duroc, Landrace and Yorkshire, Meishan and Taoyuan, TLRI Black Pig No. 1, and Kaohsiung Animal Propagation Station Black pig are studied to build a genetic reference database. Nineteen microsatellite markers (loci) provide information on genetic variation and differentiation among studied breeds. High differentiation index (F_{ST}) and Cavalli-Sforza chord distances give genetic differentiation among breeds, including Lanyu's inbred populations. Inbreeding values (F_{IS}) show that Lanyu and its derived inbred breeds have significant loss of heterozygosity. Individual assignment testing of 352 animals was done with different numbers of microsatellite markers in this study. The testing assigned 99% of the animals successfully into their correct reference populations based on 9 to 14 markers ranking D-scores, allelic number, expected heterozygosity (H_E) or F_{ST}, respectively. All miss-assigned individuals came from close lineage Lanyu breeds. To improve individual assignment among close lineage breeds, microsatellite markers selected from Lanyu populations with high polymorphic, heterozygosity, F_{ST} and D-scores were used. Only 6 to 8 markers ranking H_E, F_{ST} or allelic number were required to obtain 99% assignment accuracy. This result suggests empirical examination of assignment-error rates is required if discernible levels of co-ancestry exist. In the reference group, optimum assignment accuracy was achievable achieved through a combination of different markers by ranking the heterozygosity, F_{ST} and allelic number of close lineage populations. (**Key Words:** Microsatellite Markers, Pigs, Assignment Test)

INTRODUCTION

Genetic-structure and lineage-origin analyses in animal populations based on PCR-based molecular approaches are important in DNA tracking, forensics, phylogenetic studies, and the management of population genetics (Caratti et al., 2010; Lowenstein et al., 2010). Hence an appropriate examination system for monitoring and tracing genetic qualities at population and individual level to inform breeding programs is needed. An effective assignment system is essential to accurately manage genetic quality, track gene flow, and allocate individuals to their original populations. Assignment testing based on polymorphism of microsatellite markers has been shown to be useful in genetic characterizations of individuals for classification into different populations (Cornuet et al., 1999; Boitard et al., 2010). Besides assignment methods, assignment accuracy is affected by the number and variability of scored

* Corresponding Author: Y. T. Ju. Tel: +886-2-33664165, Fax: +886-2-27324070, E-mail: ytju@ntu.edu.tw
[1] Taitung Animal Propagation Station, Livestock Research Institute, Taitung, Taiwan.
[2] Department of Life Science, National Taiwan Normal University, Taipei, Taiwan.
[3] Institute of History and Philology, Academia Sinica, Taipei, Taiwan.
[4] Institute of Wildlife Conservation, National Pingtung University of Science and Technology, Pintung, Taiwan.
[5] Graduate Institute of Hakka Cultural Industry, National Pingtung University of Science and Technology, Pintung, Taiwan.
[6] School of Forestry and Resource Conservation, National Taiwan University, Taipei, Taiwan.
[a] These authors equally contribute to this work.
Submitted Jun. 26, 2013; Revised Sept. 12, 2013; Accepted Dec. 12, 2013

markers, degree of genetic differentiation among reference populations, and levels of co-ancestry (Bjørnstad and Røed, 2002; Guinand et al., 2006). Therefore, when assigning genetic origin to closely related animal populations, appropriate criteria and strategies are needed for examining the number and variations of microsatellite markers. In domestic pigs, assignment experiments were applied to paternity controls based on purebreds and crossbreds as reference populations. Detection of exact individuals within closely related populations becomes more difficult because of the decreasing power of assignment of selected microsatellite markers (Kim et al., 2005; Boitard et al., 2010). The rarity of ideal pig populations simultaneously possessing highly inbred and out-bred breeds makes it difficult to evaluate the affect of microsatellite-marker correlation among individuals on the accuracy of individual assignment within highly inbred pig populations.

Landrace, Yorkshire, Duroc, Berkshire, and Meishan pig breeds are used as exotic commercial pig breeds in Taiwan. These breeds have been crossed with indigenous pig breeds including Taoyuan pig and Lanyu pig to obtain heterosis and complementarity. The KHAPS Black pig and the TLRI Black Pig No. 1 were bred for their productivity and black coats, which are popular in Asia. Two founder Lanyu populations were introduced from Lanyu Islet and reared in isolation at the teaching farm of the National Taiwan University (NTU) and the Taitung Animal Propagation Station (TAPS) in 1975 and 1980, respectively. The Lanyu pigs at these locations are termed Lanyu (NTU) and Lanyu (TAPS). The Spotty Lanyu pig breed, characterized by its spotted coat, was inbred and isolated from its parental Lanyu (TAPS) population in 1993. The Binlang Lanyu pig breed, characterized by its uniform white coat, was isolated and insulated from the Spotty Lanyu population in 2001. Hence, the parental Lanyu (TAPS) breed is an ancestor of the Spotty Lanyu pig breed; and the Spotty Lanyu pig breed is an ancestor of the Binlang pig breed. The Mitsai Lanyu miniature pig breed, characterized by a white-and-brown striped coat, was generated by crossing Lanyu (TAPS) with Duroc. The Lee-Sung miniature pig breed was generated by crossing Lanyu (NTU) with Landrace and back to Lanyu (NTU) (Wu et al., 2009). It has been documented that the Lanyu breed possesses a remote genetic relationship with Asian and European type commercial breeds (Jiang et al., 2008; Chang et al., 2009). However, further assessment of genetic distance, genetic diversity, and differentiation among close lineage breeds (Spotty Lanyu and Binlang Lanyu pig) and related synthetic breeds (Mitsai Lanyu and Lee-Sung pig) remains incomplete. The complexity of the relationship among exotic commercial European pigs, Asian pigs, and Lanyu pigs through high levels of inbreeding and synthetic crossbreeding provides an excellent model for examining the criteria needed to select effective microsatellite markers and strategies needed for assigning lineage, especially when co-ancestry increases among closely related individuals.

MATERIALS AND METHODS

Blood sampling and DNA isolation

Genomic DNAs were extracted from blood cells. Blood samples from Lanyu (TAPS), Spotty Lanyu, Binlang Lanyu, and Mitsai Lanyu (50% Lanyu [TAPS] and 50% Duroc) were obtained from the Taitung Animal Propagation Station. Animal handling protocols conformed to those approved by the National Taiwan University Animal Care and Use Committee. Blood samples of Lanyu (NTU) and that of the Lee-Sung pig (75% Lanyu [NTU] and 25% Landrace) were obtained from National Taiwan University teaching farm. The blood of KHAPS Black pig (KHAPS) was obtained from Kaohsiung Animal Propagation Station. The blood samples of Taoyuan, Taiwan Livestock Research Institute Black Pig No. 1 (TLRI Black Pig No. 1), Meishan, Landrace, Yorkshire, Duroc, and Berkshire were obtained from the Taiwan Livestock Research Institute (TLRI). Besides Lanyu (TAPS), most of our samples in each population have recorded pedigrees and unrelated pigs up to 3 generations. Lanyu (TAPS) pigs were bred through natural mating in semi-wild environments. Sample sizes obtained from each population were collected respective of population size and blood relations between individuals. Thirty-two individual Lee-Sung pigs, 39 Lanyu (TAPS), 5 Lanyu (NTU) and 17 Mitsai Lanyu were sampled from respective stock populations. Samples from 30 Landrace, 30 Yorkshire, 30 Duroc, 29 Berkshire, 29 Meishan, 30 Taoyuan, 18 Spotty Lanyu, 21 Binlang Lanyu, 21 TLRI Black Pig No.1 and 21 KHAPS black pigs were selected based on there being no blood relationships among candidates, according to pig-breeding pedigree registrations in Taiwan. The non-related individuals of each population were chosen based on their pedigrees (except for the Lanyu TAPS population); therefore, the number of individuals in each population used in this study varies. Extraction and purification procedures for genomic DNA from blood samples were conducted using slightly modified QIAamp DNA Blood Maxi Kits (Qiagen, Valencia, CA, USA). First, 500 μL QIAGEN protease, 10 mL blood and 12 mL of buffer AL were mixed thoroughly in a 50 mL centrifuge tube. The mixture was then incubated at 70°C for 10 min before 10 mL of absolute ethanol was added and mixed. The solution was then transferred onto QIAamp Maxi column and centrifuged at 1,850×g for 3 min. Buffer AW1 and Buffer AW2 were separately added to the QIAamp Maxi column and centrifuged at 1,850×g for 1 min and 15 min, respectively. To elute the DNA, 1 mL buffer AE was applied to the column and centrifuge at 1,850×g for 5 min.

