Genome-based development of 15 microsatellite markers in fluorescent multiplexes for parentage testing in tigers

Xiao Zhao[1,2,3,4], Qiguan Qiu[5], Chang Li[1,4,6], Dongke Fu[2,3,4], Xuesong Hu[2,4], Shengjie Gao[2,4], Yugang Zhu[7], Haofang Mu[8], Runping Wang[9], Huanming Yang[1,4,10], Bo Li[4]

[1] BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, China

[2] Forensic Genomics International (FGI), BGI-Shenzhen, Shenzhen, China

[3] Shenzhen Key Laboratory of Forensics, BGI-Shenzhen, Shenzhen, China

[4] BGI-Shenzhen, Shenzhen, China

[5] Changsha Ecological Zoo, Changsha, China

[6] BGI-Qingdao, BGI-Shenzhen, Qingdao, China

[7] Changsha Sanzhen Tiger Park, Changsha, China

[8] Center of Forensic Sciences, BGI, Beijing, China

[9] BGI Shaanxi Xixian new area Institute of Forensic Science, Xi'an, China

[10] China National GeneBank, BGI-Shenzhen, Shenzhen, China

Corresponding Authors:

Bo Li

BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Guangdong, 518083, China

Email address: libo@genomics.cn

Formatted: Width: 8.27", Height: 11.69"

Abstract

As one of the most endangered species, tiger (*Panthera tigris*) inbreeding has become an urgent issue to address. Using a microsatellite (short tandem repeat, STR) identification system, paternity testing may be helpful to avoid inbreeding in captive breeding programs. In this study, we developed a genome-based identification system named TPI-plex (tiger pedigree identification multiplex system). By analyzing the entire tiger genome, 139,967 STR loci were identified and 12.76% of these displayed three to six alleles among three re-sequenced individual tiger genomes. A total of 204 candidate STRs were identified and screened with a reference population containing 31 unrelated captive tigers. Of these, 15 loci were chosen for inclusion in the multiplex panel. The mean allele number and mean expected heterozygosity (H_e) were 7.3333 and 0.7789, respectively. The cumulative probability of exclusion (CPE) and total probability of discrimination power (TDP) reached 0.999999472 and 0.999999999999995, respectively. The results showed that the TPI-plex system can be applied in routine pedigree identification for captive tigers. We also added a sex identification marker named TAMEL into the TPI-plex for sex determination.

1.Introduction

The tiger (*Panthera tigris*) was listed as an endangered (EN) species by the International Union for Conservation of Nature (IUCN) in 1986. As a recognized keystone species, tigers play a key role in maintaining healthy ecosystems *(Cho et al.*

Deleted: individuals

Deleted: heterozigosity

Deleted: World Conservation Union

Deleted: Three out of nine known subspecies were extinct and the South China tiger (*P.t. amoyensis*) was extinct in the wild *(Luo et al. 2004; Luo et al. 2010).*

Formatted: Font: Italic

2013). Unfortunately, the rapid loss of tigers was remarkable. Fewer than 4000 wild tigers survive in areas that occupied only 7% of their historical range and the tigers were divided into two currently recognized subspecies *(Kitchener et al. 2017; Wilting et al. 2015)*. The World Wildlife Fund (WWF) worked in alliance with local governments and agencies to protect wild tiger populations. Moreover, there has been an effort to conserve tigers through captive breeding programs and the efforts appeared to have paid off. The captive tiger population has greatly outnumbered the wild ones greatly since 2007 *(Luo et al. 2008)*. In China, the number of captive tiger population has grown rapidly since 2002 and has reached between 5,000 and 6,000 in the last two years (https://eia-international.org/where-are-the-tigers/).

For EN animals, the real purpose of protective captivity should not only concentrate on population expansion but genetic diversity preservation *(Reed & Frankham 2003*). Inbreeding led to reduced genetic diversity and had a deleterious effect on the biological fitness of the population *(Keller & Waller 2002; Reed & Frankham 2003; Ruiz - López et al. 2012)*. Research showed that tigers are one of the most inbred animals in captivity (*Begany & Criscuolo 2009*).The captive South China tiger population was found to be suffering from inbreeding depression and a decline in genetic diversity *(Xu et al. 2007)*. High levels of inbreeding brought stillbirths and high infant mortality and very low genetic diversity within the captive population of Asian lions *(Atkinson et al. 2018)*. Having a better understanding of the mating system of a species is the foundation to a successful breeding and captive management *(Ferrie et al. 2013)*. Mastering the parent-child relationship of captive

Deleted: *(Cho et al. 2013; Luo et al. 2004)*

Formatted: Font color: Text 1

Formatted: Font: Italic, Font color: Text 1

Formatted: Font color: Text 1

Deleted: a

Deleted: s

Deleted: s

Deleted: has shown

Formatted: Font color: Text 1

Formatted: Font: Italic, Font color: Text 1

Formatted: Font: Italic, Font color: Text 1

Formatted: Font color: Text 1

tigers is highly recommended in avoiding captive tiger inbreeding. The breeding programs of many captive species managed by zoos or other organizations depend on studbooks to record individual pedigrees *(Ferrie et al. 2013; Jones et al. 2002)*. Accurate and complete pedigree information is essential for effective pedigree analysis *(Ferrie et al. 2013)*. However, the recorded data in the studbook may be missing or incorrect and the information in the studbook should be supported by genetic analysis *(Ferrie et al. 2013; Xu et al. 2007).*

