Full Paper

Construction of an individual identification panel for horses using insertion and deletion markers

Teruaki TOZAKI[1#*], Aoi OHNUMA[1#], Mio KIKUCHI[1], Taichiro ISHIGE[1], Hironaga KAKOI[1], Kei-ichi HIROTA[1] and Shun-ichi NAGATA[1]

[1] Genetic Analysis Department, Laboratory of Racing Chemistry, Tochigi 320-0851, Japan

Individual identification and paternity testing are important for avoiding inbreeding in the management of small populations of wild and domestic animals. In horse racing industries, they are extremely important for identifying and registering individuals and doping control to ensure fair competition. In this study, we constructed an individual identification panel for horses by using insertion and deletion (INDEL) markers. The panel included 39 INDEL markers selected from a whole-genome INDEL database. Genotyping of 89 Thoroughbreds showed polymorphisms with minor allele frequencies (MAFs) of 0.180–0.489 in all markers. The total probability of exclusion for paternity testing, power of discrimination, and probability of identity were 0.9994271269, >0.9999999999, and 0.9999999987, respectively. The panel was applied to 13 trios (sires, dams, and foals), and no contradictions were observed in genetic inheritance among the trios. When this panel was applied to the trios (52 trios) containing false fathers, an average of 7.3 markers excluded parentage relationships. In addition, genomic DNA extracted from the urine of six horses was partially genotyped for 39 markers, and 6–28 markers were successfully genotyped. The newly constructed panel has two advantages: a low marker mutation rate compared with short tandem repeats and a genotyping procedure that is as simple as short tandem repeat typing compared with single nucleotide variant typing. This panel can be applied for individual identification, paternity determination, and urine-sample identification in Thoroughbred horses.

Key words: *insertion/deletion, parentage testing, single nucleotide variant, Thoroughbred*

J. Equine Sci.
Vol. 34, No. 3
pp. 83–92, 2023

Individual identification and paternity testing are important for avoiding inbreeding in the management of small populations of wild and domestic animals [11, 16, 19, 24]. In the horse racing industry, they are extremely important for identifying and registering individuals and doping control to ensure fair competition [6]. Short tandem repeats (STRs), also known as microsatellite DNA, contain multiple alleles, and an STR panel can be constructed using a small number of markers [20]. Therefore, STRs are used as markers to construct individual identification panels for horses, cattle, dogs, cats, and other animals [4, 12, 15, 18]. Although multiple alleles in STRs are advantageous as markers, high mutation rates are a disadvantage. Exclusion of ≥2 markers is recommended by the International Society for Animal Genetics for paternity testing in Thoroughbreds (https://www.isag.us/Docs/EquineGeneticsParentage_CT.pdf, accessed on 14 July 2023).

With improvements in analytical technologies, individual identification panels based on single nucleotide variants (SNVs) have recently been developed [3]. Although SNVs only have two alleles, they can achieve a parentage discrimination ability equivalent to that of STRs using a large number of markers. Exhaustive analysis methods that use hundreds of SNVs have also been employed.

A horse reference genome sequence has been generated and published, as EquCab2 and later as EquCab3 [13, 25]; thus, genome-wide variants are easily obtained by aligning massive numbers of overlapping sequencing reads to the horse reference genome. Through this procedure, whole-genome sequencing (WGS) of 88 horses (25 breeds) identified 23,559,582 SNVs and 2,396,022 insertions and deletions (INDELs) [8], and that of 534 horses (46 breeds)

Received: May 29, 2023
Accepted: Jul. 21, 2023
*Corresponding author. e-mail: ttozaki@lrc.or.jp
#These authors contributed equally to this work.

identified 29,038,030 SNVs and 2,102,379 INDELs [5]. In our previous studies, WGS of 101 Thoroughbred horses in Japan was analysed, and a whole-genome variant database, including loci and frequencies of approximately 12 million SNVs and 1.56 million short INDELs, was constructed [22, 23]. This made it possible to search for diverse SNVs and INDELs.

Because INDELs have potential advantages, including low mutation rates and allowing for the analyses of degraded samples, they have also been recently used as markers for individual identification in humans [2]. In this study, we used INDEL markers to construct an individual identification panel for horses. The panel we constructed here has two advantages: a low marker mutation rate and a genotyping procedure that is as simple as STR typing. We also examined the possibility of applying the constructed INDEL panel to parent–child determination tests and doping control.

Materials and Methods

Blood and urine sample collection

Blood and urine sample collection was performed in accordance with the Animal Care Committee of the Laboratory of Racing Chemistry (approval number: 20-4), with consent for sample collection and research use obtained from all owners. Blood samples from 89 Thoroughbred horses were collected from the Hidaka Training and Research Center (HTRC, Uraga, Japan) of the Japan Racing Association (JRA). Blood samples were collected from 43 Hokkaido horses of individual owners, and those of 13 trios (5 sires, 13 dams, and 13 foals) were collected from the HTRC, JRA, and Japan Bloodhorse Breeders' Association (Hidaka, Japan). Blood and urine samples from six Thoroughbred horses (BH, BP, BR, FP, LL, and PT) were collected from the JRA Horseracing School (Shiroi, Japan).

Genomic DNA extraction from urine and blood samples

Genomic DNA was extracted from whole blood (200 μl) using a DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) and diluted to 5 ng/μl using Milli-Q water (Merck, Kenilworth, NJ, USA).

Genomic DNA was extracted from urine (5 ml) using a Quick-DNA Urine Kit (Zymo Research, Irvine, CA, USA) based on the total (cellular and cell-free) DNA extraction procedure. Briefly, 350 μl of urine conditional buffer and 10 μl of cleaning beads were added to 5 m*l* of urine, and the mixture was centrifuged at 3,000 × *g* for 15 min. After discarding the supernatant and resuspending the pellet in urine pellet digestion buffer, the pellet was digested with proteinase K at 55°C for 30 min. Finally, DNA was purified from the digested product using a Zymo-Spin IC-S Column (Zymo Research, Irvine, CA, USA). The extract was eluted using 50 μl of elution buffer.

Marker selection and primer design

INDEL information was downloaded from DB101 (https://OSF.IO/QN45H/). DB101 is a database for INDELs that was constructed using the WGSs of 101 Thoroughbred horses in our previous study [23]. INDELs with reference allele frequencies in the range of 0.25–0.75 were extracted, and 39 INDELs were then selected based on chromosome and position information (non-neighbouring). Primers for PCR amplification were designed using Primer 3 (version 0.4.0; https://bioinfo.ut.ee/primer3-0.4.0/). The adaptor sequence, 5′-TGACCGGCAGCAAAATTG-3′, was conjugated to the 5′ end of forward primers.

FAM-, HEX-, or NED-labelled primers (TGACCGGCAGCAAAATTG) were prepared for adapter fluorescent-labelling PCR amplification.