Table 2 describes the number of individuals in each population and the type of pig breeds.

Microsatellite genotyping

A total of 19 microsatellite markers located on 15 chromosomes were chosen as suggested by the Domestic Animal Diversity Information System of the Food and Agriculture Organization of the United Nations (ISAG-FAO, 2004). The primers used for microsatellite loci amplification were end-labeled with FAM, TET or HEX fluorescent dye. The fragments of microsatellite DNA in the genome were amplified by PCR (PTC-200 Thermal Controller, M. J. Research, Waltham, MA, USA). Each 15 μL reaction volume contained 50 ng of genomic DNA, 1.5 μL of 10×PCR buffer 0.375 μL of 8 mM deoxynucleoside triphosphate, 4.5 pmol sense and antisense primers, and 0.6 units of DNA Taq polymerase (Amersham Biosciences, Arlington Heights, IL, USA). Thermal cycling conditions were as follows: an initial 95°C for 5 min, then 37 cycles at 95°C for 30 s, 48°C to 62°C for 30 s (depending on the locus), and 72°C for 45 s with a final extension at 72°C for 7 min. The fluorescent end-labeled PCR products were detected by capillary electrophoresis with a MegaBACE 1000 DNA sequencer (Amersham Biosciences, Arlington Heights, IL, USA) and a fluorescent-labeled marker ET-400 (Amersham Biosciences, Arlington Heights, IL, USA) was used as an internal size standard for length calibration. The allele lengths of loci were presented by Genetic-Profiler (version 2.2; Amersham Biosciences, Arlington Heights, IL).

Data analysis

The total number of alleles per microsatellite locus, observed and expected heterozygosities of alleles (Nei, 1973), polymorphic information content (PIC) of each locus (Botstein et al., 1980), and deviations from the Hardy-Weinberg equilibrium were calculated with CERVUS version 2.0 software (Marshall, 1998). The mean of the effective number of alleles (MEA) was estimated following the formula provided by Kimura and Crow (1964) and calculated by POPGENE (Yeh et al., 1999). The number of polymorphic loci (NPL) equates to the frequency of the most frequent allele at a locus being <0.95. GENEPOP 4.0 computer package was used to calculate F-statistics (F_{IT}, F_{IS}, and F_{ST}) to understand variation in heterozygosity of each locus and differentiations among populations (Rousset, 2008). To compare the number of alleles between different sample sizes, the allelic richness (AR) of each locus, which measures the number of alleles independent of the sample size, was calculated using the program FSTAT v. 2.9.3 software package (Goudet, 2001). The mean number of alleles per breed, mean observed and expected heterozygosities (Nei, 1973) per breed, and mean polymorphic information content (PIC) per breed (Botstein et al., 1980) were calculated with CERVUS version 3.0 software (Kalinowski et al., 2007). Pairwise genetic distances between sampling populations were calculated by POPULATIONS, software version 1.2.28 (Langella, 2002). A neighbor-joining (NJ) population tree, based on Cavalli-Sforza chord distances (Cavalli-Sforza and Edwards, 1967), was used to represent relationships among groups and a consensus tree was obtained by bootstrapping (1,000 replications) distance values over loci.

Multidimensional scaling (MDS) plots

Genetic distances between 14 populations are displayed in the form of a geometric picture by non-metric multidimensional scaling (MDS) techniques using the program PRIMER (Carr, 1996). Standardized Residual Sum of Squares (STRESS) shows the difference between MDS and raw data. MDS and STRESS methods are described in Krzanowski (1987). STRESS values of <0.2 mean genetic distances given by MDS can be trusted.

Assignment test

Assignment and exclusion testing for individuals into their reference populations is calculated by GeneClass v.2.0 (Piry et al., 2004). In this study, a total of 14 reference populations were defined as within-breeds, crossbreeds (synthetic breeds), or inbred breeds. The principle of assignment and the exclusion method are based on the Bayesian method (Rannala and Mountain, 1997). This method was suggested as a superior method by Cornuet et al. (1999) and Koskinen (2003). The assignment test for each individual is excluded from the dataset when performing own assignment (Efron, 1983). The number of simulated individuals was set 10,000 for Monte Carlo simulations. The criterion for assigning individuals to populations is based on the frequency of simulated genotypes in the reference populations; the degree of confidence is ≤0.001. To ensure an effective success rate in the analysis of the 19 markers and the aforementioned variability criteria, the 19 markers are ranked and assignment testing performed according to the highest allelic number (K), H_E, F_{ST} and relative discriminatory power (D-score) estimated from the 14 populations (Supplementary Table S1, approach 1: See e-version for supplement.) and 6 Lanyu related populations (i.e., Lanyu [TAPS], Lanyu [NTU], Spotty Lanyu, Binlang Lanyu, Lee-Sung and Mitsai Lanyu; Supplementary Table S1, approach 2: See e-version for supplement). The D-score of each locus to assign individuals to source cultivars was calculated with the software WHICHLOCI (Banks et al., 2003). This program ranks microsatellite loci based on their relative allelic differentials as derived from different populations (Banks and Eichert, 2000). As a control, the markers were also numbered from 1 to 19 at random and processed. The

performance of each assignment test was measured both by the small number of microsatellite loci used and the proportion of individuals correctly assigned to their original populations.

RESULTS

Statistics on information richness of the 19 microsatellite markers

The total number of alleles found for the 19 microsatellite markers from the 14 populations, including the 7 breeds, 4 synthetic breeds and 3 inbred breeds (Lanyu, Spotty Lanyu and Binlang Lanyu) was 239. The number of alleles per marker ranged from 7 (SW951, SW911) to 29 (S0005), with a global mean for the 19 markers of 12.6. On the other hand, the PIC values per locus varied from 0.555 (S0215) to 0.879 (S0005), with the mean across the 19 markers being 0.771 (Table 1). The results indicate that these microsatellite markers are polymorphic for genetic structure analysis in this study (Anderson et al., 1993). The Wright's F-statistic values for each marker are shown in Table 1. F_{IS}, the parameter of within-breed deficit in heterozygosity, resulted in a global mean of 0.110 for all loci, and ranged from –0.118 (SW24) to 0.477 (S0386). F_{IT}, the parameter of divergence between expected and observed heterozygosity for all individuals, had a global mean of 0.468 for all loci, and ranged for the different markers from 0.335 (SW24) to 0.686 (S0386). F_{ST}, the parameter for genetic differentiation among breeds, varied from 0.289 (S0005) to 0.502 (S0215), and the average F_{ST} of all loci was 0.402. These data indicate a high degree of differentiation among the populations and further analyses of genetic differentiation and diversity within and among the studied populations are required to evaluate how informative the rankings of microsatellite loci are for individual identification.

