Doctoral thesis

Genetic diversity studies of endangered Grevy's zebra (*Equus grevyi*) in the captivity

Hideyuki Ito

Wildlife Research Center

Kyoto University

Kyoto, Japan

2015

TABLE OF CONTENTS

Abstract

The Grevy's zebra (*Equus grevyi*) inhabits from arid to semi-arid region in northern Kenya and Ethiopia. During few decades, the species has suffered a strict decline due to destruction of a habitat, overhunting and competition with livestock. This equid is one of the most endangered wild equids in the world, and is one of species which depends on captive breeding for survival. Therefore, an immediate conservation programme including genetic management is required to establish the stable populations. However there is little information about the genetic diversity of the Grevy's zebra until now. Furthermore, hybridization between the Grevy's zebra and the plains zebra (*Equus quagga*) has been documented, leading to a requirement for conservation genetic management within and between the species. Also, understanding the personality of individuals is important for suitable husbandry and population management. The aim of this study was to develop molecular markers and apply them to clarify genetic structures for conservation of Grevy's zebra in the wild and captivity. Also, to obtain fundamental information to clarify the correlation between genes and personality in the Grevy's zebra, I analysed the diversity of androgen receptor gene (*AR*) as a candidate gene relating with aggressiveness and tameness.

Firstly, I developed 28 microsatellite markers based on genome sequence of Grevy's zebra and evaluated genetic diversity of these markers in this species. Then I assessed cross-amplification of these markers in plains zebra and mountain zebra (*Equus zebra*). There *de novo* microsatellite markers showed sufficient variation to enable individual identification in all populations. Comparative diversity estimates indicated greater genetic variation in plains zebra and its subspecies than in Grevy's zebra, despite potential ascertainment bias. Species and subspecies differentiation were clearly demonstrated, and

simulated F1 and F2 hybrids were correctly identified. Also, in the analysis of two regions (cytochrome b and control region) of mtDNA, Grevy's zebra showed extremely lower diversity than other two zebras. On the other hand, in genetic diversity based on traditional pedigree analysis, Grevy's zebra had higher diversity than mountain zebra. These results suggested that founders of captive population had low genetic diversity. The inconsistency of results by two analysis methods indicated that adopting molecular methods is important in captive population management of wildlife.

In *AR*, variation in number of repeats in the sequence coding glutamine was observed among zebra species and between zebra and horse. Until now, polymorphisms in horse have not been reported, while three zebra species had longer allele than horse (5 times in horse, 8-15 times in zebras). This result might indicate the possibility that *AR* have been selected in the domestication of horse.

This study indicates the genetic structures and diversity of *AR* in endangered species Grevy's zebra for the first time. The genetic diversity in captive Grevy's zebra was lower than that calculated from traditional pedigree methods, indicating that it is essential to develop captive management plan considering the results of molecular methods for conservation of Grevy's zebra. Novel molecular markers developed in this study are useful for conservation of Grevy's zebra, such as developing of more appropriate management plan, and clarifying genetic structures of wild populations and social systems. Furthermore, this study showed the possibility that *AR* gene had been selected as traits related to domestication. These findings will be fundamental information to conduct further studies focusing on the potential association between *AR* and personality traits, and to develop appropriate husbandry and population management based on individual personality traits.

Chapter 1 General introduction

Zebras belong to the taxonomic family of horses (Equidae), which is comprised of a single genus, *Equus*. Zebras are native to Africa and are characterized by their distinctive black and white striped coat of fur. They occur in a variety of habitats, including grasslands, savannas, woodlands, thorny scrublands, mountains, and coastal hills (Hack and Lorenzen, 2008; Moehlman et al., 2013; Novellie, 2008). There are three species of zebra (Fig. 1.1): the plains zebra (*Equus quagga*), the mountain zebra (*E. zebra*) and the Grevy's zebra (*E. grevyi*). The plains zebra includes five extant subspecies: Burchell's zebra *E. q. burchelli*, Grant's zebra *E. q. boehmi*, Selous' zebra *E. q. borensis*, Chapman's zebra *E. q. chapmani*, and Crawshay's zebra *E. q. crawshayi* (Hack and Lorenzen, 2008). The mountain zebra has two subspecies: Cape mountain zebra *E. z. zebra* and Hartmann's mountain zebra *E. z. hartmannae* (Novellie, 2008).

Grevy's zebra is the largest wild equid. Their unique characteristics, the stripes that are evenly divided and does not cover the belly, they have big and round ears and a brown spot on the nose. Grevy's zebras live in arid and semi-arid regions. Breeding males defend large territories of 2–12 km²; the home range size of non-territorial individuals is up to 10,000 km² (Moehlman et al., 2013). Kenya Wildlife Service (KWS) (2012) document that 'Grevy's zebra have undergone one of the most substantial reductions of range of any African mammal'. Historically the species was found in the Horn of Africa including Kenya, Ethiopia, Eritrea, Djibouti, Somalia and Sudan. However, the species is found only in Kenya and Ethiopia today. Numbers of Grevy's zebra have declined from an estimate of 15,000 in the late 1970s to the present-day estimate of 2,800 animals representing an 81% decline in

global numbers (KWS, 2012), and the Grevy's zebra is listed on CITES Appendix I. The factors which caused the decline in Grevy's zebra are reduction of available water sources, habitat degradation and loss due to overgrazing, competition for resources, hunting and disease (Moehlman et al., 2013). Hybridizations (between Grevy's zebra and plains zebra) and small population size are listing as threats to conservation currently (KWS, 2012). The rate of inbred mating, gene drop and genetic drift is high in small populations. These events cause the accumulation of deleterious trait and the loss of genetic diversity, and lead to be higher risk of extinction (Fagan and Holmes, 2006; Frankham et al., 2010). Therefore, the conservation plan including genetic management is essential, and genetic markers that can evaluate genetic diversity and identify hybridized individuals are required.