PCR amplification, electrophoresis, and genotyping

PCR was performed in a total mixture volume of 20 μl containing 5 μl of DNA extract, 0.25–2 pmol of forward primer, 0.625–5 pmol of reverse primer (Table 1), 20 pmol of FAM-labelled primer or 10 pmol of HEX- or NED-labelled primer, 2.5 mM of $MgCl_2$, 330 μM of dNTPs, 2 μl of 10× reaction buffer, and 0.1 U of AmpliTaq Gold polymerase (Thermo Fisher Scientific, Waltham, MA, USA). Thermal cycling was performed on a GeneAmp PCR System 9700 (Thermo Fisher Scientific) with the following conditions: initial denaturation at 95°C for 10 min; 15 cycles each of 95°C and 60°C at 30 sec, followed by 72°C for 1 min; 35 cycles each of 95°C and 50°C at 30 sec, followed by 72°C for 1 min, and a final extension step at 72°C for 10 min.

The PCR products (2 μl each) from each amplification (FAM, HEX, and NED) were mixed with 20 μl of Hi-Di Formamide (Thermo Fisher Scientific) and 0.25 μl of GeneScan 500 ROX Size Standard (Thermo Fisher Scientific). After denaturing at 95°C for 2.5 min, the products were subjected to electrophoresis on a 3500 Genetic Analyser (Thermo Fisher Scientific) with a 36-cm Capillary Array (Thermo Fisher Scientific). Genotyping was performed using GeneMapper (version 6; Thermo Fisher Scientific). Peaks with more than 100 relative fluorescence units (RFUs) were identified as alleles.

Statistical and parentage analyses

Minor allele frequency (MAF), expected heterozygosity (H_e), and observed heterozygosity (H_o) were calculated using SNPAlyze (ver. 9; Dynacom Co., Ltd., Chiba, Japan). Probabilities of exclusion (PEs) were calculated from the allele frequencies for the following: paternity testing (PE1), one unavailable parental genotype (PE2), and exclusion of both parents (PE3) [9]. Power of discrimination (PD) was

Table 1. Primers and product sizes of 39 insertion/deletion markers

Marker name/ PCR group	Chromosome/ position	SNP number/ allele size (b)	Forward and reverse primers	Primer concentration
ID_01	9	rs3108262338	ATGTGGCTCTTTATCTGGCTTT	2 pmol
group1	29379577	119/121	GCGAGATTTATTGTGAAGTCAAGGT	5 pmol
ID_02	12	rs3434073388	CTCTCTTAGTCCCCAGCGTG	2 pmol
group1	29009990	125/127	AAGACCCCTCAGCAAACTCG	5 pmol
ID_03	8	rs3101668550	CGGTCAGTATTCAAATTTCCCAGT	2 pmol
group1	18680694	131/135	CACAAGAACCAAATGCAAGGC	5 pmol
ID_04	28	rs3102755950	TGGCAACCTCTGGACATTTGA	2 pmol
group1	5920447	141/143	TCCAAGCCACTTACACCTATTCC	5 pmol
ID_05	13	rs3105001983	AAGCTTCTCAGTCTTGGCCC	2 pmol
group1	1778429	154/156	TTAGTCCTTGGAGCCTGGGG	5 pmol
ID_06	9	rs3436664934	CATGTCACCCAGGTATCTAAACTT	2 pmol
group1	40972218	159/162	GAGTGGAGCAGGCAGGAAAT	5 pmol
ID_07	29	rs3109233637	TTCTCCAGGATCAGGCAAGT	2 pmol
group1	11672358	168/170	CAAGTCAGATGCAAAACCTGGG	5 pmol
ID_08	7	rs3107232385	GTTTGTCTTGCCACCAGTGT	2 pmol
group1	7125111	176/180	GCTCAACACTGCTAGAGAGTCA	5 pmol
ID_09	30	rs3101636233	GCTTCGCTGCACTTTACAAA	2 pmol
group1	5148040	184/187	ACTCAGACACAGACACATAACTACT	5 pmol
ID_10	14	rs3091792469	GGCTCATATTGCTGAAAGGTGC	2 pmol
group1	8260720	189/191	TCCCTGAAGAGTGCCTGGAA	5 pmol
ID_11	10	rs3109270259	CCCCACACATGCAACTAACAC	2 pmol
group1	652088	199/201	GTGCTTCATGTTATACACCCTGT	5 pmol
ID_12	7	rs3107233025	AGTCTAGATCCTCATACCTCTCCT	2 pmol
group1	19839097	209/211	TTGAATGCTAGTTCTGTTAGTGAGT	5 pmol
ID_13	27	rs3106181037	ACGTTAATGGGAGTGCTTGTT	2 pmol
group1	4365607	216/218	GCCGTAGATTTCAGAACACTGC	5 pmol
ID_14	28	rs3103966164	TGAACTGTCAAAAGCATGGAGA	2 pmol
group2	7906782	117/119	CATCCTACCCACGGCCAAAT	5 pmol
ID_15	25	rs3429697783	GCTGCAAACTACCACCAAAGT	2 pmol
group2	20509376	125/128	CGAGAAGCAACAATGAAGGGG	5 pmol
ID_16	16	rs3107277314	TGGTGAGGAGCCAGATGAAG	2 pmol
group2	20213504	133/136	GGAAGTGCCTGTGACGATGA	5 pmol
ID_17	29	rs3101634599	ACAGAAAACACTTAACTCCTGGGT	2 pmol
group2	7731419	141/143	GCATCTACATGGACCCCTTCA	5 pmol
ID_18	13	rs3103975983	GAGTATGTATTCTAGCCAAGCACC	1 pmol
group2	3351442	146/150	CACTCTGGATGGAAGGCACA	2.5 pmol
ID_19	16	rs3107226834	TGCATCATACTTTCACCAACTTCC	2 pmol
group2	25970117	154/157	TGACGAACCAGAACCAGAATTG	5 pmol
ID_20	28	rs3434793343	GGCCCAAAATTCATGCTCCA	2 pmol
group2	9227810	162/164	GAGTTACCAACCAAAGCAAGCT	5 pmol
ID_21	30	rs3097022392	TGCTAGAAGACAGACCCAGA	2 pmol
group2	12690178	177/180	CAGGGCAGGACTACATTTTCTG	5 pmol
ID_22	27	rs3429816921	GTGCTTCCTCCATGATCACCT	2 pmol
group2	1539503	187/189	TTATGTGTGGCAGGGTGGTG	5 pmol
ID_23	11	rs3099775755	CACTGGCTATGTAACGAGGGAA	2 pmol
group2	7014702	194/196	TCCAAAACAAACGTAGCACCTC	5 pmol
ID_24	6	rs3109201461	ACCAGGCTCTCACATTGACAT	2 pmol
group2	4026094	204/206	CATCTTTGGCATCAGTTAAGCAGA	5 pmol
ID_25	23	rs3108209026	TGAGCGTCACTGATGTTTCCA	2 pmol
group2	11943998	210/213	TTTCTTGCATACCTACTACACTCC	5 pmol
ID_26	20	rs3102724697	GCCATAGGACTTAAACCCCTCC	2 pmol
group2	8942210	222/225	GAGAAGCCTTGTGATTTCGGTG	5 pmol
ID_27	26	rs3437574794	TTGGGCTGGCGATGTCTAAG	1 pmol
group3	26930329	117/119	GGCAAACAAAAGAAGTGGGGA	2.5 pmol
ID_28	27	rs3102714308	GGACAGACGTTGGGAGGAA	0.5 pmol
group3	3906270	126/129	AGATGCACCCAGACAAGGTC	1.25 pmol
ID_29	25	rs3106173216	GGCCTATGTGTTCACTCGCT	0.25 pmol
group3	32819125	135/137	AGCTCTGACCTCTCTGACCC	0.625 pmol
ID_30	28	rs3097056142	AGGTTCAAGTGCAGATGGGT	0.5 pmol
group3	5206051	145/148	TGGAATCCTATCTTTGGCTTACTG	1.25 pmol
ID_31	20	rs3433262390	CAGTGGGAGGGTGACTTTGA	2 pmol
group3	13690891	152/155	CAGGGTGCGAACTTCCTCAA	5 pmol