Table 1. Summary of observed allele number in each microsatellite marker, polymorphic information content and the Wright's F-statistics of 352 individuals based on the polymorphism of 19 microsatellite markers

Locus	N_A[a]	PIC[b]	F_{IS}[c]	F_{IT}[c]	F_{ST}[c]
SW911	**7**	0.692	–0.016	0.405	0.414
SW951	**7**	0.767	0.095	0.442	0.383
S0215	9	**0.555**	–0.107	0.449	**0.502**
S0155	10	0.737	–0.030	0.409	0.426
SW24	10	0.731	**–0.118**	**0.335**	0.405
S0386	10	0.671	**0.477**	**0.686**	0.400
SW857	11	0.835	0.035	0.344	0.321
IGF1	11	0.783	0.118	0.427	0.351
SW72	11	0.750	0.060	0.349	0.308
S0225	11	0.783	-0.009	0.493	0.497
S0002	12	0.820	0.070	0.477	0.437
S0228	12	0.829	0.276	0.625	0.482
S0355	13	0.804	0.033	0.494	0.477
SW122	14	0.814	0.117	0.385	0.304
S0226	14	0.826	0.090	0.429	0.373
S0068	15	0.858	0.011	0.454	0.448
S0218	15	0.784	0.408	0.639	0.390
S0227	18	0.735	0.203	0.561	0.449
S0005	**29**	**0.879**	0.292	0.496	**0.289**
Mean	12.6	0.771	0.110	0.468	0.402

[a] Number of alleles.
[b] Polymorphic information content.
[c] Wright's F-statistics (F_{IS}, F_{IT}, and F_{ST}).

Genetic diversity within breeds

The level of genetic diversity within each breed was determined by the number of polymorphic loci, mean

Table 2. Locations and characteristic type, summary of statistics of 14 pig populations counted based on the polymorphism of 19 microsatellite markers

Breed	Locality	Type	N[a]	NPL[b]	MPIC[c]	MNA[d]	AR[e]	MEA[f]	MH_O[g]	MH_E[h]	F_{IS}[i]
Lanyu (TAPS)	TAPS	Indigenous	39	19	0.459	3.211	2.362	2.265	0.397	0.531	0.253
Spotty Lanyu	TAPS	Inbreeding	18	19	0.368	3.000	2.097	1.955	0.275	0.433	0.364
Binlang Lanyu	TAPS	Inbreeding	21	19	0.33	2.737	1.980	1.738	0.334	0.385	0.133
Mitsai Lanyu	TAPS	Synthetic	17	18	0.385	3.158	2.173	2.045	0.48	0.448	-0.072
Lanyu (NTU)	NTU	Indigenous	5	13	0.246	2.000	1.820	1.574	0.213	0.292	0.269
Lee-Sung	NTU	Synthetic	32	17	0.364	3.526	2.129	1.959	0.364	0.404	0.098
Taoyuan	TLRI	Indigenous	30	18	0.414	3.105	2.253	2.107	0.45	0.473	0.049
Meishan	TLRI	Asia type	29	19	0.411	3.368	2.255	2.147	0.439	0.470	0.067
Landrace	TLRI	Europe type	30	17	0.464	3.947	2.498	2.529	0.471	0.512	0.079
Yorkshire	TLRI	Europe type	30	19	0.474	3.789	2.466	2.382	0.538	0.531	-0.012
Duroc	TLRI	Europe type	30	19	0.455	3.842	2.427	2.296	0.473	0.511	0.073
Berkshire	TLRI	Europe type	29	17	0.349	3.526	2.075	1.898	0.342	0.394	0.131
TLRI Black pig NO. 1	TLRI	Synthetic	21	19	0.523	4.105	2.698	2.547	0.601	0.572	-0.050
KHAPS Black pig	KAPS	Synthetic	21	19	0.572	5.158	2.920	2.970	0.596	0.624	0.044

[a] Number of individuals. [b] Number of polymorphic loci. [c] Mean polymorphism index content. [d] Mean number of alleles.
[e] Allelic richness. [f] Mean number of effect number of alleles. [g] Mean observed heterozygosity. [h] Mean expected heterozygosity. [i] Fixation indices.

number of alleles, mean polymorphism index content, allelic richness, mean effective number of alleles, expected heterozygosity, observed heterozygosity, and the F_{IS} value (Table 2). The within-breed analyses indicated that the number of polymorphic loci (the frequency of the most frequent allele being <0.95 at one locus) ranged from 13 to 19, the lowest appeared in the Lanyu (NTU) population. Mean number of alleles implies the degree of allelic diversity within breeds; this was greatest in KHAPS (5.158) and least in Lanyu (NTU) (2.000). Mean polymorphism index content ranged from 0.246 (Lanyu [NTU]) to 0.572 (KHAPS), indicating the 19 microsatellite markers used were informative for genetic analyses 'within' populations. Allelic richness measures the number of alleles per locus in a standardized uniform sample size; this was greatest in KHAPS (2.920) and least in Lanyu (NTU) (1.820). Mean effective number of alleles ranged from 1.574 (Lanyu [NTU]) to 2.970 (KHAPS). These numbers were less than the mean number of observed alleles in each population, suggesting these markers were polymorphic for genetic identification within these breeds. Mean observed heterozygosity ranged from 0.213 (Lanyu [NTU]) to 0.601 (TLRI Black pig No. 1), whereas mean expected heterozygosity ranged from 0.292 (Lanyu [NTU]) to 0.624 (KHAPS). F_{IS} were higher in Spotty Lanyu (0.364), Lanyu (TAPS) (0.253), Lanyu (NTU) (0.269), Binlang Lanyu (0.133), and Berkshire (0.131) while a near balance between observed and expected heterozygosity was observed in Taoyuan (0.049), Yorkshire (–0.012), Duroc (0.073), and Landrace (0.079). These data are consistent with the breeding history of breed generations used in this study.

Genetic differentiation among breeds and populations

Cavalli-Sforza chord distances and pairwise F_{ST} estimates in proportion to each pair of the 14 populations are shown in Table 3. The genetic distances ranged from 0.238 (between Spotty Lanyu and Binlang Lanyu) to 0.814 (Taoyuan versus Berkshire). Low genetic distances were detected between Lanyu (TAPS) versus Spotty Lanyu (0.391), and Binlang Lanyu (0.453). The shortest distance (0.238) was identified between Spotty Lanyu and Binlang Lanyu. Short genetic distances were also detected among European pig breeds (range from 0.479 to 0.598). The pairwise F_{ST} coefficients ranged from 0.142 (between Spotty Lanyu and Binlang Lanyu) to 0.568 (between Lanyu [NTU] and Berkshire). High F_{ST} coefficients were detected between several populations, including Lanyu (TAPS) and its derived breeds: Spotty Lanyu (0.210), Binlang Lanyu (0.294) and Lanyu (NTU) (0.283) as well as Lee-Sung versus Lanyu (TAPS) (0.269), and Lanyu (NTU) (0.263). Mid differentiation was detected between Binlang Lanyu and Spotty Lanyu (0.142) as well as TLRI Black pig No. 1 versus Duroc (0.165). High F_{ST} coefficients (>0.25) were identified among commercial breeds: Landrance, Duroc, Yorkshire, Meishan and Taoyuan. Separation from Lanyu (NTU) was greatest with Berkshire (0.568) and Meishan (0.523). For Lee-Sung, it was with Berkshire (0.529). Cavalli-Sforza chord distances based on polymorphisms of the 19 microsatellite markers from the 14 populations were then used to construct a Neighbor-Joining phylogenetic tree. Three major clusters with high bootstrap values (>95%) were obtained in the phylogenetic tree (Figure 1). Cluster I included European pig breeds (Landrace, Yorkshire, Duroc, and Berkshire) and TLRI Black Pig No. 1 breed. Cluster II included Meishan, Taoyuan and KHAPS Black. Lanyu, Lanyu derived breeds (Spotty Lanyu and Binlang Lanyu), Lanyu derived synthetic breeds, Mitsai Lanyu, and Lee-Sung pigs were clustered in Cluster III with a 100% bootstrap value. These results indicate mid- to high-genetic differentiation among these pig breeds.