Captive breeding has become an important tool in species conservation programmes. Now, for many species which need captive management plan, the Zoological Society of London regularly publishes the updated list of current international studbooks in the International Zoo Yearbook. Currently, there are 131 active international studbooks, including 162 species or sub-species. In Grevy's zebra at the end of 2013, 508 individuals were recorded in 112 institutions and some private collections in 28 countries (Langenhorst, 2014a). Of those, 247 individuals are kept in 51 institutions under the formally managed European Endangered Species Programmes (EEP) in European Association of Zoos and Aquaria (EAZA), while the North American Species Survival Plan (SSP) Program coordinates 179 animals in 38 institutions in the Association of Zoos and Aquariums (AZA). A third population is managed by the Japanese Association of Zoos and Aquariums (JAZA) and consists of 19 animals in eight zoos. Captive population account for about 15% of total population, and is very important for species conservation. Witzenberger and Hochkirch

(2011) indicated that inbreeding can be minimized by a thorough management of captive populations. Therefore, thorough management is essential in captive Grevy's populations, especially in JAZA population.

Recently, traditional pedigree analysis is incorporated with molecular genetic analysis in some captive breeding programme (Ferrie et al., 2013; Gautschi et al., 2003; Henkel et al., 2012; Ivy et al., 2009; McGreevy et al., 2011; Ogden et al., 2007). These combined analyses can allow us to compare genetic diversity expected by pedigree analysis with that of genetic variation observed in the molecular data, and to develop more appropriate genetic management. Molecular markers are indispensable tools for determining the genetic variation and biodiversity with high levels of accuracy and reproducibility. Microsatellite markers, that are one of molecular markers, are a class of short nucleotides (2 – 6 bp) that are repeated in tandem. Because microsatellite markers are codominant and usually ubiquitous in the genome, and have high rate of mutation, these markers have been used to study parentage and kinship, genetic diversity of various species. As these genetic informations are important to conserve the wildlife, microsatellite markers are one of the most popular genetic markers for conservation genetic studies (Guichoux et al., 2011). Although cross-species applicability is possible in some closely related taxa, they may not be very efficient in some instances (Inoue-Murayama et al., 2001), and microsatellite markers have to be developed *de novo* in many species. In equid, 24,475 microsatellite markers have been registered in the DDBJ (http://www.ddbj.nig.ac.jp/, accessed at 28th October 2015) in the horse which is the domesticated animals. However, the microsatellite markers have not been developed in the other equids until now, there are only reports used microsatellite markers that were developed in horse (Lorenzen et al., 2008; Moodley and Harley, 2006). There are no reports

about genetic diversity of Grevy's zebra populations using microsatellite markers. Also, mitochondrial DNA (mtDNA) is a useful marker for population genetics studies, with some characteristics such as maternal-inheritance and non-recombinant with a high mutation rate (Avise, 1987; Chen and Hebert, 1999; Larizza et al., 2002). Furthermore, because of easier amplification due to the presence of multiple copies within a cell, mtDNA has been used extensively to study populations and to trace maternal lineages. The facility of PCR amplification is major advantage in the research for the endangered species (not only in the captive but also in the wild), because many of samples from these species is collected by non-invasive method. Although mtDNA have been widely used the research of evolution and relation among breeds/species in the equid, the studies of Grevy's zebra are only a few; there is one recent report about genetic diversity of the wild Grevy's zebra (Kebede et al., 2014) but no report in the captivity. Therefore, these genetic diversity measures of populations may serve as useful information for conservation.

Animals in the captivity are kept and cared by human, so understanding personality of each individual can enable better routine management of them. Also, understanding of personality is important to manage not only individuals but also population. In cheetah (*Acinoynx jubtus*) and black rhino (*Diceros bicornis*), the relationship between personality (tence-fearful, dominant) and reproductive success were reported (Carlstead et al., 1999; Wielebnowski, 1999). Kuhar et al. (2006) suggested personality is useful to develop the bachelor-group in western lowland gorilla (*Gorilla gorilla gorilla*). Personality has affected by both environment and genetic factors. Around 20 - 50 % of the phenotypic variation in animal personality traits has genetic basis (van Oers and Mueller, 2010). Equidae are comprised of eight species, which all belong to a single genus, *Equus* (Fig. 1.1). *Equus*

contains four sub-genera, *Equus*, *Asinus*, *Dolichohippus*, and *Hippotigris*. Sub-genus *Equus* includes domestic horse and wild horse. Sub-genus *Asinus* includes donkey, African wild ass, Asian wild ass and kiang. *Dolichohippus* includes only Grevy's zebra. *Hippotigris* includes plains zebra and mountain zebra. Together with other factors such as ecology and phylogeny, social structure affects or reflects species difference in personality, such as levels of aggression, affiliation, and pair-bonding. Similarly, social structures could also be reflected in genetic differences. Members of the genus *Equus* show two patterns of social organization (Rubenstein, 2011). In the sub-genus *Equus* and *Hippotigris*, a single breeding male is found in constant association with a fixed group of 1-6 unrelated females and their offspring; this is the forming of a harem. On the other hand, in the sub-genus *Asinus* and *Dolichohippus*, breeding males have territories, and females do not form stable groups. The diversity of these social systems is not consistent with genetic phylogeny. In addition, humans succeeded only in the domestication of the horse in sub-genus *Equus* and donkey in sub-genus *Asinus*, which display different social systems. The other species, including wild asses and zebras, have been considered not suitable for domestication because of their untameable nature. As genus *Equus* displays different social systems, it is a suitable system to investigate the function of personality- and sociality-related genes in different sub-genera. Because androgen receptor gene (*AR*) variation may influence aggressiveness along with other traits, they are among the ideal candidate genes to study the basis for social system and behavioural differences in equids, which in turn can help us to understand the functions and evolution of these genes.