Table 1. Primers and product sizes of 39 insertion/deletion markers (continued)

Marker name/ PCR group	Chromosome/ position	SNP number/ allele size (b)	Forward and reverse primers	Primer concentration
ID_32	25	rs3109279099	GCCAGATTTTCAGCCTCAGG	2 pmol
group3	16357709	159/162	TCATCTCGGAAATAGGACAACTGA	5 pmol
ID_33	26	rs3101664183	TCTGATACAAAGGAGATGCTTGGT	2 pmol
group3	5126429	166/169	ACTAGGGGTAGCTAAGTCAGTAGG	5 pmol
ID_34	26	rs3097091955	ATGATTAAACCATTGCTCTGCTTTT	2 pmol
group3	5647032	175/178	GCCAGTCCCAGGAGAAGC	5 pmol
ID_35	21	rs3102768946	CCCCAACCCCAAAAGCCAAT	2 pmol
group3	33970288	186/190	AAGTCACTCTGTTCATTTAGCCTGT	5 pmol
ID_36	22	rs3109253811	CCCTTTGTATTCCTGGTCTGCT	2 pmol
group3	22558757	190/192	ACTAGCAAGCCCGATGAAGTC	5 pmol
ID_37	15	rs3107270471	GGGATCGAAGGCTACAACCC	2 pmol
group3	4421217	201/203	CTGCTATGAGTCACGAGTGGA	5 pmol
ID_38	24	rs3429680474	TCACCCAGAGCTTATGAGAGAT	2 pmol
group3	17054211	207/210	TCGGGTTGAATTAGCAGGCT	5 pmol
ID_39	24	rs3437727710	TGCACTGTCCACATCGTAAAAC	2 pmol
group3	8130109	216/218	TTCCCCTGACACTGGCTCTA	5 pmol

calculated using allele frequencies [14], while probability of identity (PID) was calculated from the allele frequency [17]. The average and cumulative PE, PD, and PID were calculated as the mean and total PE, PD, and PID, respectively. Based on Mendel's law of inheritance, parentage (trio analysis) was determined based on allele sharing among the sires, dams, and foals.

Results

Construction of an INDEL marker panel

First, 1,274,708 diallelic INDELs of the autosome (chromosomes 1–31) were extracted from DB101. Subsequently, INDELs with a size difference of 1 bp between insertion and deletion were excluded from the marker candidates, as experimentally discriminating each allele was difficult; this left 366,752 INDELs with a size difference of 2–4 bp, 97,443 of which were in the range of 0.25–0.75 in terms of MAF, excluding all-heterozygous INDELs. Finally, 39 INDELs were identified as marker candidates by selection from different chromosome positions, and primers for their amplification were designed (Table 1).

A multiplex PCR of 39 INDELs was performed for each set (FAM, group 1; HEX, group 2; NED, group 3; Table 1). After amplification, all PCR products were mixed and subjected to capillary gel electrophoresis. Although the amount of amplified product detected (peak height) differed between the markers, 39 INDEL markers were successfully detected (Fig. 1).

Statistical analyses of the INDEL panel

Genotyping of 89 Thoroughbreds showed polymorphisms in all markers (MAF: 0.180 at ID_39 to 0.489 at ID_12 and ID_15; Table 2). The MAFs in DB101 (MAF: 0.253 at ID_39 to 0.495 at ID_11) were similar, although differences between some markers were observed. Three markers (ID_01, ID_09, and ID_31) showed no polymorphisms in the 43 Hokkaido horses, whereas the remaining 36 markers showed polymorphisms (MAF:0.012 at ID_34 and ID_39 to 0.477 at ID_16 and ID_36; Table 2). In all INDEL markers, except ID_05, ID_06, ID_07, ID_13, ID_18, ID_27, ID_30, and ID_37, deletion (del) was identified as a minor allele in Thoroughbred horses. In all INDEL markers, except ID_04, ID_05, ID_07, ID_10, ID_11, ID_13, ID_16, ID_18, ID_21, ID_25, ID_30, ID_32, ID_35, ID_36, and ID_38, deletion (del) was identified as a minor allele in Hokkaido horses.

The cumulative values for PE, PD, and PID are listed in Table 3. The total probabilities of exclusion for paternity testing (PE1s) in DB101 and the 89 Thoroughbreds were 0.9994630250 and 0.9994271269, respectively. The total PD was >0.9999999999 in DB101 and the 89 Thoroughbreds. The total PIDs in DB101 and the 89 Thoroughbreds were 0.9999999982 and 0.9999999987, respectively.

Parentage analyses using the INDEL panel

Genetic inheritance was examined using 13 trios (sire, dam, and foal), and no contradictions were observed in genetic inheritance among the trios. Subsequently, we investigated the parent–child relationships (52 pairs) by assigning false fathers to trios with true mothers and children. Of the 39 markers, an average of 7.3 markers (minimum 1 marker and maximum 13 markers) excluded parentage relationships.