The relative position of pig populations defined by MDS

Non-metric MDS placed 14 points that represented the

Table 3. Cavalli-Sforza chord distances (Cavalli-Sforza and Edwards, 1967; below the diagonal) and mean F_{ST} estimates (above the diagonal) in proportion to the 14 pig populations calculated based on the polymorphism of the 19 microsatellite markers

Items	Lanyu (TAPS)	Spotty Lanyu	Binlang Lanyu	Mitsai Lanyu	Lanyu (NTU)	Lee-Sung	Landrace	Yorkshire	Berkshire	Duroc	TLRI Black pig NO. 1	Taoyuan	Meishan	KHAPS Black pig
Lanyu (TAPS)		0.210	0.294	0.328	0.283	0.269	0.418	0.373	0.441	0.408	0.386	0.428	0.391	0.330
Spotty Lanyu	0.391		0.142	0.361	0.380	0.344	0.428	0.381	0.473	0.413	0.372	0.486	0.457	0.358
Binlang Lanyu	0.453	0.238		0.423	0.416	0.417	0.467	0.432	0.515	0.454	0.426	0.514	0.509	0.409
Mitsai Lanyu	0.545	0.524	0.560		0.475	0.432	0.378	0.363	0.436	0.355	0.336	0.472	0.429	0.308
Lanyu (NTU)	0.505	0.550	0.546	0.647		0.263	0.491	0.463	0.568	0.477	0.437	0.516	0.523	0.406
Lee-Sung	0.453	0.496	0.550	0.576	0.378		0.473	0.451	0.529	0.456	0.413	0.503	0.484	0.409
Landrace	0.733	0.704	0.742	0.638	0.798	0.734		0.237	0.345	0.262	0.251	0.464	0.462	0.275
Yorkshire	0.685	0.674	0.701	0.644	0.779	0.731	0.479		0.319	0.311	0.296	0.429	0.436	0.319
Berkshire	0.691	0.677	0.708	0.658	0.797	0.724	0.531	0.554		0.382	0.364	0.540	0.506	0.409
Duroc	0.729	0.665	0.697	0.600	0.786	0.720	0.503	0.598	0.586		0.165	0.463	0.422	0.228
TLRI Black pig NO.1	0.736	0.684	0.719	0.611	0.760	0.687	0.559	0.644	0.627	0.446		0.416	0.407	0.205
Taoyuan	0.734	0.764	0.779	0.771	0.769	0.754	0.792	0.767	0.814	0.799	0.718		0.426	0.367
Meishan	0.684	0.731	0.757	0.697	0.810	0.737	0.786	0.764	0.767	0.747	0.750	0.670		0.248
KHAPS Black pig	0.676	0.658	0.690	0.617	0.764	0.705	0.586	0.669	0.696	0.512	0.485	0.687	0.544	

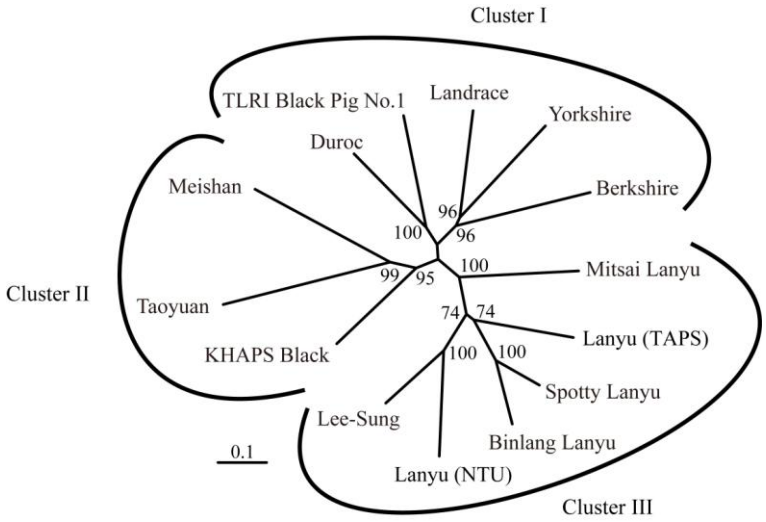

Figure 1. Phylogenetic trees constructed from Cavalli-Sforza Edward chord genetic distances based on the polymorphism of 19 microsatellite markers. The numbers at the branch nodes are the percentages of group occurrence in 1,000 bootstrap replications of resampled loci. The scale bar represents the 0.1 genetic distance of Cavalli-Sforza Edward chord.

Cavalli-Sforza chord distances among the 14 populations in a two-dimensional ordination plot (Figure 2). The STRESS value in Figure 2 was exceptionally low (0.09). This indicates that the diagram gives an excellent representation of the data (Bond et al., 2002). The three clusters are clearly indicated within the ovals in Figure 2. European breeds, TLRI Black Pig No. 1 and KHAPS Black are clustered in one cluster. Asian pig breeds other than Lanyu and Lanyu derived pig breeds are distributed in 2 different clusters. The MDS data correlated with the Neighbor-Joining tree except for the KHAPS Black pig breed that was clustered in the Asian pigs' clade in the NJ tree. These data showed closer genetic relationships among Lanyu and its derived breeds than other breeds showed among their relationships. Accordingly, Lanyu and its derived breeds provide an excellent model for testing the strategy of selecting microsatellite markers for improving assignment testing.

Principles of marker selection versus the accuracy of assignment tests

An effective assignment test can be defined as that test which requires the least number and best combination of microsatellite markers for assigning individuals to their original populations. The efficiency of assignment testing is affected by the level of H_E, allele number (K), F_{ST} or D-score values of microsatellite markers among used

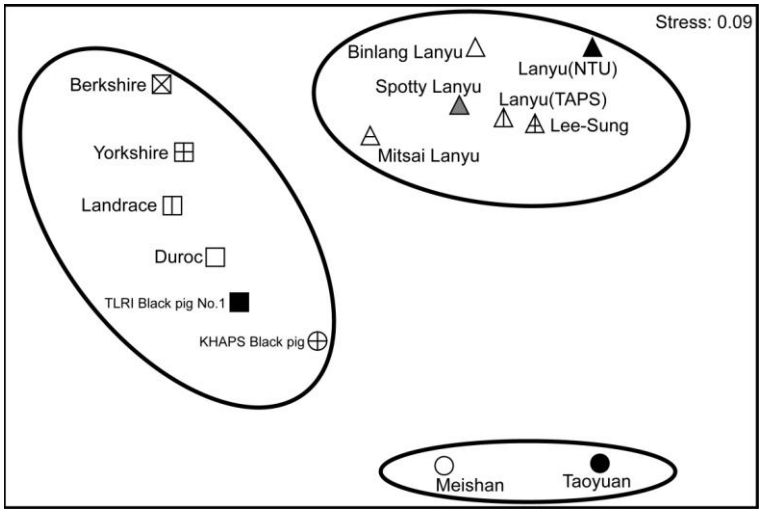

Figure 2. Two dimensional representation of a multi dimensional scaling (MDS) plot showing the relative relationships of 14 pig populations. The circling of the groups is arbitrary. It shows the general pattern. The groups are: Lanyu, Lanyu derived breeds and its synthesis breeds, and Asian type breeds, and European type breeds. STRESS testing indicates the difference between MDS and raw data. 0.09 is very low indicating highly accurate groupings (Krzanowski, 1987).