Microsatellites (short tandem repeats, STRs) proved to be one of most powerful genetic markers for kinship analysis of animals and have been generally applied for this purpose *(Ichikawa et al. 2001; Luikart et al. 1999; Pei et al. 2018; Queller et al. 1993; Webster & Reichart 2005).* Isolating polymorphic microsatellites efficiently from the species genome is a crucial precondition of paternity testing for the proven STR method *(Webster & Reichart 2005)*. Until now, there have been two alternative sources of microsatellite loci for the STR-based method of tiger paternity identification. First, one may select a set of microsatellite primers derived from the domestic cat *(Felis catus) (Menotti-Raymond et al. 1999)* to amplify the tiger's microsatellite DNA *(Wu et al. 2011; Zhang et al. 2003)*. These investigations introduced an initial single-locus amplification in the target DNA and failed to provide the accuracy of paternity testing by calculating the cumulative probability of exclusion (CPE). Second, one may isolate microsatellite loci from the tiger genome *(Sharma et al. 2008; Williamson et al. 2002; Wu et al. 2009; Zhang et al. 2006a; ZHANG et al. 2006b)*. These investigations preselected specific repeated motifs as probes without knowing their abundance in the tiger genome and therefore inevitably introduced biases and limited the microsatellite types into a small fixed subset *(Castoe et al. 2010; Malausa et al. 2011)*. These methods are both quite labor-intensive.

Therefore, establishing a more efficient method for screening polymorphic loci unbiasedly from all types of microsatellite loci present in the tiger genome were essential for establishing a paternity test ~~for the tiger~~. Here, we developed a multiplex system in a single reaction tube that can serve as a convenient, effective, and accurate method for paternity testing in the captive tiger. In this study, we screened highly polymorphic microsatellite loci ~~in the tiger~~ on a genome-wide scale using bioinformatics analysis. We also validated the tiger amelogenin locus based on homology analysis for sex identification. We incorporated all the selected autosomal STR loci and the sex determination locus into one polymerase chain reaction (PCR) to establish an STR five-color fluorescent-multiplex system for simple and effective use. We used a reference population that consists of a group of unrelated ~~tiger~~ individuals to investigate alleles, allelic frequencies, genotypes, genotype frequencies of each STR locus and assessed the validity and accuracy of the multiplex system for paternity testing and individualization in captive tigers.

Deleted: attempted to

Formatted: Strikethrough, Highlight

Formatted: Strikethrough, Highlight

Formatted: Strikethrough, Highlight

2. Materials and Methods

2.1 Sample collection

A total of 42 captive continental tigers were sampled ~~in this study~~ (*Table S1*). 20

Formatted: Font color: Text 1

Deleted: captive tiger samples

Deleted: ~~samples~~

Formatted: Font color: Text 1, Strikethrough, Highlight

Formatted: Strikethrough, Highlight

Formatted: Font color: Text 1

blood samples were collected from individuals T01-T20 via the femoral vein after anesthetization and 22 hair samples were collected from individuals RT01-RT22 (*Table S1*). Individuals T01-T20 are from Changsha Ecological Zoo and their parent-child relationships ~~were expected to be identified~~. Individuals T12-T20 are parents and do not have blood relationship, but they may mate and produce offspring (T01-T11). There into, T01, T02, T06, T07, T08 and T09 six tigers were born from the same womb, and T03, T04 and T05 three tigers were from the same womb. However, ~~which parents were~~ the ~~right pair of their~~ biological parents were in doubt. The available recorded information was provided by the zoo staff (*Table S2*). Individuals RT01 to RT22 are other 22 unrelated tigers from different places. The study was approved by the *Institutional Review Board on Bioethics and Biosafety of BGI* (FT 16084).

2.2 DNA isolation

Genomic DNAs from whole blood were extracted by TIANamp Blood DNA Kits® (TIANGEN Biotech Beijing), following the manufacturer's instructions. The hair shaft of each hair sample was cut off and the remaining part containing the hair follicle was placed into a 1.5 ml Eppendorf tube and washed with double distilled water and absolute ethyl alcohol, respectively and digested by proteinase K. The genomic DNA from the hair samples was isolated using Chelex 100.

Deleted: 20 blood samples were collected from Bengal tigers (*Panthera tigris tigris*, T01-T20,

Deleted: via the femoral vein after anesthetization and 22 hair samples were collected from other individuals (RT01-RT22, *Table S1*). The 20 Bengal tigers from Changsha Ecological Zoo kept unclear relationships and their partially known and probable genealogic information was provided by the zoo staff (*Table S2*).

Formatted: Font color: Text 1

Commented [AA1]: I do not understand – why expected? Do you mean they served as positive test for paternity assessment?

Formatted: Font color: Text 1, Strikethrough, Highlight

Formatted: Font color: Text 1

Deleted: don't

Formatted: Font color: Text 1

Deleted: Thereinto

Formatted: Font color: Text 1

Formatted: Font color: Text 1, Strikethrough, Highlight

Formatted: Font color: Text 1

Formatted: Font color: Text 1, Strikethrough, Highlight

Formatted: Font color: Text 1

Deleted: manufaturer's

Deleted: Eppenddorf

2.3 Selection of tiger markers

We downloaded the tiger reference genome (GCA_000464555.1 PanTig1.0) from the NCBI database and used Tandem Repeats Finder (v4.09) *(Benson 1999)* for annotating the STR loci in the reference genome. We also downloaded the re-sequenced data from three tigers from the NCBI database (SRR640236, SRR640237 and SRR640238) *(Xu et al. 2013*) and filtered the raw data by SOAPnuke (v2.0) *(Chen et al. 2017)*. Based on the annotation results, we ~~allelo~~typed the STRs of the three tigers using lobSTR *(Gymrek et al. 2012)*. We screened all of the valid STR loci ~~for those~~ with high polymorphisms that satisfied the following criteria: (1) the locus is a tetra-nucleotide or penta-nucleotide repeat, (2) the repeat units repeat 10 to 20 times in the reference genome, (3) three to six alleles in three re-sequenced individuals exist at the site.