In this study, I aimed to develop microsatellite markers that can be useful for the characterization of genetic diversity of endangered Grevy's zebra in order to manage captive

populations and to conserve wild populations. As far as I know, this study is the first attempt to develop molecular markers based on wild equid species, not only Grevy's zebra. This study is organized into two parts. In chapter 2 of this thesis, I developed novel microsatellite markers in the Grevy's zebra using next-generation sequencing, and considered usefulness of developed markers. Also, I assessed the utility of novel markers for the conservation of three zebra species (identification individuals, species and subspecies, and detection hybrid individuals), and determined the genetic diversity of the two captive Grevy's zebra populations using mitochondrial DNA and developed microsatellite markers. In chapter 3, I investigated the polymorphisms within *AR*, to clarify genetic information about *AR* gene in three zebra. These findings will be useful for basic understanding of the relationship between *AR* genotypes and aggressiveness/tameness in zebra species. Moreover it will be helpful to comprehend the personality of individuals, and to develop suitable keeping ways for individuals. As a first step toward this approach, I analysed the repeat of glutamine in *AR* in three zebra species for comparison with those of horses. In chapter 4, I discuss usefulness of novel genetic markers in this study for the conservation and indicate future perspective. This information will provide a foundation for developing genetic conservation programmes of Grevy's zebra in captivity, and application tool to wild population. Moreover, this study will provide useful information to the captive breeding plan in other species.

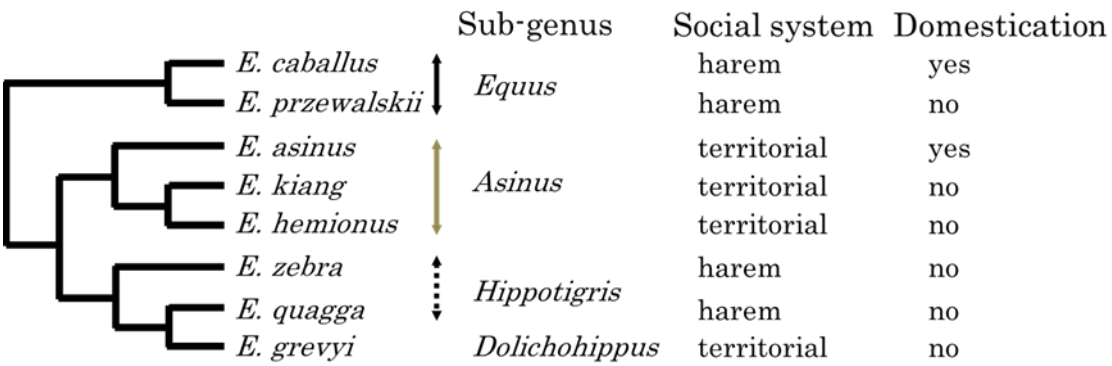


Fig. 1.1 Phylogeny and sociality in equids. Phylogenetic tree follows Steiner and Ryder (2011), based on mtDNA and nuclear DNA sequences.

Chapter 2 Evaluation of genetic diversity with DNA markers in Grevy's zebra

2.1 Using next generation sequencing to develop microsatellite markers for the endangered Grevy's zebra

2.1.1 Introduction

Today the Grevy's zebra distribution ranges from arid to semi-arid areas of Ethiopia and northern Kenya. Since the seventies the species has suffered a sharp decline due to destruction of habitat, overhunting and competition with livestock. The most recent survey counted about 2,500 and 280 individuals in Kenya and Ethiopia, respectively (KWS, 2012), making this equid one of the most endangered mammals in the world; classified in the CITES Appendix I and 'Endangered' in the IUCN's red list (Moehlman et al., 2013). A conservation programme is in action on the ground but so far, this does not include any genetic management. In captivity, 508 individuals were recorded in the international studbook by the end of 2013 (Langenhorst, 2014a). While some of these animals are genetically managed in organized breeding programmes, some of the pedigree information in the studbook is ambiguous. The genetic information gained from molecular-biological study is expected to complement the studbook data. In this study I developed microsatellite markers using next-generation sequencing to identify the genetic diversity of Grevy's zebra which will help conservation management plans in the future.