Table 4. The result of assignment tests by different criteria based on the Bayesian method among 14 pig populations[1]

Items	Number of loci be used for assignment test																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Approach 1																			
F_{ST}	38.6	67.6	83.2	93.8	94.9	96.6	98.0	98.3	98.9	98.3	98.3	98.0	98.9	**99.1**	**99.1**	**99.1**	**99.1**	**99.1**	**99.4**
K	61.6	80.7	91.2	94.3	96.0	97.7	98.9	98.6	98.9	**99.1**	**99.1**	**98.9**	**99.1**	**99.4**	**99.1**	**99.4**	**99.4**	**99.4**	**99.4**
H_E	61.6	86.4	92.9	94.9	96.9	98.6	98.9	98.9	98.9	**99.1**	**99.4**	**99.7**	**99.7**	**99.7**	**99.7**	**99.7**	**99.4**	**99.4**	**99.4**
Dscore	71.9	85.8	92.6	94.3	96.9	98.0	98.6	98.9	**99.4**	**99.4**	**99.1**	**99.7**	**99.4**	**99.4**	**99.4**	**99.4**	**99.4**	**99.4**	**99.4**
Random	35.8	61.6	79.5	91.5	94.9	97.4	98.0	98.3	98.6	98.6	98.6	98.3	98.9	98.9	98.9	**99.4**	**99.4**	**99.4**	**99.4**
Approach 2 (within Lanyu)																			
F_{ST}	47.4	80.1	92.9	95.2	96.9	98.9	**99.1**	**99.1**	**99.1**	**99.1**	**99.4**	**99.1**	**99.4**	**99.1**	**99.1**	**99.4**	**99.4**	**99.4**	**99.4**
K	61.6	85.5	93.5	96.0	97.7	98.3	98.9	**99.4**	**99.7**	**99.7**	**99.7**	**99.7**	**99.7**	**99.7**	**99.4**	**99.4**	**99.4**	**99.4**	**99.4**
H_E	50.6	67.6	89.8	93.8	97.4	**99.1**	98.9	**99.1**	**99.7**	**99.7**	**99.7**	**99.7**	**99.7**	**99.4**	**99.4**	**99.4**	**99.4**	**99.4**	**99.4**
Dscore	51.1	77.0	90.9	96.9	98.6	98.3	98.3	98.3	98.9	**99.4**	**99.4**	**99.4**	**99.4**	**99.4**	**99.4**	**99.4**	**99.4**	**99.4**	**99.4**

[1] The table shows the relation between the percentage of correctly assigned individuals and the number of markers used. Markers were selected and ranked by two approaches, approach 1 ranking by the highest F_{ST}, H_E, K (allelic numbers), D-score, and random ordering from 14 populations; approach 2 (within Lanyu), the markers were selected and ranked from the statistic within 6 Lanyu and its derived lines and breeds, and ranking from the highest F_{ST}, H_E, K, and D-score.

reference populations. This process was evaluated and analyzed for varying conditions based on the Bayesian method (Rannala and Mountain, 1997). The performance of each assignment test was measured by the proportion of individuals correctly assigned to their original populations (Bank et al., 2003).

When random ordering criteria were used to select markers for assignment testing, 99% assignment accuracy was achieved when 7 and 16 markers were chosen, respectively (Table 4). Six markers were enough for commercial pig breeds (Landrace, Yorkshire, Berkshire, Duroc) to obtain 100% assignment accuracy; 5 markers were adequate for TLRI Black pig No.1 and KHAPS Black pig to obtain 100% assignment accuracy; and 2 markers were enough for Taoyuan and Meishan to obtain 100% assignment accuracy (Supplementary Table S2, A: See e-version for supplement.). In addition, according to analysis results from the 14 populations (Supplementary Table S1, approach 1: See e-version for supplement.), marker subsets were selected by ranking all 19 microsatellite markers according to highest H_E, K, $F_{ST,}$ and D-score values to conduct assignment tests. Nine markers with highest D-scores, 10 markers with highest H_E or K and 14 markers with highest F_{ST} were required to achieve 99% assignment accuracy. The commercial pig breeds needed less than 7 markers ranked according to highest H_E, K, $F_{ST,}$ and D-score values from 14 populations to obtain 100% assignment accuracy (Supplementary Table S2 B, C, and D: See e-version for supplement.). After checking row data, most of incorrect individual assignments came from missed assignment among close lineage breeds (Lanyu, Binlang Lanyu and Spotty Lanyu, Table 5; Supplementary Table S2 and S3: See e-version for supplement). This result suggests very close lineage among reference populations may affect assignment accuracy.

To increase the performance of assignment testing, markers were selected by ranking highest H_E, allele number, F_{ST} and D-scores obtained from 6 Lanyu related populations (Lanyu [TAPS], Lanyu [NTU], Spotty Lanyu, Binlang Lanyu, Lee-Sung and Mitsai Lanyu) and conducting

Table 5. Percentage of all individuals assigned from each reference populations based on the 19 microsatellite markers

Items	Lanyu (TAPS)	Spotty Lanyu	Binlang Lanyu	Mitsae Lanyu	Lanyu (NTU)	Lee-Sung	Landrace	Yorkshire	Berkshire	Duroc	TLRI Black pig NO.1	Taoyuan	Meishan	KHAPS Black pig
Lanyu (TAPS)	100	-	-	-	-	-	-	-	-	-	-	-	-	-
Spotty Lanyu	5.56	94.44	-	-	-	-	-	-	-	-	-	-	-	-
Binlang Lanyu	-	4.76	95.24	-	-	-	-	-	-	-	-	-	-	-
Mitsai Lanyu	-	-	-	100	-	-	-	-	-	-	-	-	-	-
Lanyu (NTU)	-	-	-	-	100	-	-	-	-	-	-	-	-	-
Lee-Sung	-	-	-	-	-	100	-	-	-	-	-	-	-	-
Landrace	-	-	-	-	-	-	100	-	-	-	-	-	-	-
Yorkshire	-	-	-	-	-	-	-	100	-	-	-	-	-	-
Berkshire	-	-	-	-	-	-	-	-	100	-	-	-	-	-
Duroc	-	-	-	-	-	-	-	-	-	100	-	-	-	-
TLRI Black pig NO. 1	-	-	-	-	-	-	-	-	-	-	100	-	-	-
Taoyuan	-	-	-	-	-	-	-	-	-	-	-	100	-	-
Meishan	-	-	-	-	-	-	-	-	-	-	-	-	100	-
KHAPS Black pig	-	-	-	-	-	-	-	-	-	-	-	-	-	100

assignment testing among the 14 populations (Table 4, approach 2; Supplementary Table S1, approach 2: See e-version for supplement.). Figure 3 and Table 4 show that 99% accuracy in assignment was obtained using the following marker requirements: 6 markers with highest H_E, 7 markers with highest F_{ST}, 8 markers with highest K, and 10 markers with highest D-scores. If 9 markers with highest H_E or K were selected, we obtained 99.7% assignment-test accuracy. On the other hand, assignment accuracy for commercial breeds decreased when marker selection was from just 6 Lanyu related populations by ranking highest H_E, allele number, F_{ST} and D-scores (Supplementary Table S2 E, F, and G: See e-version for supplement.). One-hundred percent assignment accuracy among commercial breeds was achieved using: 9 markers with highest H_E, 7 markers with highest F_{ST}, and 9 markers with highest K (Supplementary Table S2 E, F and G: See e-version for supplement.). Overall assignment performance significantly improved after subset markers were selected based on highest ordering of H_E, allele numbers, F_{ST} and D-scores from close lineage Lanyu populations.