DNA extraction from horse urine samples and genotyping

An INDEL panel was applied to six horses with both urine and blood samples. Genomic DNA extracted from the blood was successfully and completely genotyped for 39

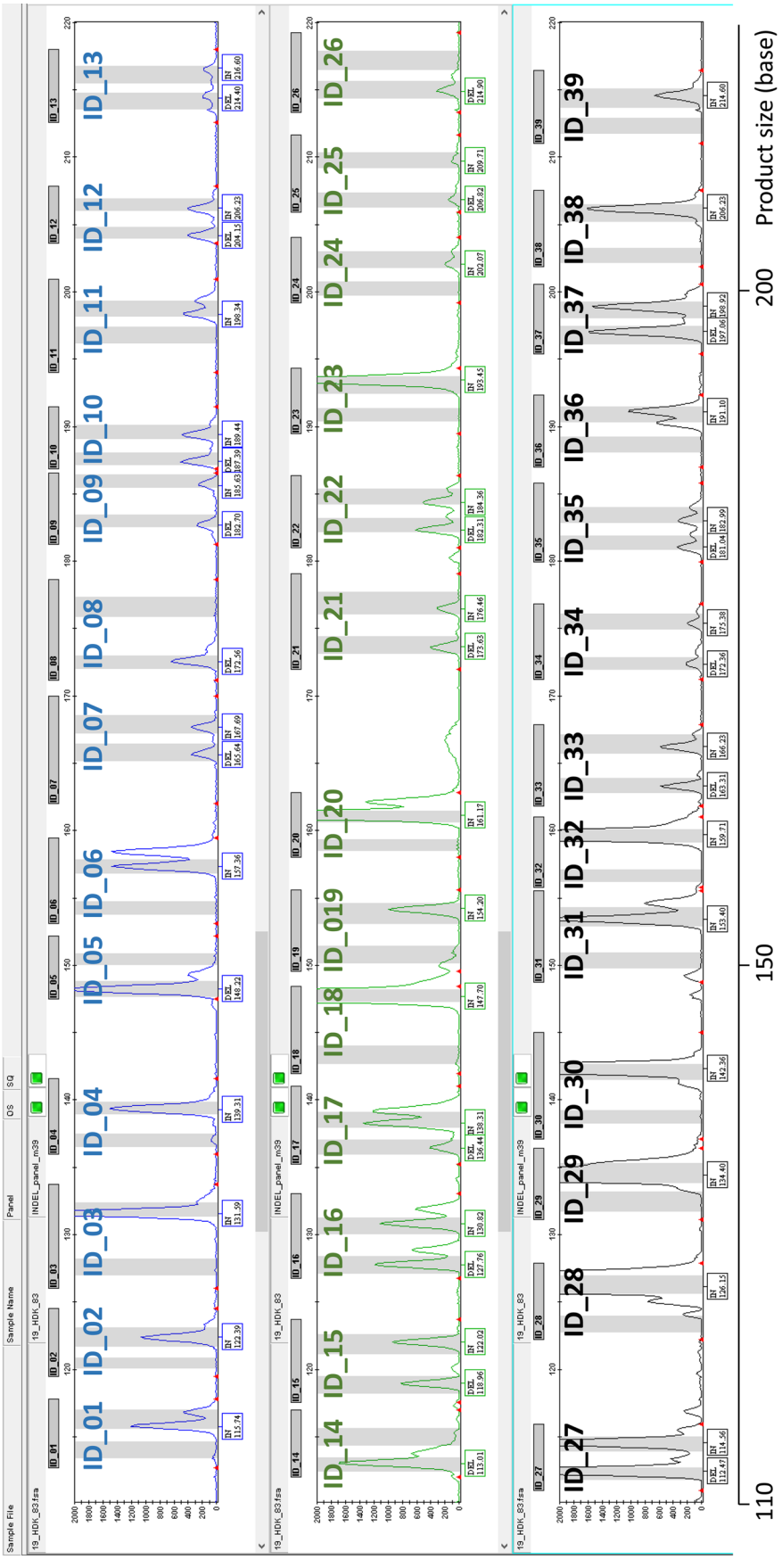

Fig. 1. Electropherogram of 39 insertion/deletion markers. ID_01 to ID_13 were detected using FAM (blue), ID_04 to ID_26 were detected using HEX (green), and ID27 to 39 were detected using NED (black).