2.1.2 Materials and Methods

DNA was extracted from blood of male Grevy's zebra which was kept in Japan using

QIAGEN DNeasy Blood and Tissue Kit (Qiagen, Germany). After checking the quality of genomic DNA by resolution on a 0.5% agarose gel and spectrophotometry (Nanodrop, USA), 500ng of the genomic DNA was nebulized at 0.24 MPa for 1 min, and purified using the MinElute PCR Purification kit (Qiagen). The fragments were end-repaired, A-tailed and ligated to the Rapid Library Adapter with RL Ligase (Roche, Switzerland). Short fragments were removed using AMPure XP beads, and the quality and quantity of the library were assessed using Agilent 2100 Bioanalyser (Agilent, USA). Library fragments were mixed with capture beads and clonally amplified through emulsion PCR using the GS-Junior Titanium emPCR kit (Roche). Captured fragments were enriched and annealed with sequencing primers and sequenced using GS-Junior bench-top sequencer (Roche). I obtained 92,254 reads, and the reads containing microsatellite were screened by MSATCOMMANDER ver.0.0.2 (Faircloth, 2008). Repeats including 2-6 nucleotides repeat were searched for with the following settings: more than seven di-repeats and more than four repeats for the other repeat types. Among reads containing microsatellite, sixty-six primers were designed using the PRIMER3 (Rozen and Skaletsky, 2000). And these primer pairs were tested for amplification and polymorphism in 41 samples (17 samples in Japan, blood n = 1, muscles n = 4, hairs n = 5, feces n = 4; 24 samples in the United Kingdom, feces n = 27).

PCR amplifications were performed by modified protocol of the Qiagen Multiplex PCR Kit (Qiagen) in a final volume of 10 µl, which contained 20 ng of DNA, 5 µl Multiplex PCR Master Mix, 400 µM of each dNTP, 0.4 µM of forward (fluorescently labeled) and reverse primers. In fecal samples, instead of 20 ng DNA, 2 µl of extracted DNA solution and 0.1 µg of T4 Gene 32 Protein (Nippon Gene, Japan) were added. The PCR conditions were as

follows: 95 °C for 15 min; 35 cycles at 94 °C for 30 s, 60 °C for 90 s, and 72 °C for 60 s with a final extension at 60 °C for 30 min. In fecal samples, the PCR conditions were as follows: 95 °C for 15 min; 15 cycles at 94 °C for 30 s, 57 °C for 90 s, and 72 °C for 60 s; then 30 cycles at 94 °C for 30 s, 52 °C for 90 s, and 72 °C for 60 s with a final extension at 60 °C for 30 min. The size of the PCR products was measured using the ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems) and GENEMAPPER ver.4.0 (Applied Biosystems). And the number of alleles (*Na*), expected heterozygosity (*He*), observed heterozygosity (*Ho*), probability of identity (*PID*) and *PID* among siblings (*PID-sib*) were calculated using Genalex ver. 6.41 (Peakall and Smouse, 2012). Deviation from Hardy–Weinberg Equilibrium (HWE) and linkage disequilibrium were tested for using GENEPOP ver.4.0.10 (Rousset, 2008) after Bonferroni correction.

2.1.3 Results and discussion

Sequence characteristics and repeat structure

A total of 92,254 sequence reads were generated from next generation sequencing (NGS) with an average read length of 355.6 bp and total bases were 32.8 Mb. The numbers of read including 2 - 6 tandem repeats were 2,725 reads (3.0%). Repeat motif and repeat numbers were indicated in Fig. 2.1.1. In microsatellite markers that were distributed widely on the genome, there are differences of repeat numbers and repeat motifs among species (Abdelkrim et al., 2009; Dieringer and Schlotterer, 2003; Santana et al., 2009). In class of repeat unit, di-, tri- and tetra-repeats was abundant, penta- and hexa-repeats is extremely low (Table 2.1.1). Among the di-repeats, AC/GT (792 reads, 65.5%) was the most frequent whereas GC was the least frequent (one read, 0.08%). AAT/ATT, AAAT/ATTT motifs were

most abundant for tri- and tetra-repeats, respectively. Across many vertebrate taxa, AC and AAT were found to be the most frequent for di-repeat and tri-repeat, respectively (Meglecz et al., 2012; Toth et al., 2000). These are also true for the Grevy's zebra according to my results. The accumulation of these data from various species leads to the effective development of microsatellite markers.

Development of markers

21 of tested 66 loci show polymorphic. The profile of 21 loci in 41 individuals shows Table 2.1.2. All loci showed no significant deviations from HWE and linkage disequilibrium was not detected between pairs of loci. The *Na*, *Ho* and *He* ranged from 2 to 9 (mean 4.57), 0.075-0.854 (mean 0.523), 0.072-0.784 (mean 0.529), respectively. Cumulative *PID* and *PID-sib* for all loci were 1.78×10^{-13} and 2.78×10^{-6}, respectively. This *PID-sib* value means distinguishable 359,712 individuals for the purpose of calculation and show that these markers are able to identify all individuals in wild and captivity.

I developed the first microsatellite markers for Grevy's zebra derived from zebra genome information. These markers will be useful for the conservation genetic study in wild and captive populations, e.g. for parentage analysis, individual identification, and hybrid identification.