DISCUSSION

The Lanyu pig is a unique miniature pig breed. It is genetically separated from its European cousins and other Asian breeds by quite some distance (Wu et al., 2007; Jiang et al., 2008; Chang et al., 2009). Based on coat color and its original habitat (Lanyu [Orchid] Islet), the Lanyu pig is generally assigned as a domesticated Asian pig breed. In addition, its native habitat is geographically isolated, suggesting the possibility of a unique island domestication event (Larson et al., 2010). According to linguistic studies, Taiwan is thought to be the launching point of Austronesian peoples throughout the Asia Pacific region (Gray et al., 2009). The spread of domestic pigs has been used as a tracer to reconstruct the trajectory of Austronesian migration in prehistoric times (Larson et al., 2007; Larson et al., 2010). Understanding the origin and dispersion of Lanyu pigs provides useful information for resolving trajectory issues relating to Austronesian migration.

Originally, Lanyu pigs were conserved for development as future laboratory animals. To this end, the parental Lanyu population was subjected to inbreeding in isolation to standardize particular phenotypes and performance. A crossbreeding program was also used to obtain complementary heterosis and new breeds. Over two decades, complete pedigree information and identical performance were established among Lanyu inbred breeds and out-bred synthetic breeds. The Lanyu is now not only a laboratory pig but also an ideal model population for genetic studies. Additionally, effective population analysis based on

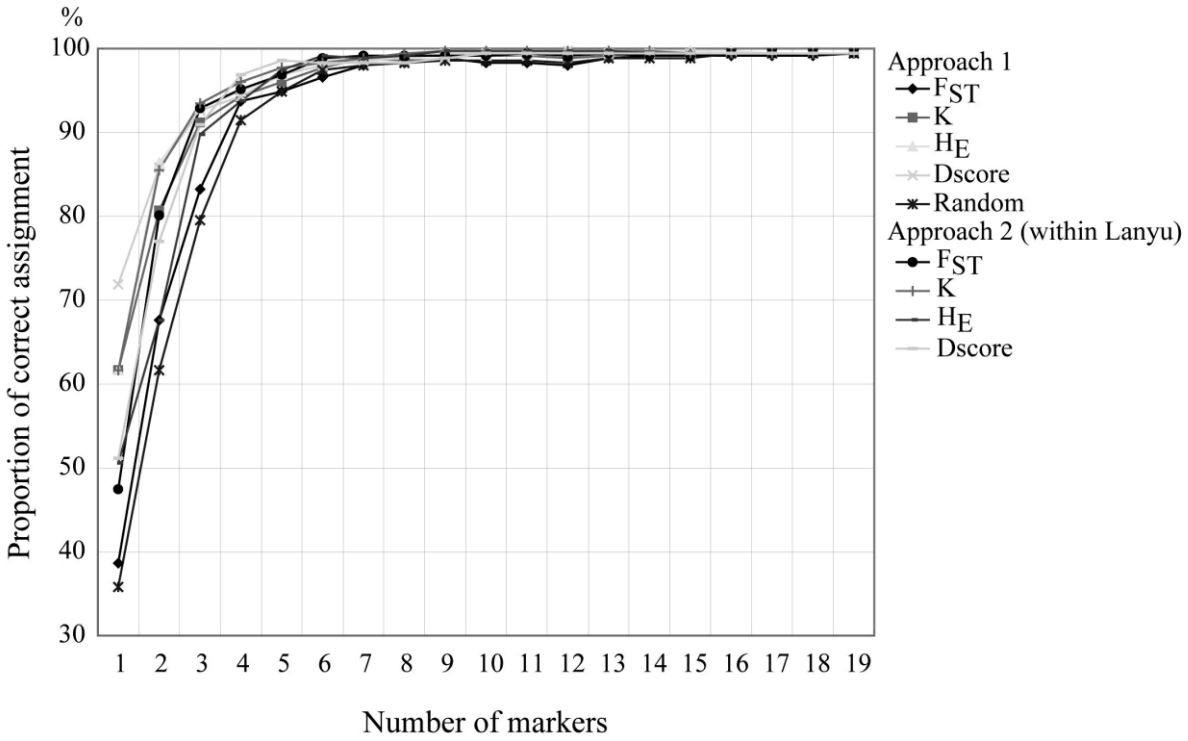

Figure 3. The results of assignment tests based on the Bayesian method among 14 pig populations. The figure shows the relation between the proportion of correctly assigned individuals and the number of markers used. In approach 1, markers were selected by ranking the highest orders of H_E, K (allele numbers), F_{ST}, D-scores and random selection. In approach 2 (within Lanyu), the ranking of assignment-test markers was selected based on the ranking's highest ordering of H_E, K, F_{ST}, and D-scores from 6 Lanyu related populations.

selected genetic markers and good genetic information is important for genetic management of the Lanyu population.

Significant loss of heterozygosity among microsatellite loci within Lanyu populations compared with Asian and European type commercial pig breeds

Inbreeding of the Lanyu and its derived breeds, Spotty Lanyu and Binlang Lanyu resulted in low allelic richness (2.097 and 1.980) and mean effective allele numbers (1.955 and 1.738) (Table 2). Lanyu (NTU) possesses low numbers for polymorphic loci (13), allelic richness (1.82), and mean effective allele numbers (1.574). This might be due to the initial size of the breeding population being small and founder effect (Wu et al., 2007). Fixation indexes (F_{IS}) were significantly derived from 0 in Lanyu (NTU) and Lanyu-derived populations, suggesting a loss of heterozygosity in these populations. Berkshire breeds (not a popular commercial breed in Taiwan) also possess low allelic richness and mean effective allele numbers. This may also be due to high intensive inbreeding and founder effect (Wu et al., 2007). Crossbred synthetic breeds, including Lee-Sung, Mitsai Lanyu, TLRI Black Pig No.1 and KHAPS Black pig, show reasonably high numbers for polymorphic loci, allelic richness and effective allele numbers even though all of them are established from small populations experiencing founder effect. This suggests that heterosis within these two very small unrelated populations have overcome founder effect in the inbreeding index. A similar result was obtained when the hermaphroditic freshwater snail lymnaea stagnalis was used as a model to test heterosis and inbreeding depression in bottlenecked populations (Coutellec and Caquet, 2011). Both the TLRI Black Pig No. 1 (75% Duroc and 25% Taoyuan) and KHAPS Black pig (50% Meishan and 50% Duroc) commercial proposed breeds possess higher figures for mean allele numbers (4.105 and 5.158, respectively), allelic richness (2.698 and 2.920, respectively) and effective allele numbers (2.547 and 2.970, respectively) than the Lee-Sung and Mitsai Lanyu populations. These differences may result from variations in population size and inbreeding practices after crossing of original parents. Information richness from the 19 microsatellite markers was enough to allow analysis of Taoyuan derived and Duroc derived synthetic breeds. Hale et al. (2012) randomly sub-sampling 5 to 100 individuals from 4 empirical microsatellite genotype datasets (*Formica lugubris*, *Sciurus vulgaris*, *Thalassarche melanophris*, and *Himantopus novaezelandia*) showed 25 to 35 individuals per population are enough to accurately estimate allele frequencies in genetic studies of populations. In this present study, the numbers used from each population varied; therefore, to compensate for this drawback, Nei (1978) and Kalinowski (2005) suggested using a large number of markers when estimating the average heterozygosity of a population.