To identify a tiger's sex, we validated the amelogenin locus based on homology analysis. We downloaded human amelogenin sequences (AMELX and AMELY) from the NCBI database, aligned them to the tiger genome, and found two homologous sequences in the tiger. Clustal X (v2.1) *(Jeanmougin et al. 1998)* was used to find the deletion polymorphism of amelogenin sequences. Primer pairs were designed in flanking regions using OLIGO 7 (v7.56) *(Rychlik 2007)* and the size range of the amplification products was controlled between 100-500 bp. The specificity of the primers was validated by PCR, and loci that could be easily amplified were reserved. The forward primers of the normal primer pairs were then labeled with different fluorescent-dye. Loci with inefficient amplification or split

Formatted: Strikethrough, Highlight

Formatted: Strikethrough, Highlight

Deleted: included the

Deleted: two

Deleted: in the tiger

Deleted: Normal

Deleted: the appearance of a

peaks were excluded.

2.4 Multiplex amplification assay

The selected loci were incorporated into a multiplex amplification system. For the loci whose primers were labeled with the same dye color, their allele size ranges did not overlap each other. All of the fluorescence-labeled primer pairs were mixed in proportions that were determined from the results of multiple experiments guaranteeing good amplification at each site and peak height during capillary electrophoretic separation. The final primer concentration of each primer pair is in the range of 0.2 to 1.3μM/μL. The primer mixture was used to amplify genomic DNAs by PCR and amplification was performed in a 10 μL reaction volume. The reaction mixture contains 50 mM KCl, 10 mM Tris-HCl (pH 8.3, 25°C), 2.0 mM $MgCl_2$, 0.1 mg/ml BSA, 0.2 mM dNTPMix (dATP, dTTP, dCTP and dGTP mixed equally), 0.2 units DNA polymerase (EzAmp® Fast Taq DNA Polymerase) and 0.1-2 ng genomic DNA. For each reaction, the PCR conditions were as follows: 1 cycle of 95 °C for 5 min, 30 cycles of 95 °C for 10 s, 58 °C for 1 min, 70 °C for 20 s, and 1 cycle of 60 °C for 1h with Applied Biosystems® Veriti® Thermal Cycler.

2.5 Electrophoresis separation and data analysis

The PCR products (1 μl) were mixed with loading buffer composed of Hi-DiTM formamide and internal size standard Salmon 500 Plus at a 9:0.3 (v/v) ratio. The electrophoretic separations were performed on an ABI 3500 Genetic Analyzer

Deleted: should

Deleted: The distances between the adjacent size ranges should remain in consideration of other alleles that were not included in the defined allele size ranges or even for new alleles that may be discovered in the future.

Deleted: an applicable

Deleted: as

Deleted: .

Deleted: This proportion

Deleted: s

Deleted: a

Deleted: effect

Deleted: clear signals of

Formatted: Font color: Text 1

Deleted: 、

Deleted: 、

(Thermo Fisher, USA). Then the collected data were analyzed with GeneMapper® ID-X Software Version 1.5.

2.6 Allele sequencing and genotyping

A homozygote was selected at each locus and the amplification products were sent to Sangon Biotech (Sangon, Shanghai, China) for Sanger sequencing in both forward and reverse directions after performing agarose gel electrophoresis for validation. The sequenced alleles were named by the number of repetitions of the motif according to the nomenclature of STR recommended by the International Society for Forensic Genetics (ISFG). The repetitions and names of all the remaining unsequenced alleles were deduced according to their observed size and the sequenced alleles. The panel and bin files were programmed for the GeneMapper® ID-X Software Version 1.5 to genotyping analysis.

2.7 Population genetic analysis

From the 42 captive continental tigers, a total of 31 unrelated tigers (T12-T20 and RT01-RT22) were selected as the reference population for population genetic analyses of the 15 autosomal STR loci. The Hardy–Weinberg equilibrium (HWE) in each locus was tested using the x^2-test. The sample size of the investigated population was not large as it was restricted by objective limiting conditions and there were cases in which the number of genotypes was less than five and even some alleles were not observed. The data structure was then adjusted by merging alleles with frequencies of

Formatted: Font color: Text 1

Deleted: The allele at each locus that was amplified from the homozygote was sent to Sangon Biotech (Sangon, Shanghai, China) for Sanger sequencing in both forward and reverse directions after performing agarose gel electrophoresis for validation.

Deleted: repeat

Deleted: unit

Deleted:

Deleted: r

Deleted: ules

Commented [AA2]: Reference needed.

Commented [AA3]: In contradiction with lines 263, 264

Formatted: Highlight

Commented [AA4]: Proper genotyping/ allele assignment would require allelic ladders. It seems they were not made/used. Am I right?

Formatted: Highlight

Formatted: Font color: Text 1

Deleted: individuals

Deleted: , including eighteen Bengal tigers (T12-T20, RT08, RT11-RT15, RT18, RT19, and RT22), eleven Amur tigers (RT01-RT07, RT09, RT10, RT17, and RT21) and two South Chinese tigers (RT16 and RT20)

Deleted: enough in our study

Commented [AA5]: I do not understand – how could some alleles not be observed?

less than 0.1 into one group, in order to allow performing χ^2-test. Expected heterozygosity (H_e: Equation 1), probability of exclusion (PE: Equation 2), and discrimination power (DP: Equation 3) of each locus, cumulative probability of exclusion (CPE: Equation 4), and cumulative discrimination power (TDP: Equation 5) according to the formulas below. Undetected and null alleles were removed from consideration in all of these calculations.