Table 2.1.1. Number of reads contained microsatellite

Number of read	Motif (repeats)	Number of reads contained microsatellite	*Number of potentially amplifiable reads (%)
92254	Hexa (4≤)	11	2 (18.2)
	Penta (4≤)	55	33 (66.0)
	Tetra (4≤)	517	263 (50.9)
	Tri (4≤)	932	516 (55.4)
	Di (7≤)	1210	360 (29.8)
	total	2725	1174 (43.1)

*: The number of reads that from which primers were designed by PRIMER3

Table 2.1.2 Profile of 21 microsatellite markers

Locus	Repeat unit	*n*	*Na*	*He*	*Ho*	*PID*	*PID-sib*	Hardy-Weinberg Equilibrium (*p*-value)
EGR01	$(CAG)_5$	40	2	0.250	0.255	0.588	0.769	0.901
EGR02	$(AAT)_{11}$	40	4	0.450	0.527	0.328	0.569	0.587
EGR03	$(AGG)_8$	39	3	0.385	0.512	0.357	0.583	0.295
EGR05	$(CTT)_9$	41	4	0.610	0.609	0.234	0.504	0.447
EGR07	$(GGT)_8$	41	3	0.585	0.534	0.326	0.565	0.261
EGR08	$(AGAT)_{10}$	39	6	0.615	0.709	0.132	0.428	0.177
EGR09	$(AGAT)_{11}$	40	4	0.650	0.598	0.226	0.508	0.332
EGR10	$(AGGC)_8$	40	5	0.300	0.267	0.554	0.755	1.000
EGR11	$(ATCT)_{11}$	41	4	0.634	0.613	0.198	0.493	0.778
EGR12	$(ATCT)_9$	40	6	0.750	0.769	0.090	0.388	0.474
EGR13	$(ATCT)_9$	40	4	0.500	0.609	0.226	0.502	0.526
EGR14	$(CA)_{13}$	41	4	0.585	0.585	0.233	0.516	0.911
EGR15	$(CA)_{14}$	40	7	0.850	0.784	0.079	0.378	0.495
EGR16	$(CA)_{18}$	41	8	0.683	0.657	0.151	0.459	1.000
EGR18	$(CAG)_4$	40	4	0.175	0.207	0.637	0.806	0.159
EGR21	$(CAG)_4$	40	4	0.075	0.073	0.860	0.928	1.000
EGR24	$(AAC)_{10}$	40	2	0.075	0.072	0.863	0.930	0.805
EGR30	$(CA)_{13}$	41	9	0.854	0.741	0.091	0.402	0.791
EGR31	$(CA)_{14}$	37	5	0.568	0.668	0.155	0.455	0.261
EGR32	$(CA)_{15}$	40	5	0.700	0.714	0.129	0.425	0.732
EGR33	$(CA)_{16}$	39	8	0.718	0.638	0.175	0.475	0.978

Abbreviations: *n*, number of genotyped individuals *Na*, observed no. of alleles; *Ho*, observed heterozygosity; *He* expected heterozygosity; *PID*, probability of identity (unrelated); *PID-sib*, probability of identity of siblings.