Table 2. Statistical information of the 39 insertion/deletion markers

Marker	Breed	MAF	He	Ho	PE1	PE2	PE3	PD	PID
ID_01	DB101	0.460	0.497	0.505	0.187	0.123	0.280	0.623	0.377
	Thoroughbred	0.472	0.498	0.449	0.187	0.124	0.281	0.624	0.376
	Hokkaido	0	0	0	0	0	0	0	1
ID_02	DB101	0.257	0.382	0.356	0.155	0.073	0.242	0.545	0.455
	Thoroughbred	0.287	0.409	0.303	0.163	0.084	0.251	0.567	0.433
	Hokkaido	0.012	0.023	0.023	0.011	0.000	0.022	0.045	0.955
ID_03	DB101	0.386	0.474	0.436	0.181	0.112	0.273	0.611	0.389
	Thoroughbred	0.298	0.418	0.416	0.165	0.087	0.254	0.574	0.426
	Hokkaido	0.116	0.206	0.140	0.092	0.021	0.159	0.348	0.652
ID_04	DB101	0.302	0.422	0.426	0.166	0.089	0.256	0.577	0.423
	Thoroughbred	0.382	0.472	0.517	0.180	0.111	0.272	0.610	0.390
	Hokkaido	0.279	0.402	0.326	0.161	0.081	0.249	0.562	0.438
ID_05	DB101	0.297	0.418	0.396	0.165	0.087	0.254	0.574	0.426
	Thoroughbred	0.287	0.409	0.416	0.163	0.084	0.251	0.567	0.433
	Hokkaido	0.023	0.045	0.047	0.022	0.001	0.043	0.088	0.912
ID_06	DB101	0.376	0.469	0.396	0.180	0.110	0.272	0.608	0.392
	Thoroughbred	0.421	0.488	0.528	0.184	0.119	0.277	0.619	0.381
	Hokkaido	0.279	0.402	0.372	0.161	0.081	0.249	0.562	0.438
ID_07	DB101	0.381	0.472	0.446	0.180	0.111	0.272	0.610	0.390
	Thoroughbred	0.298	0.418	0.438	0.165	0.087	0.254	0.574	0.426
	Hokkaido	0.465	0.498	0.372	0.187	0.124	0.280	0.624	0.376
ID_08	DB101	0.262	0.387	0.327	0.156	0.075	0.243	0.549	0.451
	Thoroughbred	0.281	0.404	0.382	0.161	0.082	0.249	0.563	0.437
	Hokkaido	0.105	0.187	0.209	0.085	0.018	0.148	0.322	0.678
ID_09	DB101	0.401	0.480	0.485	0.183	0.115	0.275	0.615	0.385
	Thoroughbred	0.388	0.475	0.573	0.181	0.113	0.273	0.611	0.389
	Hokkaido	0	0	0	0	0	0	0	1
ID_10	DB101	0.465	0.498	0.535	0.187	0.124	0.280	0.624	0.376
	Thoroughbred	0.382	0.472	0.404	0.180	0.111	0.272	0.610	0.390
	Hokkaido	0.430	0.490	0.581	0.185	0.120	0.278	0.620	0.380
ID_11	DB101	0.495	0.500	0.515	0.187	0.125	0.281	0.625	0.375
	Thoroughbred	0.449	0.495	0.494	0.186	0.122	0.280	0.622	0.378
	Hokkaido	0.186	0.303	0.326	0.129	0.046	0.209	0.468	0.532
ID_12	DB101	0.485	0.500	0.535	0.187	0.125	0.281	0.625	0.375
	Thoroughbred	0.489	0.500	0.551	0.187	0.125	0.281	0.625	0.375
	Hokkaido	0.058	0.110	0.116	0.052	0.006	0.096	0.201	0.799
ID_13	DB101	0.421	0.487	0.465	0.184	0.119	0.277	0.618	0.382
	Thoroughbred	0.483	0.499	0.472	0.187	0.125	0.281	0.625	0.375
	Hokkaido	0.419	0.487	0.605	0.184	0.118	0.277	0.618	0.382
ID_14	DB101	0.416	0.486	0.495	0.184	0.118	0.277	0.618	0.382
	Thoroughbred	0.360	0.461	0.517	0.177	0.106	0.269	0.603	0.397
	Hokkaido	0.384	0.473	0.302	0.181	0.112	0.273	0.610	0.390
ID_15	DB101	0.421	0.487	0.525	0.184	0.119	0.277	0.618	0.382
	Thoroughbred	0.489	0.500	0.483	0.187	0.125	0.281	0.625	0.375
	Hokkaido	0.360	0.461	0.488	0.177	0.106	0.269	0.603	0.397
ID_16	DB101	0.426	0.489	0.455	0.185	0.120	0.278	0.619	0.381
	Thoroughbred	0.388	0.475	0.483	0.181	0.113	0.273	0.611	0.389
	Hokkaido	0.477	0.499	0.442	0.187	0.124	0.281	0.624	0.376
ID_17	DB101	0.272	0.396	0.366	0.159	0.079	0.247	0.557	0.443
	Thoroughbred	0.382	0.472	0.517	0.180	0.111	0.272	0.610	0.390
	Hokkaido	0.186	0.303	0.372	0.129	0.046	0.209	0.468	0.532
ID_18	DB101	0.436	0.492	0.475	0.185	0.121	0.279	0.621	0.379
	Thoroughbred	0.483	0.499	0.449	0.187	0.125	0.281	0.625	0.375
	Hokkaido	0.349	0.454	0.279	0.176	0.103	0.267	0.599	0.401
ID_19	DB101	0.322	0.436	0.426	0.171	0.095	0.261	0.587	0.413
	Thoroughbred	0.348	0.454	0.472	0.175	0.103	0.267	0.599	0.401
	Hokkaido	0.140	0.240	0.233	0.106	0.029	0.178	0.394	0.606
ID_20	DB101	0.460	0.497	0.564	0.187	0.123	0.280	0.623	0.377
	Thoroughbred	0.404	0.482	0.404	0.183	0.116	0.276	0.615	0.385
	Hokkaido	0.047	0.089	0.093	0.042	0.004	0.079	0.166	0.834

Table 2. Statistical information of the 39 insertion/deletion markers (continued)

Marker	Breed	MAF	He	Ho	PE1	PE2	PE3	PD	PID
ID_21	DB101	0.436	0.492	0.455	0.185	0.121	0.279	0.621	0.379
	Thoroughbred	0.483	0.499	0.472	0.187	0.125	0.281	0.625	0.375
	Hokkaido	0.302	0.422	0.512	0.166	0.089	0.256	0.577	0.423
ID_22	DB101	0.307	0.425	0.396	0.167	0.091	0.257	0.579	0.421
	Thoroughbred	0.315	0.431	0.472	0.169	0.093	0.259	0.584	0.416
	Hokkaido	0.128	0.223	0.209	0.099	0.025	0.169	0.372	0.628
ID_23	DB101	0.441	0.493	0.485	0.186	0.121	0.279	0.621	0.379
	Thoroughbred	0.455	0.496	0.461	0.186	0.123	0.280	0.623	0.377
	Hokkaido	0.430	0.490	0.581	0.185	0.120	0.278	0.620	0.380
ID_24	DB101	0.267	0.392	0.376	0.158	0.077	0.245	0.553	0.447
	Thoroughbred	0.230	0.355	0.348	0.146	0.063	0.231	0.521	0.479
	Hokkaido	0.035	0.067	0.023	0.033	0.002	0.062	0.128	0.872
ID_25	DB101	0.441	0.493	0.485	0.186	0.121	0.279	0.621	0.379
	Thoroughbred	0.404	0.482	0.494	0.183	0.116	0.276	0.615	0.385
	Hokkaido	0.314	0.431	0.488	0.169	0.093	0.259	0.583	0.417
ID_26	DB101	0.421	0.487	0.386	0.184	0.119	0.277	0.618	0.382
	Thoroughbred	0.348	0.454	0.382	0.175	0.103	0.267	0.599	0.401
	Hokkaido	0.058	0.110	0.116	0.052	0.006	0.096	0.201	0.799
ID_27	DB101	0.287	0.409	0.376	0.163	0.084	0.251	0.567	0.433
	Thoroughbred	0.309	0.427	0.506	0.168	0.091	0.257	0.581	0.419
	Hokkaido	0.430	0.490	0.628	0.185	0.120	0.278	0.620	0.380
ID_28	DB101	0.272	0.396	0.366	0.159	0.079	0.247	0.557	0.443
	Thoroughbred	0.258	0.383	0.404	0.155	0.073	0.242	0.546	0.454
	Hokkaido	0.151	0.257	0.302	0.112	0.033	0.187	0.414	0.586
ID_29	DB101	0.480	0.499	0.485	0.187	0.125	0.281	0.625	0.375
	Thoroughbred	0.472	0.498	0.562	0.187	0.124	0.281	0.624	0.376
	Hokkaido	0.047	0.089	0.093	0.042	0.004	0.079	0.166	0.834
ID_30	DB101	0.446	0.494	0.475	0.186	0.122	0.279	0.622	0.378
	Thoroughbred	0.376	0.469	0.438	0.180	0.110	0.272	0.608	0.392
	Hokkaido	0.419	0.487	0.465	0.184	0.118	0.277	0.618	0.382
ID_31	DB101	0.347	0.453	0.475	0.175	0.103	0.266	0.598	0.402
	Thoroughbred	0.326	0.439	0.539	0.171	0.097	0.262	0.589	0.411
	Hokkaido	0	0	0	0	0	0	0	1
ID_32	DB101	0.332	0.443	0.485	0.173	0.098	0.263	0.592	0.408
	Thoroughbred	0.309	0.427	0.303	0.168	0.091	0.257	0.581	0.419
	Hokkaido	0.419	0.487	0.465	0.184	0.118	0.277	0.618	0.382
ID_33	DB101	0.401	0.480	0.525	0.183	0.115	0.275	0.615	0.385
	Thoroughbred	0.461	0.497	0.607	0.187	0.123	0.280	0.623	0.377
	Hokkaido	0.035	0.067	0.070	0.033	0.002	0.062	0.128	0.872
ID_34	DB101	0.441	0.493	0.525	0.186	0.121	0.279	0.621	0.379
	Thoroughbred	0.365	0.464	0.506	0.178	0.107	0.270	0.605	0.395
	Hokkaido	0.012	0.023	0.023	0.011	0.000	0.022	0.045	0.955
ID_35	DB101	0.421	0.487	0.446	0.184	0.119	0.277	0.618	0.382
	Thoroughbred	0.427	0.489	0.427	0.185	0.120	0.278	0.619	0.381
	Hokkaido	0.198	0.317	0.395	0.133	0.050	0.215	0.483	0.517
ID_36	DB101	0.272	0.396	0.347	0.159	0.079	0.247	0.557	0.443
	Thoroughbred	0.309	0.427	0.303	0.168	0.091	0.257	0.581	0.419
	Hokkaido	0.477	0.499	0.535	0.187	0.124	0.281	0.624	0.376
ID_37	DB101	0.272	0.396	0.347	0.159	0.079	0.247	0.557	0.443
	Thoroughbred	0.197	0.316	0.326	0.133	0.050	0.215	0.482	0.518
	Hokkaido	0.419	0.487	0.372	0.184	0.118	0.277	0.618	0.382
ID_38	DB101	0.272	0.396	0.406	0.159	0.079	0.247	0.557	0.443
	Thoroughbred	0.337	0.447	0.539	0.174	0.100	0.264	0.594	0.406
	Hokkaido	0.419	0.487	0.558	0.184	0.118	0.277	0.618	0.382
ID_39	DB101	0.252	0.377	0.386	0.153	0.071	0.240	0.541	0.459
	Thoroughbred	0.180	0.295	0.270	0.126	0.043	0.205	0.459	0.541
	Hokkaido	0.012	0.023	0.023	0.011	0.000	0.022	0.045	0.955
Mean	DB101	0.373	0.456	0.445	0.175	0.105	0.267	0.597	0.403
	Thoroughbred	0.367	0.451	0.452	0.174	0.103	0.265	0.593	0.407
	Hokkaido	0.221	0.285	0.287	0.113	0.059	0.178	0.395	0.605