$H_e = 1 - \sum_{i=1}^{k} p_i^2$ (Equation 1, k is the number of alleles and p_i the allele frequency of the i^{th} allele at the target locus)

$PE = \sum_{i=1}^{k} p_i (1 - p_i)^2 + 1/2 \sum_{i=1}^{k-1} \sum_{\substack{j=i+1 \\ i \neq j}}^{k} p_i^2 p_j^2 (4 - 3p_i - 3p_j)$ (Equation 2, p_i and p_j are respectively the allele frequency of the i^{th} and the j^{th} allele at the target locus, with i not equal to j)

$CPE = 1 - \prod_{i=1}^{k} (1 - PE_k)$ (Equation 3, PE_k is PE for each of k loci)

$DP = 1 - \sum_{i=1}^{k} p_i^2$ (Equation 4, k is the number of phenotypes and p_i the genotype frequency of the i^{th} genotype at the target locus.)

$TDP = 1 - \prod_{i=1}^{k} (1 - DP_k)$ (Equation 5, DP_k is DP for locus k.)

2.8 Sensitivity testing

To evaluate the sensitivity of the TPI-plex amplification system, the DNA from the individual of T08 was chosen as the control DNA and used to perform the sensitivity testing experiment. A series of template DNA quantities were diluted in a 10 μL PCR reaction system and the DNA concentrations, from high to low, were as follows: 2 ng 10 μL^{-1}, 1 ng 10 μL^{-1}, 0.5 ng 10 μL^{-1}, 0.25 ng 10 μL^{-1}, 0.125 ng 10 μL^{-1}, 0.0625 ng 10

Commented [AA6]: Debatable – you could have performed an exact test as provided, for instance by Arlequin or Genepop software.

Deleted: the

Deleted: to verify whether the investigated population conformed to the Hardy-Weinberg equilibrium

Deleted:

Deleted: total …umulative probability of

Commented [AA7]: Reference(s) needed for the formulas used

Deleted: of the genetic marker system were calculated on the basis of the observed allele frequency and genotype frequency.

Formatted: Font color: Red

Formatted: Highlight

Commented [AA8]: I do not understand – do you mean that a formal model of total codominance was applied?

Formatted: Highlight

Deleted: here

Deleted: at the target locus

Deleted: is

Deleted: .

Deleted: and …ith i is …ot equal to j.

Deleted: here

Deleted: locus … loci.

Commented [AA9]: This formula is not correct nor the description; please check the used reference

Formatted: Highlight

Formatted: Highlight

Formatted

Deleted: here … is the number of phenotypes of the target locus

Formatted

Deleted: is

Formatted

Deleted: here

μL^{-1}, and 0.03125ng 10 μL^{-1}. Each quantity of DNA was analyzed in triplicate and a negative control group was set up.

2.9 Specificity testing

As DNA from human or other non-human species may be mix in detected material, the TPI-plex amplification system was tested with DNA from a range of species including human, sheep, chick, duck, dog and rat, under the same PCR amplification condition to estimate potential interference. Genomic DNAs from human and dog samples was extracted from whole blood with TIANamp Blood DNA Kits® (TIANGEN Biotech Beijing), while others were extracted from fresh tissue (purchased from markets) using TIANamp Genomics DNA Kits® (TIANGEN Biotech Beijing). Each DNA was analyzed in duplicate and a negative control group was set up.

3. Results and discussion

3.1 Establishment of TPI-plex identification system

3.11 Selection of tiger markers

Microsatellites have a high mutation rate as compared to other types of known genetic polymorphisms *(Ballantyne et al. 2010; Webster & Reichart 2005)*. Microsatellites are codominant markers, heterozygotes can be distinguished from homozygotes at microsatellite loci, which contribute to accurate genotyping *(Webster & Reichart*

Deleted: variation

Formatted: Highlight

Formatted: Highlight

2005). In addition, microsatellites analyses are based on simple PCR, which allows for typing in samples of low DNA quality or concentration *(Dawson et al. 2013; Webster & Reichart 2005)*. Therefore, microsatellites are the most commonly used genetic marker in parentage identification of animals *(Vignal et al. 2002; Webster & Reichart 2005)*, the paternity testing in tigers is no exception *(Wu et al. 2011; Zhang et al. 2003)*. A total of 139,967 valid STR loci were identified across the whole tiger genome and 31.48% and 12.76% of these STR loci display two and three to six alleles (*Fig. S1A*), respectively. As the tiger genome sequence shows a 95.6% similarity to that of the domestic cat *(Cho et al. 2013)*, we mapped all the detected STR loci from the tiger genome onto a cat karyotype, based on their genomic homology, to show their distribution on chromosomes *(Fig. S1B)*. We generated 204 candidate STR loci (*Table S3*) of which 84.11% displayed three alleles and covered nineteen chromosomes of the tiger (*Fig. 1*). Then, primers were designed for 49 STR loci of 204 candidate STR loci based on the tiger reference and 27 pairs of primers on 27 STR loci were effective. Three loci were excluded because of weak or split peaks after the detection of fluorescent-labeled PCR products. The remaining 24 loci were distributed on 14 chromosomes and were named as DA1S1290, DA1S1470, DA2S1059, DA2S1575, DA3S1145, DA3S461, DA3S1123, DB1S1259, DB1S1096, DB1S542, DB2S734, DB2S23, DB3S187, DB4S1505, DB4S2706, DB4S2753, DC1S1364, DD2S793, DD3S899, DD3S86, DD4S705, DE1S613, DF1S579 and DF2S497, respectively (*Fig. 1*).