Close lineage but high differentiation among Lanyu and Lanyu derived breeds

An index of genetic distance and differentiation among populations is given in Table 3. The genetic distances and differentiations range from 0.238 to 0.814 and 0.142 to 0.568, respectively. Lanyu and its derived populations showed little genetic distance but mid (0.142 between Binlang Lanyu and Spotty Lanyu) to high differentiation coefficients (0.475 between Lanyu [NTU] versus Mitsai Lanyu). In addition, there was lower observed heterozygosity (0.213 to 0.397) within Lanyu inbreeding populations (Table 2). Interesting, Lanyu (NTU) was co-ancestor to Lanyu (TAPS), Spotty Lanyu and Binling Lanyu but with high F_{ST} (from 0.283 to 0.416) values. These results indicate speciation among Lanyu and its derived breeds caused by intense inbreeding and founder effect. High genetic differentiation occurs when a highly intense inbreeding program is performed even among two pig populations with high levels of co-ancestry. High genetic distance and mid to high differentiation coefficients also exist between Duroc and Duroc synthetic breeds (Mitsai Lanyu and TLRI Black pig No. 1), ranging from 0.66 and 0.446 to 0.165 and 0.355, respectively (Table 3). In addition, Neighbor-Joining tree analysis formed 3 sub-clusters (Figure 1): the Lanyu, its derived breeds, and derived synthetic breeds. These sub-clusters formed a unique clade with a 100% bootstrap value. The European breeds and TLRI Black Pig No. 1, Asian pig breeds and KHAPS Black pig formed the other two clusters. The results from MDS plot analysis show similar results to Neighbor-Joining tree analysis except for the KHAPS Black pig breed. This might be due to close lineage and small genetic distances between KHAPS Black pig versus Duroc (0.512) and Meishan (0.544). The data indicate mid to high genetic divergence among these groups. Therefore, Lanyu populations provided both close genetic breeds and highly differentiated related breeds. This ideal animal model made it possible to examine and conduct different strategies for assignment testing. It allowed the development of an efficient method for identifying individuals based on our reference data.

The F_{ST}, allele number, and heterozygosity of microsatellite loci in close lineage reference populations significantly affect assignment accuracy

Some factors such as assignment principles, degree of genetic differentiation among populations, sample size per population, number of markers used, and level of variability in markers affect the accuracy and efficiency of assignment testing (Cornuet et al., 1999; Bjørnstad and Røed, 2002).

The rate at which accuracy of allele frequency-estimates increases should level out as sample size increases with the cost of genotyping more individuals increasing congruently (Hale et al., 2012). In addition, loci with high expected heterozygosity, K, and F_{ST} values across reference populations provides the best means of selecting loci to improve assignment testing efficiency (Rosenberg et al., 2001; Tadano et al., 2008). The Bayesian statistical approach proved more efficient than the genetic distance approach in calculating assignment of individuals to populations (Rannala and Mountain, 1997; Cornuet et al., 1999; Primmer et al., 2000). Even with as few as 10 microsatellite loci with heterozygosity ≈0.6, assignment accuracy performed by the Bayesian method can reach 100% based on 30 to 35 individuals from each of 10 populations when F_{ST} is near 0.1 (Cornuet et al., 1999). In this study, high heterozygosity (mean = 0.798) and a high level of genetic differentiation (mean F_{ST} = 0.403) among 14 pig populations increased the success rate of the Bayesian statistical approach in assigning individuals to populations (Supplementary Table S1, approach 1: See e-version for supplement.). Through applying the direct approach to the selection criteria needed for microsatellite markers, we could assign more than 99% of 352 animals to their correct reference populations. To make these assignments more efficient, we required 14 markers for ranking highest F_{ST}, 10 markers for highest H_E, 10 markers for highest K, and 9 markers for highest D-scores to obtain 99.1% correct assignment. This result is better than the assignment accuracy of previous studies (91.3% in Kim et al., 2005; 92.14% in Yang et al., 2003). Individual assignment based on markers of highest F_{ST} from 14 populations did not give ideal results.

Assignment error rates increase depending upon genetic similarities between populations and co-ancestry between individuals (Guinand et al., 2004). A perfect marker for understanding error-rates is mono-morphism within given breeds but polymorphism across breeds (Reed, 1973). Among the 352 individuals tested, one Spotty Lanyu was assigned to Lanyu in the TAPS population, and one Binlang Lanyu was assigned to the Spotty Lanyu population. The commercial breeds with high genetic differentiation obtained better assignment. These results reveal the closeness of these inbred groups and their preservation of original genetic characteristics (Table 4). This result led us to hypothesize that loci with high polymorphism, heterozygosity, F_{ST} and D-scores among Lanyu, its derived breeds, and synthetic breeds may provide a better ability to discriminate between individuals. High mean heterozygosity (0.631) and F_{ST} (0.305) were calculated for Lanyu and its derived populations, suggesting that genetic variation existed in related populations (Supplementary Table S1, approach 2: See e-version for supplement.). We, therefore, selected these microsatellite markers to correct individual assignments. By using markers within the Lanyu and Lanyu derived breeds and ranking these for highest H_E, highest K, high F_{ST} and high D-score. We found 6 markers with high H_E (S0218, SW951, S0355, SW122, SW857, S0266), 7 markers for high F_{ST} and 8 markers for high K (S0005, S0227, SW857, S0266, S0218, S0355, SW122, SW951) were required to obtain 99% assignment accuracy (Table 4 and Figure 3). High D-score microsatellite markers within Lanyu breeds did not increase individual-discrimination efficiency. Interestingly, microsatellite markers (SW857, S0355, SW951, and S0227) are used in the paternity control system for DNA tracking of pigs under selection (MAPS) by Spain and France (Boitard et al., 2010), indicating that these markers are informative for both Asian and European type pig breeds.

CONCLUSIONS

Effective assignment testing is defined by the selection of the minimum number of microsatellite markers that gives the maximum assignment success to a genetic population. But developed markers cannot be universally applied to all populations. The inbreeding of Lanyu derived breeds and out-bred synthetic populations provide an ideal model for examining strategies in selecting informative microsatellite markers for individual identification. Microsatellite markers are informative when variation in genetic characteristics among the reference populations can be used to make assignments. To avoid assignment errors, estimation of genetic differentiation among reference populations and identification of microsatellite markers of high H_E, F_{ST} or K among close lineage populations with high levels of co-ancestries are required to increase assignment efficiency and accuracy. These findings will be useful in paternity control, breed identification, conservation and forensic investigations.

CONFLICT OF INTEREST STATEMENT

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

ACKNOWLEDGEMENT

This work was supported by a grant from the National Science Council of Taiwan, Republic of China (NSC 94-2317-B-002-020 and NSC 96-2317-B-002-003) and Council of Agriculture of Taiwan, Republic of China (95AS-11.1.2-AD-U1, 96AS-11.1.4-AD-U1, and 97AS-

11.1.3-AD-U1). We thank Daniel Flynn for editing the English manuscript.

REFERENCES

Anderson, J. A., G. A. Churchill, J. E. Autrique, S. D. Tanksley, and M. E. Sorrells. 1993. Optimizing parental selection for genetic linkage maps. Genome 36:181-186.

Banks, M. A. and W. Eichert. 2000. WHICHRUN (version 3.2), a computer program for population assignment of individuals based on multilocus genotype data. J. Hered. 91:87-89.

Banks, M. A., W. Eichert, and J. B. Olsen. 2003. Which genetic loci have greater population assignment power? Bioinformatics 19:1436-1438.

Bjørnstad, G. and K. H. Røed. 2002. Evaluation of factors affecting individual assignment precision using microsatellite data from horse breeds and simulated breed crosses. Anim. Genet. 33:264-270.

Boitard, S., C. Chevalet, M. J. Mercat, J. C. Meriaux, A. Sanchez, J. Tibau, and M. Sancristobal. 2010. Genetic variability, structure and assignment of Spanish and French pig populations based on a large sampling. Anim. Genet. 41:608-618.

Bond, J. M., E. M. Veenendaal, D. D. Hornby, and A. J. Gray. 2002. Looking for progenitors, a molecular approach to finding the origins of an invasive weed. Biol. Invasions 4:349-357.

Botstein, D., R. L. White, H. Skolmick, and R. W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am. J. Hum. Genet. 32:314-331.