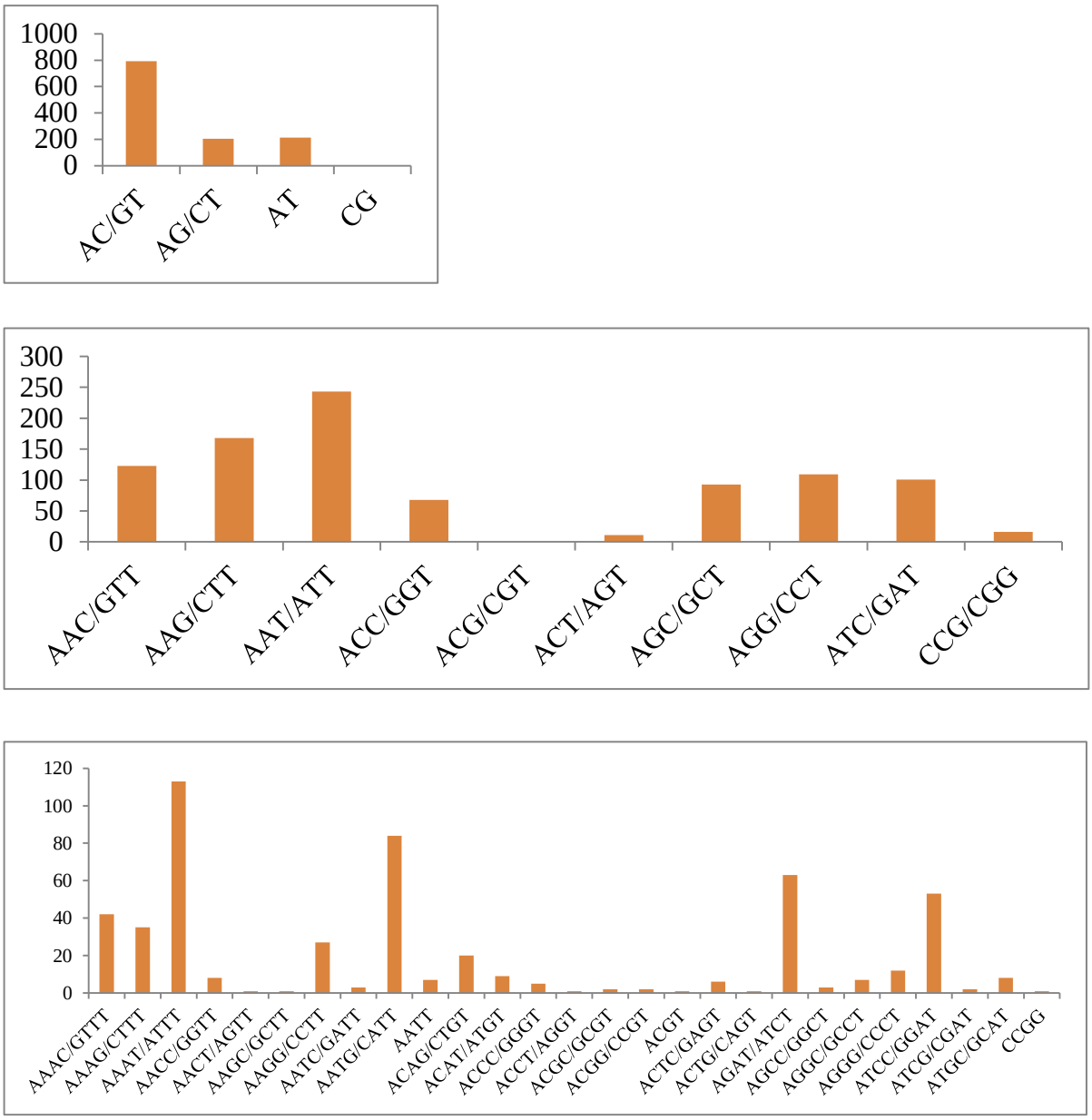


Fig. 2.1.1 The number of reads of different repeat types of the Grevy's zebra based on partial genome data.

2.2 Population genetic diversity and hybrid detection in captive zebras

2.2.1 Introduction

Zebras belong to the taxonomic family of horses (Equidae), which is comprised of a single genus, *Equus*. There are three species of zebra: the plains zebra, the mountain zebra and the Grevy's zebra (see chapter 1). While plains zebras are plentiful, various anthropogenic factors (over hunting, competition with livestock, habitat loss, etc.) have had a severe impact on Grevy's zebra and mountain zebra populations, which are now listed as Endangered or Vulnerable respectively on the IUCN red list (Moehlman et al., 2013; Novellie, 2008). There is relatively little information on the genetic diversity of these species to support conservation management in the wild or in captivity; the development and application of molecular genetic tools is therefore an important consideration.

Analysis of genetic structure using microsatellite markers and mtDNA has been reported in mountain zebra (Moodley and Harley, 2006) and plains zebra (Lorenzen et al., 2008). However, in Grevy's zebra, although a limited number of microsatellite markers have been identified (Ogden et al., 2007), there has been no detailed analysis of genetic structure with nuclear genetic markers. Furthermore, hybridization between the Grevy's zebra and the Grant's zebra, a sub-species of plains zebra has been reported in the wild, resulting in fertile F1 individuals (Cordingley et al., 2009). Hybridization may be a potential risk factor in the conservation of Grevy's zebra and it is therefore essential to be able to genetically identify these admixed individuals and to investigate how these hybridized individuals affect Grevy's zebra populations.

Previously, I developed 21 microsatellite markers in the Grevy's zebra using next-generation sequencing (Chapter 2.1) (Ito et al., 2013). The aim of this study was to

investigate the utility of an expanded Grevy's zebra microsatellite marker panel to identify subspecies, detect hybrids and assess population genetic diversity in all three zebra species. To achieve this I have characterized an additional seven novel microsatellite markers, evaluated cross-species amplification in the plains and mountain zebras and conducted a series of simulation studies to determine the power of these genetic tools to detect hybridization in captive zebra populations.

2.2.2 Materials and Methods

Samples

This study was conducted in strict accordance with the guidelines for the ethics of animal research by the Wildlife Research Center of Kyoto University. The sampling and methods were approved by each zoo providing samples and the Wildlife Research Center of Kyoto University. I obtained 59 samples from Grevy's zebra in Japan (blood, n = 1; muscle, n = 4; hairs, n = 6; and feces, n = 11) and in the United Kingdom (feces, n = 37). I collected 33 samples from grant's zebra in Japan (hair, n = 1; feces, n = 32), 18 samples from Chapman's zebra in Japan (blood, n =1; hair, n = 6; feces, n = 8) and in the United Kingdom (feces, n = 3). I collected 10 feces from Hartmann's mountain zebra in Japan (n = 8) and in the United Kingdom (n = 2). All individuals were kept in zoos in Japan or the United Kingdom. Invasive sampling was minimized, with blood samples collected as bi-products during health examination and hairs samples (ca. 10 hairs) plucked by keepers and muscle samples obtained *post mortem*. DNA was extracted from whole blood, muscle and hair using the QIAGEN DNeasy Blood and Tissue Kit (Qiagen), and from faeces using the QIAGEN DNeasy Stool Kit (Qiagen).

Development of microsatellite markers

In addition to the 21 Grevy's zebra microsatellite markers previously published (Chapter 2.1) (Ito et al., 2013), a further seven markers were used in this study, developed at the same time using the same method. The methods of development of microsatellite markers refer to the chapter 2.1. I selected novel seven microsatellite markers (four polymorphic markers and three monomorphic markers) in this study. All 28 markers were tested for cross-species amplification in plains and mountain zebra and the additional seven markers also examined for the first time in Grevy's zebra.

PCR amplifications were performed by a method same as Chapter 2.1. PCR and genotyping were replicated 3 to 9 times depending on the genotype observed. The samples that failed to genotype at more than 5 loci were subsequently excluded from the study. Allelic richness (*Ar*) per locus were calculated using HP-Rare ver.1.1 (Kalinowski, 2005). The number of alleles (*Na*), expected heterozygosity (*He*), observed heterozygosity (*Ho*), probability of identity (*PID*) and *PID* among siblings (*PID-sib*) were calculated for each species using GenALEx ver.6.41 (Peakall and Smouse, 2012). Deviation from Hardy–Weinberg Equilibrium (*HWE*) and linkage disequilibrium were tested for using GenALEx ver.6.41 (Peakall and Smouse, 2012) and GENEPOP ver.4.0.10 (Rousset, 2008) after Bonferroni correction, respectively.