In humans, sex identification in forensic multiplexes is based upon the

Commented [AA10]: Nonsense – please omit; STRs are not intrinsically codominant – it depends on the typing methodology

Formatted: Highlight

Deleted: We also used microsatellite makers in our study.

Deleted: for

Deleted: peak

Deleted: the appearance of a

amelogenin gene on both the X and Y chromosomes, which is commonly used in sex genotyping *(Akane et al. 1991; Nakahori et al. 1991)*. Similarly, we discovered two sequences of the tiger amelogenin gene on scaffolds of ATCQ01070658.1 and ATCQ01738.1 by homology searching (*Fig. S2*). Interestingly, on the tiger amelogenin locus, we found a deletion polymorphism (20 bp) which may be used to identify the tiger's sex. We named the tiger amelogenin locus *TAMEL*, amelogenin X-linked as *TAMELX* on ATCQ01070658.1 and amelogenin Y-linked as *TAMELY* on ATCQ01738.1, respectively.

3.12 TPI-plex identification system development

We established an STR five-color fluorescent-multiplex system, with 15 autosomal STRs and a sex identification locus. (*Table 1*). The 16 loci (DA1S1290, DA1S1470, DA2S1059, DA2S1575, DA3S1123, DA3S1145, DB1S542, DB1S1259, DD2S793, DD3S86, DD4S705, DE1S613, DF2S497, DF1S579, and TAMEL; *Fig. 1*) are distributed on 12 chromosomes and the 16-plex identification system used fluorescent forward primers labeled at the 5' end with blue (6'-FAM), green (HEX), yellow (TAMRA), or red (ROX) dyes (*Table 1*).

To characterize each locus, we sequenced PCR amplification products from homozygotes *(Xu et al. 2005)*, and sequencing results (*Table S4*), provided the repeat numbers of motif units (*Table S5*). Sequenced alleles were defined on the basis of the nomenclature of STRs *(Mayr et al. 1995)*. Most of these loci are simple repeats of tetra-nucleotide or penta-nucleotide motifs (*Table S5*). DE1S613, DF1S579,

Deleted: ..

Deleted: . There are

Deleted: loci

Deleted: could effectively amplify in a multiplex system

Deleted: A total of

Deleted: were eventually retained

Deleted: labeled in black in

Deleted: nd were

Deleted: , which together formed a

Deleted: .

Deleted: of the 16 loci were

Deleted: explore

Deleted: stics of

Deleted: .

Deleted: From the

Deleted: we got the

Deleted: times

Deleted: basic

Deleted: with

DA3S1145, and DD2S793 are compound repeats due to their two different types of motifs. DB1S1259 and DD2S705 are complex repeats composed of tetra- and a few penta-nucleotide motifs. Their alleles were designated using the method which assumed the region was a general tetrameric repeat structure *(Hellmann et al. 2006)*. Eleven out of the fifteen loci had the same repeat sequences as the reference genome, while four loci (DB1S1259, DE1S613, DF1S579, and DA3S1145) displayed polymorphisms in repeat sequences.

3.2 Evaluation of the TPI-plex identification system

3.21 The CPE and TDP of TPI-plex system

To assess the TPI-plex, we calculated the allele (*Table 2*) and genotype frequencies (*Table S6*) in the reference population (n=31). We deduced repeat motifs of unsequenced alleles through their observed size based on the homologous allele's observed size and corresponding sequenced motif (*Table S5*). We provided the peak ratio of each locus for correct allele calling *(Table S5)*. We ran this TPI-plex for each individual in the reference population (n=31), and found that all of the STRs are polymorphic (*Tables 2 and S6*), and allele number ranged from five to eleven. DA3S1145, DA1S1290, DA1S1470, DA2S1059, DB1S1259, and DF1S579 displayed a departure from HWE (*P* < 0.05, *Table 3*).

We evaluated the TPI-plex power in individualization and paternity testing. The expected heterozygosity ranged from 0.6852 to 0.8902, and the average H_e was 0.7789. The PE of each locus ranged from 0.4183 to 0.8183, average PE being 0.6028

Deleted: categorized as

Deleted: belong to

Deleted: in which the repeat regions are

Deleted: nucleotide motifs

Commented [AA11]: Requires better explanation; I cannot see how fragment length in bp is translated into repeat numbers. Please provide also a table where amplicon length is provided along with numerical nomenclature for each allele. Table S5 does not seem to apply correctly the nomenclature rules.

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Commented [AA12]: See previous comment

Formatted: Highlight

Deleted: .

Deleted: A

Deleted: in the population

Deleted: Interestingly, all of the DB1S1259, DE1S613, and DD4S705 have eleven alleles.

Formatted: Font color: Red

Commented [AA13]: Too many loci show significant deviation; need explanation. A genotype/phenotype distribution table would be required.

Formatted: Font color: Red

Deleted: further

Deleted: 's

Deleted: fitness

Deleted: determining

Deleted: for

Deleted: (H_e)

Deleted: . The

Deleted: of 15 loci was

and the CPE reached 0.999999472 (*Table 3*). The DP ranged from 0.8262 to 0.9417 with an average of 0.8808. The TDP of all the 15 loci reached 0.999999999999995 (*Table 3*). In a previous study using 21 microsatellite loci and more than one multiplex panel for the Amur tiger *(Wu et al. 2009)*, the CPE reached 0.9999, and more than one multiplex panel was used in the method. Compared with this report, our method greatly simplifies this process and is more efficient.