Caratti, S., L. Rossi, B. Sona, S. Origlia, S. Viara, G. Martano, C. Torre, and C. Robino. 2010. Analysis of 11 tetrameric STRs in wild boars for forensic purposes. Forensic Sci. Int. Genet. 4: 339-342.

Carr, M. R. 1996. PRIMER user manual. Ver 4.0. Plymouth routines in multivariate ecological research. Plymouth Marine Laboratory, Plymouth, UK.

Cavalli-Sforza, L. L. and A. W. F. Edwards. 1967. Phylogenetic analysis, models and estimation procedures. Am. J. Med. Genet. 19:233-257.

Chang, W. H., H. P. Chu, Y. N. Jiang, S. H. Li, Y. Wang, C. H. Chen, K. J. Chen, C. Y. Lin, and Y. T. Ju. 2009. Genetic variation and phylogenetics of Lanyu and exotic pig breeds in Taiwan analyzed by nineteen microsatellite markers. J. Anim. Sci. 87:1-8.

Cornuet, J. M., S. Piry, G. Luikart, A. Estoup, and M. Solignac. 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 153: 1989-2000.

Coutellec, M. A. and T. Caquet. 2011. Heterosis and inbreeding depression in bottlenecked populations, a test in the hermaphroditic freshwater snail *Lymnaea stagnalis*. J. Evol. Biol. 24:2248-2257.

Efron, B. 1983. Estimating the error rate of a prediction rule, Improvement on cross-validation. J. Am. Stat. Assoc. 78:316-331.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3) [Internet]. [cited 2014 Jan 22] Available from: http://www2.unil.ch/popgen/softwares/fstat.htm

Gray, R. D., A. J. Drummond, and S. J. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. Science 323:479-483.

Guinand, B., K. T. Scribner, K. S. Page, K. Filcek, and L. Main. 2006. Effects of coancestry on accuracy of individual assignments to population of origin: examples using great lakes lake trout (*Salvelinus namaycush*). Genetica 127:329-340.

Guinand, B., K. T. Scribner, A. Topchy, K. S. Page, W. Punch, and M. K. Burnham-Curtis. 2004. Sampling issues affecting accuracy of likelihood-based classification using genetical data. Environ. Biol. Fish. 69:245-259.

Hale, M. L., T. M. Burg, and T. E. Steeves. 2012. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PLoS ONE 7(9):e45170.

ISAG-FAO. 2004. Secondary Guidelines: Measurement of Domestic Animal Diversity (MoDAD): Recommended Microsatellite Markers.

Jiang, Y. N., C. Y. Wu, C. Y. Huang, H. P. Chu, M. W. Ke, M. S. Kung, K. Y. Li, S. H. Li, C. H. Wang, Y. Wang, and Y. T. Ju. 2008. Inter-population and intra-population maternal lineage genetics of Lanyu pig (*Sus scrofa*) by analysis of mitochondrial cytochrome *b* and control region sequences. J. Anim. Sci. 86:2461-2470.

Kalinowski, S. T. 2005. Do polymorphic loci require large sample sizes to estimate genetic distances? Heredity 94:33-36.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16:1099-1106.

Kim, T. H., K. S. Kim, B. H. Choi, D. H. Yoon, G. W. Jang, K. T. Lee, H. Y. Chung, H. Y. Lee, H. S. Park, and J. W. Lee. 2005. Genetic structure of pig breeds from Korea and China using microsatellite loci analysis. J. Anim. Sci. 83:2255-2263.

Kimura, M. and J. F. Crow. 1964. The number of alleles that can be maintained in a finite population. Genetics 49:725-738.

Koskinen, M. T. 2003. Individual assignment using microsatellite DNA reveals unambiguous breed identification in the domestic dog. Anim. Genet. 34:297-301.

Krzanowski, W. 1987. Principles of multivariate analysis, a user's perspective. Clarendon Press, Oxford.

Langella, O. 2002. POPULATIONS 1.2.28, Population genetic software, individuals or populations distances based on allelic frequencies, phylogenetic trees, file conversions [Internet]. [cited 2011 Feb 15] Available from: http://bioinformatics.org/~tryphon/populations/

Larson, G., T. Cucchi, M. Fujita, E. Matisoo-Smith, J. Robins, A. Anderson, B. Rolett, M. Spriggs, G. Dolman, T. H. Kim, N. T. D. Thuy, E. Randi, M. Doherty, R. A. Due, R. Bollt, T. Djubiantono, B. Griffin, M. Intoh, E. Keane, P. Kirch, K. T. Li, M. Morwood, L. M. Pedriña, P. J. Piper, R. J. Rabett, P. Shooter, G. V. den Bergh, E. West, S. Wickler, J. Yuan, A. Cooper, and K. Dobneybc. 2007. Phylogeny and ancient DNA of *Sus* provides insights into neolithic expansion in Island Southeast Asia and Oceania. Proc. Natl. Acad. Sci. USA 104:

4834-4839.

Larson, G., R. Liu, X. Zhao, J. Yuan, D. Fuller, L. Barton, K. Dobney, Q. Fan, Z. Gu, X. H. Liu, Y. Luo, P. Lv, L. Andersson, and N. Li. 2010. Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. Proc. Natl. Acad. Sci. USA 107:7686-7691.

Lowenstein, J. H., J. Burger, C. W. Jeitner, G. Amato, S. O. Kolokotronis, and M. Gochfeld. 2010. DNA barcodes reveal species-specific mercury levels in tuna sushi that pose a health risk to consumers. Biol. Lett. 6:692-695.

Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7:639-655.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA 70:3321-3323.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590.

Piry, S., A. Alapetite, J. M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GeneClass2, A Software for genetic assignment and first-generation migrant detection. J. Hered. 95: 536-539.

Primmer, C. R., T. T. K. Mikko, and J. Piironen. 2000. The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. Proc. R. Soc. Lond. 267:1699-1704.

Rannala, B. and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proc. Natl. Acad. Sci. USA. 94: 9197-9201.

Reed, T. E. 1973. Number of gene loci required for accurate estima tion of ancestral population proportions in individual human hybrids. Nature 244:575-576.

Rosenberg, N. A., E. Woolf, J. K. Pritchard, T. Schaap, D. Gefel, I. Shpirer, U. Lavi, B. Bonné-Tamir, J. Hillel, and M. W. Feldman. 2001. Distinctive genetic signatures in the Libyan Jews. Proc. Natl. Acad. Sci. USA 98:858-863.

Rousset, F. 2008. Genepop, a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resour. 8:103-106.

Tadano, R., M. Nishibori, and M. Tsudzuki. 2008. High accuracy of genetic discrimination among chicken lines obtained through an individual assignment test. Anim. Genet. 39:567-571.

Wu, C. Y., Y. N. Jiang, H. P. Chu, S. H. Li, Y. Wang, Y. H. Li, Y. Chang, and Y. T. Ju. 2007. The type I Lanyu pig has a maternal genetic lineage distinct from Asian and European pigs. Anim. Genet. 38:499-505.

Wu, M. C., H. P. Chu, and K. J. Chen. 2009. Lanyu pig transforming from the conserved animal to new breeds for medical biology industry. Sci-Tech Policy Rev. 5:88-93.

Yang, S. L., Z. G. Wang, B. Liu, G. X. Zhang, S. H. Zhao, M. Yu, B. Fan, M. H. Li, T. A. Xiong, and K. Li. 2003. Genetic variation and relationships of eighteen Chinese indigenous pig breeds. Genet. Sel. Evol. 35:657-671.

Yeh, F. C., R. C. Yang, and T. Boyle. 1999. POPGENE (Version 1.31), Microsoft Window-bases freeware for population genetic analysis, University of Alberta and the Centre for International Forestry Research.