Genetic structure analysis

Microsatellite data were also analyzed using the programme STRUCTURE ver.2.3.3 (Pritchard et al., 2000) using the admixture model to estimate population genetic structure

and individual ancestries, among and within species. I conducted an analysis with 10 iterations for each population size (*K*) of 1 to 8, and with Markov chain Monte Carlo (MCMC) running for 500,000 generations and initial burn-in of 250,000 generations. The *K* values described by Evanno *et al*. (2005) were then calculated to identify the most reasonable *K* using the programme Structure Harvester (Earl and vonHoldt, 2012). Runs were averaged using CLUMPP ver.1.1.2 (Jakobsson and Rosenberg, 2007), and results were visualized using DISTRUCT ver.1.1 (Rosenberg, 2004). I assessed the average coefficient of membership (*Qi*) of each sampled population to the inferred clusters. Then, I assessed each genotyped to the inferred clusters, based on threshold values of the individual proportion of membership (*qi*). Moreover, the pattern of allelic differentiation between species was explored through Principal Coordinate Analysis (PCoA) by GenALEx ver.6.41 (Peakall and Smouse, 2012) based on calculated genetic distances.

To test the ability of the markers to identify subspecies (Grant's zebra and Chapman's zebra) in the plains zebra, STRUCTURE analysis and PCoA were conducted in the same way using microsatellite data of plains zebra. To validate the utility of the markers to detect hybridization between Grevy's zebra and Grant's zebra (the partially sympatric subspecies), I simulated four hybrid populations (Grevy's zebra x Grant's zebra = F1; F1 x F1 = F2; F1 x Grevy's zebra backcross = BxGy, F1 x Grant's zebra = BxGt). For the F1 population 20 hybrid individuals were simulated using allele frequencies from Grevy's zebra and Grant's zebra with the software HYBRIDLAB ver.1.0 (Nielsen et al., 2006). F2 and the two backcross populations (n = 20) were developed from the two pure species and the simulated F1 population. I performed STRUCTURE analysis and PCoA as described above using all six populations (Grevy's zebra, Grant's zebra, and the four simulated hybrid

populations).

2.2.3 Results

I excluded 38 of 101 faecal samples that had failed to genotype at more than five loci. The numbers of samples used subsequent analysis in Grevy's zebra, plains zebra (Grant's zebra, Chapman's zebra) and Hartmann's mountain zebra were 52, 27 (15, 12) and 6 respectively.

Validation of utility for cross-species amplification

The results of cross-species amplification for all 28 loci are shown Table 2.2.1 (for details, see Supplementary Table 2.2.1). For the estimation of population genetic indices, the plains zebra was divided into Grant's and Chapman's subspecies and analyzed separately. The data for mountain zebra are presented here for information, however the sample size is too small to allow accurate estimates to be obtained. Seven loci x species (sub-species) combinations showed deviations from Hardy–Weinberg Equilibrium after sequential Bonferroni correction (three loci in Grant's zebra, two loci in Grant's zebra, and two loci Chapman's zebra), but no single locus showed deviation in more than one species. *Ar*, *Na*, *Ho* and *He* in combined all loci are indicated in Table 2.2.1 (The indices in each locus are indicated in Supplementary Table 2.2.1). Polymorphism is generally conserved across species, with 25 polymorphic markers in Grevy's zebra (the discovery species) compared to 27 markers in plains zebra (Grant's zebra $n = 27$; Chapman's zebra $n = 26$) and 23 markers in mountain zebra. A greater number of private alleles was observed in plains zebra ($n = 63$) than in Grevy's zebra ($n = 37$) or mountain zebra ($n = 19$) and allelic richness and mean heterozygosity were higher in plains zebra (and its individual subspecies) than Grevy's zebra

(Table 2.2.1).

Cumulative *PID* and *PID-sib* across all loci ranged from 2.28 x 10^{-14} to 7.63 x 10^{-21} and 1.06 x 10^{-6} to 5.67 x 10^{-9} respectively, supporting the use of the marker panels for individual identification in all populations.

Differentiation of species and sub-species

The clustering of three zebra species observed in the STRUCTURE analysis (Fig. 2.2.1a) and principal component analysis (Fig. 2.2.1b) demonstrates clear separation of the three species using microsatellite data. In Structure Harvester, greatest support was found for K=2 clusters, resulting in the separation of Grevy's zebra from the other two zebra species (Fig. 2.2.1a); at K = 3 (Fig. 2.2.1a), plains zebra and mountain zebra were subsequently differentiated unambiguously. The confidence in individual assignment was high, with the Grevy's zebra assigned to cluster I with average proportion of membership Q_I = 99.6%, plains zebra assigned to cluster II with average proportion of membership Q_{II} = 99.4%, and mountains zebra assigned to cluster III with average proportion of membership Q_{III} = 99.5%. Additionally, PCoA separated the three species clearly. Percentages of variation explained by the first 2 axes were 22.4% and 7.0%, respectively.

The results of STRUCTURE analysis and PCoA in the two sub-species of plains zebra show separation of the Grant's zebra from Chapman's zebra (Fig. 2.2.2). At K = 2, average proportion of cluster membership in both Grant's zebra Q_I = 98.2%, and Chapman's zebra Q_{II} = 99.1% was high.