Deleted: The

Deleted: isolated

Deleted: in all

Deleted:

Deleted: e previous study,

Deleted: accurate

3.22 Sensitivity testing

Sensitivity testing can be used to find the DNA template usage limitation of the multiplex typing assay *(Zhang et al. 2015)*. Here, we used DNA from the individual of T08 as control DNA. We performed the typing assay in triplicate in the range of total input DNA (0.03125~2 ng 10 μL^{-1}) under the same PCR conditions. We calculated the mean percentage of detected sites in the sensitivity testing and found that all 16 loci can be detected with DNA from 2 ng down to 0.5 ng. When the DNA amount were 0.25ng, 0.125 ng, 0.0625 ng, and 0.03125 ng, the mean percentages of detected loci were 93.75%, 93.75%, 37.5%, and 18.75%, respectively. When the DNA template amount decreased to 0.125ng, only one loci could not be detected. Furthermore, we calculated the mean peak height in the sensitivity testing. When the DNA template amount varied from 2 ng down to 0.5 ng, the mean peak height was from 10,950.37 relative fluorescence units (RFU) down to 5,752.40 RFU. In summary, the minimal DNA template was 0.5 ng for the TPI-plex.

3.23 Species specificity

We performed the TPI-plex on the common species of human, sheep, chick, duck, dog, and rat and showed that there was an off-range peak 180 to 210 bases long at DA2S1059 in human (193.0 bases), sheep (193.2 bases), dog (207.9 bases), and rat (182.9 bases). We detected a second OL peak at DC1S1364 in sheep (357.0 bases), and dog (372.6 bases).

Commented [AA14]: No ladder was made/employed.

Deleted: ladder (OL)

Deleted: ranging from

3.3 Applications of the TPI-plex identification system

3.31 Sex determination

Tigers have a chromosomal XY sex-determination system. The sex of tigers can be identified by analyzing the amplification products of the sex identification marker, TAMEL. We added the sex identification marker to the multiplex system for the first time and designed paired primers at both ends of the 20 bp deletion sequences. At the TAMEL locus, a male and female tiger display heterozygous (87 bp and 109 bp, namely, Y and X, *Fig. 2A,* boxed in carmine) and homozygous bands (109 bp, namely, X, *Fig. 2B,* boxed in carmine), respectively. Furthermore, we validated this sex marker by identifying the sexes of a group of tigers with known sex information (*Table S1*: RT01, RT04, RT05, RT06, RT07, RT08, RT10, RT11, RT14 and RT15) using the TPI-plex system and compared our detected sex with information from these tiger’s breeders. We found that our detection on all individuals was correct (*Fig. S3*), suggesting that the TAMEL locus can identify the sex correctly.

Deleted: an

3.32 Tiger pedigree identification

Using STR genotyping results, we determined the parent-child relationship between the 20 tigers from Changsha Ecological Zoo and validated the pedigree reconstruction using information provided by the zoo. As shown in *Fig. 3*, there were three families in all (Family 1: T13, T19, T01, T02, T06, T07, T08 and T09, *Fig. 3A*; Family 2: T13, T20, T03, T04 and T05, *Fig. 3A*; and Family 3: T15, T18 and T10, *Fig. 3B*) and T13 connects two families as having a common father. Another 4 individuals, T12, T14, T16 and T17, have no blood relationships with the others. Table S7 lists genotyping information of the three families.

We checked information from the zoo and found that T01, T02, T06, T07, T08, and T09 tigers were born from the same womb (*Table S2*). These six individuals and T11 shared the same dam, T19. Our genotyping results show that T19 was indeed biological mother of the seven tigers, since one of two alleles at each locus in the seven offspring individuals inherited from T19, which was in accordance with Mendelian inheritance. In addition, T03, T04 and T05 were three tigers born in a single birth.

Interestingly, zoo staffs cannot confirm whether white tiger, T10, is from orange tiger family (T16 and T14) or from white tiger family (T18 and T15). Our identification results showed that T15 and T18 were biological parents of T10.

Deleted: from

Conclusions

We used the bioinformatics analysis method to identify tiger microsatellite loci on a genome-wide scale for the first time and screened 15 highly polymorphic microsatellite loci distributed on 11 chromosomes from the tiger genome. We obtained a sex determination locus by validating the tiger amelogenin locus based on homology analysis. The 15 loci together with the sex determination locus were incorporated into one PCR reaction and a STR five-color fluorescent-multiplex system named TPI-plex was established. The TPI-plex system's CPE and TDP reached 0.999999472 and 0.999999999999995, respectively, suggesting that this TPI-plex can be applied for routine pedigree identification and individualization for tigers. The sex identification locus provided sex information of individuals.

Compared with other methods, our identification process is time- and cost-saving, as the TPI-plex system is a single reaction multiplex system. Our research could contribute to the supplementation and correction of studbook records by identifying and verifying the pedigree relationships among captive individuals and could also play a positive role in promoting pedigree management and breeding control of tigers in captive institutions. It is of great significance to effectively avoid inbreeding in order to protect the genetic diversity of captive tigers.

Acknowledgements

We thank Changsha Ecological Zoo and Sanzhen Tiger Park for providing the biological samples for the tigers. We thank Rihua Yang, Yixin Zhu, Songjiao Liu and Dami Jiang from the Key Laboratory of Forensics of BGI-Shenzhen for providing technical support and equipment debugging. We also thank Dong Li from Forensic Genomics International (FGI) of BGI-Shenzhen for providing us samples of human, sheep, chick, duck, dog, and rat.