MAF: minor allele frequency, He: expected heterozygosity, Ho: observed heterozygosity, PE: probability of exclusion, PD: power of discrimination, PID: probability of identity.

Table 3. Cumulated probabilities of exclusion, power of discrimination, and probability of identity

Breed	Total PE1	Total PE2	Total PE3	Total PD	Total PID
DB101*	0.9994630250	0.9867623114	0.9999944587	>0.9999999999	0.9999999982
Thoroughbred	0.9994271269	0.9857432635	0.9999939355	>0.9999999999	0.9999999987
Hokkaido	0.9918696924	0.9102398590	0.9996430787	0.9999999998	>0.9999999999

*DB101: Allele frequency was extracted from a whole-genome variant database from 101 Thoroughbred horses (Tozaki *et al*., 2023 [23]).
PE: probability of exclusion, PD: power of discrimination, PID: probability of identity.

markers, whereas genomic DNA extracted from the urine was partially genotyped for 39 markers. Genotyping of urine extracts was performed four times, and the final genotype was confirmed synthetically (identical genotype with ≥2 markers). Finally, 23, 12, 11, 28, 6, and 20 markers were genotyped for BH, BP, BR, FP, LL, and PT, respectively. The genotypes of the markers obtained, except for one, were generally consistent between the blood and urine extracts. In the outlier case, a heterozygote (D/I) was genotyped as a homozygote (D/D). This finding can be attributed to the fact that the extracted genome was degraded, making the amplification of one allele difficult.

Discussion

As our first aim, we constructed a panel of 39 INDEL markers for parentage testing in Thoroughbred horses. The International Stud Book Committee has suggested >0.9995 as the total PE1 for a parentage marker panel using STRs in Thoroughbred registration [20], but the total PE1 for DB101 was lower than that (0.9994630250). However, because we confirmed genetic inheritance between 13 parent–child pairs and exclusion of 52 false parent–child pairs, the panel can be used for paternity testing in horses. At least, it could be used as a complement panel to the current paternity testing panel using STRs.

The constructed panel has two advantages: a low marker mutation rate and a genotyping procedure that is as simple as STR typing compared to SNP typing. STRs used for parentage testing in racehorses have multiple alleles; therefore, a panel can be constructed using a small number of markers [12]. However, this method has the disadvantage of a heightened mutation rate, consequently leading to Mendelian errors in parentage verification. To address this problem, panels using SNPs have been developed [7, 10]; however, SNP genotyping is not as straightforward as STR genotyping. The genotyping procedure for STRs only involves PCR amplification of markers and capillary electrophoresis, whereas those for SNPs (e.g., SNaPshot and MassArray) involve PCR amplification, enzymatic treatment, extension reaction, another enzymatic treatment, and then capillary electrophoresis or mass spectrometry analyses [7, 10]. The INDEL marker panel developed in the present study solved the problems with respect to both the mutation rate and operability of genotyping.

In our previous studies, we identified 12,173,068 SNVs and 1,566,396 INDELs in a population of 101 Thoroughbreds [22, 23]. Approximately 7.77 times fewer INDELs were detected than SNVs, indicating that they are less likely to occur than single-base substitutions. In humans (Danish), the de novo mutation rates of SNVs and INDELs were reportedly 1.27e-8 and 1.5e-9 per nucleotide per generation, respectively [1]. Their ratio of 8.47 (=1.27e-8 / 1.5e-9) was similar to the SNV/INDEL distribution ratio in the Thoroughbred genome. These observations suggest that INDELs are more stable markers than SNVs.

As our second aim, we considered using the constructed panel to complement doping testing. Doping tests in horse racing generally use urine as a sample [26]. Samples should generally be handled based on ISO 17025; however, unforeseen circumstances may obscure the origins of samples. In this case, distinguishing between urine samples may be possible using the developed INDEL panel, because of the total PD of >0.9999999999 and total PID of 0.9999999987. Although several markers were not well genotyped when using extracts from urine, a relatively appreciable number of markers were genotyped in three horses (BH, FP, and PT). This difference could be attributed to the number of cells in individual urine samples and their properties, which would explain the differences in the amounts of genomic DNA extracted. In addition, in this study, one genotype differed between urine and blood samples. Therefore, for genotyping using urine-derived extracts, one must consider that genotypes of amplified products may not necessarily match. Moreover, allelic discrimination of amplified products requires caution, as this study did not assess species specificity in PCR amplification.