References

Akane A, Shiono H, Matsubara K, Nakahori Y, Seki S, Nagafuchi S, Yamada M, and Nakagome Y. 1991. Sex identification of forensic specimens by polymerase chain reaction (PCR): Two alternative methods. *Forensic science international* 49:81-88. https://doi.org/10.1016/0379-0738(91)90174-H

Atkinson KE, Kitchener AC, Tobe SS, and O’Donoghue P. 2018. An assessment of the genetic diversity of the founders of the European captive population of Asian lion (*Panthera leo leo*), using microsatellite markers and studbook analysis. *Mammalian biology* 88:138-143.

Ballantyne KN, Goedbloed M, Fang R, Schaap O, Lao O, Wollstein A, Choi Y, van Duijn K, Vermeulen M, and Brauer S. 2010. Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *The American Journal of Human Genetics* 87:341-353.

Begany L, and Criscuolo C. 2009. Accumulation of deleterious mutations due to inbreeding in tiger population.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27:573-580.

Castoe TA, Poole AW, Gu W, Jason de Koning A, Daza JM, Smith EN, and Pollock DD. 2010. Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Molecular Ecology Resources* 10:341-347.

Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, and Li Z. 2017. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* 7:gix120.

Cho YS, Hu L, Hou H, Lee H, Xu J, Kwon S, Oh S, Kim H-M, Jho S, and Kim S. 2013. The tiger genome and comparative analysis with lion and snow leopard genomes. *Nature communications* 4:1-7.

Dawson DA, Ball AD, Spurgin LG, Martín-Gálvez D, Stewart IR, Horsburgh GJ, Potter J, Molina-Morales M, Bicknell AW, and Preston SA. 2013. High-utility conserved avian microsatellite markers enable parentage and population studies across a wide range of species. *BMC Genomics* 14:176.

Ferrie GM, Cohen OR, Schutz P, Leighty KA, Plasse C, Bettinger TL, and Hoffman EA. 2013.

Field Code Changed
Formatted: Font: Times Font
Formatted: Font: Times Font
Formatted: Font: Times Font
Formatted: Font: Times Font, Italic
Formatted: Font: Times Font
Deleted: D
Formatted: Font: Times Font
Deleted: M
Formatted: Font: Times Font
Deleted: D
Formatted: Font: Times Font
Deleted: I
Formatted: Font: Times Font
Deleted: T
Formatted: Font: Times Font
Deleted: P
Formatted: Font: Times Font
Formatted: Font: Times Font, Italic
Formatted: Font: Times Font

Identifying parentage using molecular markers: improving accuracy of studbook records for a captive flock of marabou storks (*Leptoptilus crumeniferus*). *Zoo biology* 32:556-564.

Gymrek M, Golan D, Rosset S, and Erlich Y. 2012. lobSTR: a short tandem repeat profiler for personal genomes. *Genome research* 22:1154-1162.

Hellmann AP, Rohleder U, Eichmann C, Pfeiffer I, Parson W, and Schleenbecker U. 2006. A proposal for standardization in forensic canine DNA typing: allele nomenclature of six canine-specific STR loci. *Journal of forensic sciences* 51:274-281.

Ichikawa Y, Takagi K, Tsumagari S, Ishihama K, Morita M, Kanemaki M, Takeishi M, and Takahashi H. 2001. Canine parentage testing based on microsatellite polymorphisms. *Journal of Veterinary Medical Science* 63:1209-1213.

Jeanmougin F, Thompson JD, Gouy M, Higgins DG, and Gibson TJ. 1998. Multiple sequence alignment with Clustal X. *Trends in Biochemical Sciences* 23:403-405. https://doi.org/10.1016/S0968-0004(98)01285-7

Jones KL, Glenn TC, Lacy RC, Pierce JR, Unruh N, Mirande CM, and Chavez-Ramirez F. 2002. Refining the whooping crane studbook by incorporating microsatellite DNA and leg-banding analyses. *conservation biology* 16:789-799.

Keller LF, and Waller DM. 2002. Inbreeding effects in wild populations. *Trends in Ecology & Evolution* 17:230-241.

Kitchener AC, Breitenmoser-Würsten C, Eizirik E, Gentry A, Werdelin L, Wilting A, Yamaguchi N, and Johnson WE. 2017. A revised taxonomy of the Felidae: The final report of the Cat Classification Task Force of the IUCN Cat Specialist Group. *Cat News*.

Luikart G, Biju - Duval MP, Ertugrul O, Zagdsuren Y, Maudet C, and Taberlet P. 1999. Power of 22 microsatellite markers in fluorescent multiplexes for parentage testing in goats (*Capra hircus*). *Animal Genetics* 30:431-438.

Luo S-J, Johnson WE, Martenson J, Antunes A, Martelli P, Uphyrkina O, Traylor-Holzer K, Smith JL, and O'Brien SJ. 2008. Subspecies genetic assignments of worldwide captive tigers increase conservation value of captive populations. *Current biology* 18:592-596.

Malausa T, Gilles A, Meglécz E, Blanquart H, Duthoy S, Costedoat C, Dubut V, Pech N, Castagnone-Sereno P, and Delye C. 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* 11:638-644.

Mayr W. 1995. DNA Recommendations-1994 Report Concerning Further Recommendations of the DNA Commission of the ISFH Regarding PCR-Based Polymorphisms in STR (Short Tandem Repeat) Systems. *Vox Sanguinis* 69:70-71.

Menotti-Raymond M, David VA, Lyons LA, Schäffer AA, Tomlin JF, Hutton MK, and O'Brien SJ. 1999. A genetic linkage map of microsatellites in the domestic cat (*Felis catus*). *Genomics* 57:9-23.

Nakahori Y, Takenaka O, and Nakagome Y. 1991. A human X-Y homologous region encodes "amelogenin". *Genomics* 9:264-269. https://doi.org/10.1016/0888-7543(91)90251-9

Pei J, Bao P, Chu M, Liang C, Ding X, Wang H, Wu X, Guo X, and Yan P. 2018. Evaluation of 17 microsatellite markers for parentage testing and individual identification of domestic yak (*Bos grunniens*). *PeerJ* 6:e5946.

Queller DC, Strassmann JE, and Hughes CR. 1993. Microsatellites and kinship. *Trends in Ecology & Evolution* 8:285-288.

Formatted: Font: Times Font, Italic
Formatted: Font: Times Font
Deleted: -
Formatted: Font: (Default) SimSun
Formatted: Font: Times Font
Formatted: Font: Times Font
Formatted: Font: Times Font
Deleted: -
Formatted: Font: (Default) SimSun
Formatted: Font: Times Font
Deleted: -
Formatted: Font: (Default) SimSun
Formatted: Font: Times Font
Formatted: Font: (Default) SimSun
Formatted: Font: Times Font
Formatted: Font: Times Font, Italic
Formatted: Font: Times Font
Deleted: ASTAGNONE - SERENO P
Formatted: Font: Times Font
Formatted: Font: (Default) SimSun
Formatted: Font: Times Font
Deleted: -
Formatted: Font: (Default) SimSun
Formatted: Font: Times Font
Deleted: -
Formatted: Font: (Default) SimSun
Formatted: Font: Times Font
Deleted: -
Deleted: -
Formatted: Font: (Default) SimSun
Formatted: Font: Times Font
Formatted: Font: Times Font, Italic
Formatted: Font: Times Font
Formatted: Font: Times Font
Formatted: Font: Times Font
Formatted: Font: Times Font, Italic
Formatted: Font: Times Font

Reed DH, and Frankham R. 2003. Correlation between fitness and genetic diversity. *conservation biology* 17:230-237.

Ruiz-López MJ, Gañan N, Godoy JA, Del Olmo A, Garde J, Espeso G, Vargas A, Martinez F, Roldán ER, and Gomendio M. 2012. Heterozygosity-fitness correlations and inbreeding depression in two critically endangered mammals. *conservation biology* 26:1121-1129.

Rychlik W. 2007. OLIGO 7 primer analysis software. *PCR Primer Design*: Springer, 35-59.

Sharma R, Stuckas H, Moll K, Khan I, Bhaskar R, Goyal SP, and Tiedemann R. 2008. Fourteen new di- and tetranucleotide microsatellite loci for the critically endangered Indian tiger (*Panthera tigris tigris*). *Molecular Ecology Resources* 8:1480-1482.

Vignal A, Milan D, SanCristobal M, and Eggen A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol* 34:275-305. 10.1051/gse:2002009

Webster MS, and Reichart L. 2005. Use of microsatellites for parentage and kinship analyses in animals. *Methods in Enzymology*: Elsevier, 222-238.

Williamson J, Huebinger R, Sommer J, Louis Jr E, and Barber R. 2002. Development and cross-species amplification of 18 microsatellite markers in the Sumatran tiger (*Panthera tigris sumatrae*). *Molecular Ecology Notes* 2:110-112.

Wilting A, Courtiol A, Christiansen P, Niedballa J, Scharf AK, Orlando L, Balkenhol N, Hofer H, Kramer-Schadt S, and Fickel J. 2015. Planning tiger recovery: understanding intraspecific variation for effective conservation. *Science advances* 1:e1400175.

Wu J-H, Lei Y-L, Fang S-G, and Wan Q-H. 2009. Twenty-one novel tri-and tetranucleotide microsatellite loci for the Amur tiger (*Panthera tigris altaica*). *Conservation Genetics* 10:567-570.

Wu Y, Bao W, Zhang H, Zheng W, Li W, Xu Q, and Chen G. 2011. Analysis of genetic diversity and relationship of *Panthera tigris altaica* using microsatellite markers. *Journal of Yangzhou University, Agricultural and Life Sciences Edition* 32:87-91.

Xu X, Dong G-X, Hu X-S, Miao L, Zhang X-L, Zhang D-L, Yang H-D, Zhang T-Y, Zou Z-T, and Zhang T-T. 2013. The genetic basis of white tigers. *Current biology* 23:1031-1035.

Xu Y, Fang S, and Li Z. 2007. Sustainability of the South China tiger: implications of inbreeding depression and introgression. *Conservation Genetics* 8:1199-1207.

Xu YC, Li B, Li WS, Bai SY, Jin Y, Li XP, Gu MB, Jing SY, and Zhang W. 2005. Individualization of tiger by using microsatellites. *Forensic science international* 151:45-51.

Zhang S, Bian Y, Tian H, Wang Z, Hu Z, and Li C. 2015. Development and validation of a new STR 25-plex typing system. *Forensic Science International: Genetics* 17:61-69.

Zhang W, Zhang Z, Wei K, Shen F, Hou R, Zhang L, and Yue B. 2006a. Isolation and characterization of polymorphic tri- and tetra- nucleotide microsatellite loci for the south China tiger *Panthera tigris amoyensis*. *Journal of Natural History* 40:2259-2263.

Zhang Y, Li D, Rao L, Xiao Q, and Liu D. 2003. Identification of polymorphic microsatellite DNA loci and paternity testing of Amur tigers. *Dong wu xue bao[Acta zoologica Sinica]* 49:118-123.

ZHANG ZH, ZHANG WP, YUE BS, SHEN FJ, Zhang L, Hou R, and ZHU MY. 2006b. Twelve polymorphic microsatellite loci for the South China tiger *Panthera tigris amoyensis*. *Molecular Ecology Notes* 6:24-26.