Genotyping yielded better results with smaller markers than with larger markers, indicating the possibility of genomic DNA fragmentation in urine samples. Therefore, if urine-derived genomic DNA is used as a template, it may be necessary to redesign the primers to smaller sizes. In this study, 39 markers were constructed as a panel; however, it is possible to use them as markers because there are many other INDELs with MAFs of 0.25–0.75. Through PCR amplification with fluorescent-adapter labelling [21], individual markers can be genotyped by designing only nonlabelled forward primers (18-bp adaptor + specific sequences) and reverse primers (specific sequences). Therefore, modifying and adding markers at low cost is possible. Moreover,

one can possibly increase the total PD and total PID by redesigning the primers and/or by adding novel markers from the INDEL database, even if all markers in this panel were not successfully genotyped. When adding markers, PD and PID can be calculated from the allele frequency in the INDEL database.

An advantage of using INDELs as markers is that allele discrimination and the nomenclature are not difficult. The results (genotypes) of individual identification may be shared worldwide among various laboratories, as Thoroughbred horses are imported and exported. Because STRs have a tandem repeat structure, matching the PCR amplification size, number of repeats, and allele nomenclature between laboratories is difficult. In the case of INDELs, which have two alleles (deletion and insertion alleles), the allele from the PCR amplicon can be easily identified by comparison with the published reference genome sequence.

In this study, only Thoroughbreds and Japanese horses (Hokkaido) were analysed; therefore, it may not be possible to achieve similar discriminative ability if the identification panel is applied to other breeds. Three markers showed no polymorphisms in the Hokkaido population. Other breeds may have different allele frequencies. For this reason, prior investigation is required for application to other breeds.

Furthermore, as the INDEL marker in the panel basically has two alleles, genotypes can be expressed numerically by setting In/In=1, In/Del=2, Del/Del=3, and Null=9 as follows: 1232321232312-3123232123212-2221231293132 (13, 13, and 13 markers as FAM, HEX, and NED, respectively). Developing programmes for individual identification and paternity testing is easier by managing all racehorses using these numbers. Individuals can be identified and managed using DNA information as identification (numerical) information. Microchips are currently being used for individual identification of animals; however, the numbers proposed in this study may be used as alternative information.

In conclusion, we constructed an individual identification and paternity testing panel by selecting INDELs from an INDEL database constructed in a previous study as markers, and genotyping was performed on a separate population for validation, which generally matched the statistical information from the database. Therefore, the INDEL database may be used as a resource for various applications, such as parentage determination, individual identification, and sample identification.

Conflict of Interest

The authors declare no competing interests concerning this study, including patents, products in development, or marketed products.

Acknowledgments

We would like to thank Noriko Tanaka and Harue Sato for their assistance in this study.

References

1. Besenbacher, S., Liu, S., Izarzugaza, J.M., Grove, J., Belling, K., Bork-Jensen, J., Huang, S., Als, T.D., Li, S., Yadav, R., Rubio-García, A., Lescai, F., Demontis, D., Rao, J., Ye, W., Mailund, T., Friborg, R.M., Pedersen, C.N., Xu, R., Sun, J., Liu, H., Wang, O., Cheng, X., Flores, D., Rydza, E., Rapacki, K., Damm Sørensen, J., Chmura, P., Westergaard, D., Dworzynski, P., Sørensen, T.I., Lund, O., Hansen, T., Xu, X., Li, N., Bolund, L., Pedersen, O., Eiberg, H., Krogh, A., Børglum, A.D., Brunak, S., Kristiansen, K., Schierup, M.H., Wang, J., Gupta, R., Villesen, P., and Rasmussen, S. 2015. Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. *Nat. Commun.* **6**: 5969. [Medline] [CrossRef]
2. Czado, N., LaRue, B., Wheeler, A., Houston, R., Holmes, A., Grisedale, K., and Hughes, S. 2023. The effectiveness of various strategies to improve DNA analysis of formaldehyde-damaged tissues from embalmed cadavers for human identification purposes. *J. Forensic Sci.* **68**: 596–607. [Medline] [CrossRef]
3. de Groot, M., Anderson, H., Bauer, H., Bauguil, C., Bellone, R.R., Brugidou, R., Buckley, R.M., Dovč, P., Forman, O., Grahn, R.A., Kock, L., Longeri, M., Mouysset-Geniez, S., Qiu, J., Sofronidis, G., van der Goor, L.H.P., and Lyons, L.A. 2021. Standardization of a SNP panel for parentage verification and identification in the domestic cat (Felis silvestris catus). *Anim. Genet.* **52**: 675–682. [Medline] [CrossRef]
4. DeNise, S., Johnston, E., Halverson, J., Marshall, K., Rosenfeld, D., McKenna, S., Sharp, T., and Edwards, J. 2004. Power of exclusion for parentage verification and probability of match for identity in American Kennel Club breeds using 17 canine microsatellite markers. *Anim. Genet.* **35**: 14–17. [Medline] [CrossRef]
5. Durward-Akhurst, S.A., Schaefer, R.J., Grantham, B., Carey, W.K., Mickelson, J.R., and McCue, M.E. 2021. Genetic variation and the distribution of variant types in the horse. *Front. Genet.* **12**: 758366. [Medline] [CrossRef]
6. Fragkaki, A.G., Kioukia-Fougia, N., Kiousi, P., Kioussi, M., and Tsivou, M. 2017. Challenges in detecting substances for equine anti-doping. *Drug Test. Anal.* **9**: 1291–1303. [Medline] [CrossRef]
7. Hirota, K., Kakoi, H., Gawahara, H., Hasegawa, T., and Tozaki, T. 2010. Construction and validation of parentage testing for thoroughbred horses by 53 single nucleotide polymorphisms. *J. Vet. Med. Sci.* **72**: 719–726. [Medline] [CrossRef]
8. Jagannathan, V., Gerber, V., Rieder, S., Tetens, J., Thaller, G., Drögemüller, C., and Leeb, T. 2019. Comprehensive

characterization of horse genome variation by whole-genome sequencing of 88 horses. *Anim. Genet.* **50**: 74–77. [Medline] [CrossRef]
9. Jamieson, A., and Taylor, S.C. 1997. Comparisons of three probability formulae for parentage exclusion. *Anim. Genet.* **28**: 397–400. [Medline] [CrossRef]
10. Kakoi, H., Kijima-Suda, I., Gawahara, H., Kinoshita, K., Tozaki, T., Hirota, K., and Yoshizawa, M. 2013. Individual identification of racehorses from urine samples using a 26-plex single-nucleotide polymorphism assay. *J. Forensic Sci.* **58**: 21–28. [Medline] [CrossRef]
11. Kakoi, H., Kikuchi, M., Tozaki, T., Hirota, K.I., and Nagata, S.I. 2019. Evaluation of recent changes in genetic variability in Japanese thoroughbred population based on a short tandem repeat parentage panel. *Anim. Sci. J.* **90**: 151–157. [Medline] [CrossRef]
12. Kakoi, H., Nagata, S., and Kurosawa, M. 2001. DNA typing with 17 microsatellites for parentage verification of racehorses in Japan. *Anim. Sci. J.* **72**: 453–460.
13. Kalbfleisch, T.S., Rice, E.S., DePriest, M.S. Jr., Walenz, B.P., Hestand, M.S., Vermeesch, J.R. Jr., O Connell, B.L., Fiddes, I.T., Vershinina, A.O., Saremi, N.F., Petersen, J.L., Finno, C.J., Bellone, R.R., McCue, M.E., Brooks, S.A., Bailey, E., Orlando, L., Green, R.E., Miller, D.C., Antczak, D.F., and MacLeod, J.N. 2018. Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun. Biol.* **1**: 197. [Medline] [CrossRef]
14. Kimberly, A.H. 1998. Statistical analysis of STR data. *Profiles DNA* **1**: 14–15.
15. Lipinski, M.J., Amigues, Y., Blasi, M., Broad, T.E., Cherbonnel, C., Cho, G.J., Corley, S., Daftari, P., Delattre, D.R., Dileanis, S., Flynn, J.M., Grattapaglia, D., Guthrie, A., Harper, C., Karttunen, P.L., Kimura, H., Lewis, G.M., Longeri, M., Meriaux, J.C., Morita, M., Morrin-O'donnell, R.C., Niini, T., Pedersen, N.C., Perrotta, G., Polli, M., Rittler, S., Schubbert, R., Strillacci, M.G., Van Haeringen, H., Van Haeringen, W., and Lyons, L.A. 2007. An international parentage and identification panel for the domestic cat (Felis catus). *Anim. Genet.* **38**: 371–377. [Medline] [CrossRef]
16. Masuda, M., Nikadori, A., Nikadori, E., Uchida, N., Takizawa, Y., Ishimaru, M., Yamamoto, T., Yuasa, R., Kugai, H., Nagahama, Y., Takasu, M., and Tozaki, T. 2023. Evaluation of genetic diversity using 31 microsatellites in Miyako horses. *J. Equine Sci.* **34**: 1–6. [Medline] [CrossRef]
17. Paetkau, D., and Strobeck, C. 1994. Microsatellite analysis of genetic variation in black bear populations. *Mol. Ecol.* **3**: 489–495. [Medline] [CrossRef]
18. Peelman, L.J., Mortiaux, F., Van Zeveren, A., Dansercoer, A., Mommens, G., Coopman, F., Bouquet, Y., Burny, A., Renaville, R., and Portetelle, D. 1998. Evaluation of the genetic variability of 23 bovine microsatellite markers in four Belgian cattle breeds. *Anim. Genet.* **29**: 161–167. [Medline]
19. Tende, T., Hansson, B., Ottosson, U., Akesson, M., and Bensch, S. 2014. Individual identification and genetic variation of lions (Panthera leo) from two protected areas in Nigeria. *PLoS One* **9**: e84288. [Medline] [CrossRef]
20. Tozaki, T., Kakoi, H., Mashima, S., Hirota, K., Hasegawa, T., Ishida, N., Miura, N., Choi-Miura, N.H., and Tomita, M. 2001. Population study and validation of paternity testing for Thoroughbred horses by 15 microsatellite loci. *J. Vet. Med. Sci.* **63**: 1191–1197. [Medline] [CrossRef]
21. Tozaki, T., Mashima, S., Hirota, K., Miura, N., Choi-Miura, N.H., and Tomita, M. 2001. Characterization of equine microsatellites and microsatellite-linked repetitive elements (eMLREs) by efficient cloning and genotyping methods. *DNA Res.* **8**: 33–45. [Medline] [CrossRef]
22. Tozaki, T., Ohnuma, A., Kikuchi, M., Ishige, T., Kakoi, H., Hirota, K.I., Kusano, K., and Nagata, S.I. 2021. Rare and common variant discovery by whole-genome sequencing of 101 Thoroughbred racehorses. *Sci. Rep.* **11**: 16057. [Medline] [CrossRef]
23. Tozaki, T., Ohnuma, A., Kikuchi, M., Ishige, T., Kakoi, H., Hirota, K.I., Takahashi, Y., and Nagata, S.I. 2023. Short insertion and deletion discoveries via whole-genome sequencing of 101 Thoroughbred racehorses. *Genes (Basel)* **14**: 638. [Medline] [CrossRef]
24. Yoshihara, T., Kikuchi, M., Akiba, Y., Horiguchi, M., Takasu, M., and Tozaki, T. 2022. Genetic diversity analysis and parentage verification of Taishu horses using 31 microsatellites. *J. Equine Sci.* **33**: 63–69. [Medline] [CrossRef]
25. Wade, C.M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T.L., Adelson, D.L., Bailey, E., Bellone, R.R., Blöcker, H., Distl, O., Edgar, R.C., Garber, M., Leeb, T., Mauceli, E., MacLeod, J.N., Penedo, M.C., Raison, J.M., Sharpe, T., Vogel, J., Andersson, L., Antczak, D.F., Biagi, T., Binns, M.M., Chowdhary, B.P., Coleman, S.J., Della Valle, G., Fryc, S., Guérin, G., Hasegawa, T., Hill, E.W., Jurka, J., Kiialainen, A., Lindgren, G., Liu, J., Magnani, E., Mickelson, J.R., Murray, J., Nergadze, S.G., Onofrio, R., Pedroni, S., Piras, M.F., Raudsepp, T., Rocchi, M., Røed, K.H., Ryder, O.A., Searle, S., Skow, L., Swinburne, J.E., Syvänen, A.C., Tozaki, T., Valberg, S.J., Vaudin, M., White, J.R., Zody, M.C., Lander, E.S., Lindblad-Toh, K., Broad Institute Genome Sequencing Platform Broad Institute Whole Genome Assembly Team. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **326**: 865–867. [Medline] [CrossRef]
26. Wong, J.K.Y., Chan, G.H.M., Choi, T.L.S., Kwok, K.Y., Lau, M.Y., Leung, G.N.W., Wan, T.S.M., and Ho, E.N.M. 2020. A high-throughput and broad-spectrum screening method for analysing over 120 drugs in horse urine using liquid chromatography-high-resolution mass spectrometry. *Drug Test. Anal.* **12**: 900–917. [Medline] [CrossRef]