Detection of hybridized individuals

The results of STRUCTURE analysis and PCoA in the six populations comprised of two pure species (Grevy's zebra and Grant's zebra) and four hybridized populations (Grevy's zebra x Grant's zebra = F1; F1 x F1 = F2; F1 x Grevy's zebra backcross = BxGy, F1 x Grant's zebra = BxGt) indicate a sharp distinction between the pure species and two of the hybridized populations (F1 and F2 populations). As expected, the backcrossed populations were less clearly differentiated, with the F1 x Grant's zebra backcross partly overlapping with Grant's zebra (identification rates, 0.8; four BxGt individuals were misidentified as Grant's zebra) (Fig. 2.2.3). STRUCTURE results for assignment to Grevy's zebra for actual and simulated individuals were as follows: For Grevy's zebra *Qi* (population average) and *qi* (individual range) scores were: Grevy's zebra 0.986 (0.954 – 0.991); Grant's zebra, 0.016 (0.009 – 0.024); simulated F1s, 0.404 (0.347 – 0.531); F1 x Grevy's zebra backcrosses, 0.681 (0.493 – 0.826); F1 x Grant's zebra backcrosses, 0.125 (0.034 – 0.275); and in simulated F2s, 0.406 (0.243 – 0.539).

2.2.4 Discussion

This study has demonstrated the utility of a 28 marker microsatellite panel for assessing subspecies identity and hybridization in zebras, including endangered species where hybridization is recognized as a potential conservation issue. All markers successfully were cross-amplified from Grevy's zebra to plains and mountain zebra with high levels of conserved polymorphism. While direct comparison of population genetic diversity among subspecies is potentially affected by ascertainment bias, given that the markers were isolated in Grevy's zebra, it is interesting to note that indices of population genetic diversity for plains zebra (and its two subspecies) are higher (also in the case of removing monomorphic

three loci in Grevy's zebra; *EGR25, EGR28, EGR29*), suggesting comparatively less genetic diversity exists in the captive Grevy's zebra population. No comparative data exists for captive population founder size in these two species, but as the number of wild plains zebra (~600,000) far exceeds that of Grevy's zebra (~2,500) (Hack and Lorenzen, 2008; Moehlman et al., 2013), this result is perhaps not surprising, and indicates that an assessment of the wild Grevy's zebra population would be useful to evaluate the proportion of species genetic diversity that is represented in the captive conservation breeding programme.

The power of the microsatellite panels for individual DNA profiling and differentiation of siblings indicated that these markers should be suitable for the identification of all individuals in captivity and the wild. Overall the results demonstrate that these novel microsatellite markers are useful for the analysis of genetic diversity and identification of individuals in three zebra species, although the numbers of mountain zebra should be increased to strengthen these findings.

Species identification using STRUCTURE was accurate and unambiguous, with PCoA analysis also showing clear differentiation of species clusters. Between subspecies of plains zebra, all individuals were also assigned correctly but with a lower level of confidence. However the proportion of membership (*qi*) of all individuals was higher than 0.8 (minimum 0.83) used as a target for admixed individuals (Khosravi et al., 2013; Ramadan, 2012; Randi, 2008). This result demonstrates that captive Grant's zebra and Chapman's zebra can be readily separated using DNA markers. While allopatric distribution of these two sub-species minimizes the risk of hybridization in the wild, they are commonly kept in captivity and not managed by studbook, so there is a risk that these sub-species might have hybridized in the past. The lack of studbook data for Grant's zebra and Chapman's zebra also means that I

cannot quantify the proportion of captive subspecies diversity sampled in this study. As samples were derived from relatively few zoos, further sampling would be recommended to increase confidence in these findings.

The index of admixture in Bayesian analysis has been frequently used to identify individual introgression (Barilani et al., 2006; Khosravi et al., 2013; Sanz et al., 2008; Scandura et al., 2009), and has been used in the detection of hybridized individuals by defining threshold values of *qi* (Randi, 2008). In order to increase the exclusion of potential hybrids, some studies have used stringent *qi* values (> 0.95) to define non-hybridized individuals (Barilani et al., 2007; Negri et al., 2012) and I adopted this high *qi* threshold value (> 0.95) to confidently identify pure individuals in this study. Assignment of simulated hybrid individuals to their hybrid category (F1, F2, BxGy, BxGt) deviated from expectations (0.5, 0.5, 0.75 and 0.25, respectively) with lower average assignment to the Grevy's zebra genetic lineage. This deviation may be due to the fact that a greater number of species-specific alleles were observed in Grant's zebra than Grevy's zebra, skewing the assignment of simulated individuals.

Pure species (Grevy's zebra and Grant's zebra) could be clearly separated from the simulated F1 and F2 populations. Cordingley *et al.* (2009) reported observing F1 individuals derived from natural matings between male Grevy's zebra and female Grant's zebra in Kenya, with the F1 individuals subsequently staying in the Grant's zebra group. Although effect of hybridized individuals on the population of Grevy's zebra is still unclear, it is important to discriminate between Grevy's zebra and hybridized individuals for the conservation of the pure population of Grevy's zebra. This study demonstrates that microsatellite markers can distinguish Grevy's zebra from hybridized individuals in captive

populations. In addition to their usefulness in captive population genetic management in all three species, these markers are expected to distinguish hybrids between these taxa in wild populations. However as the population allele frequencies generated within this study are unlikely to accurately reflect the situation in the wild due to founder effects and drift, further analysis using wild population samples from potential hybrid zones is strongly recommended.

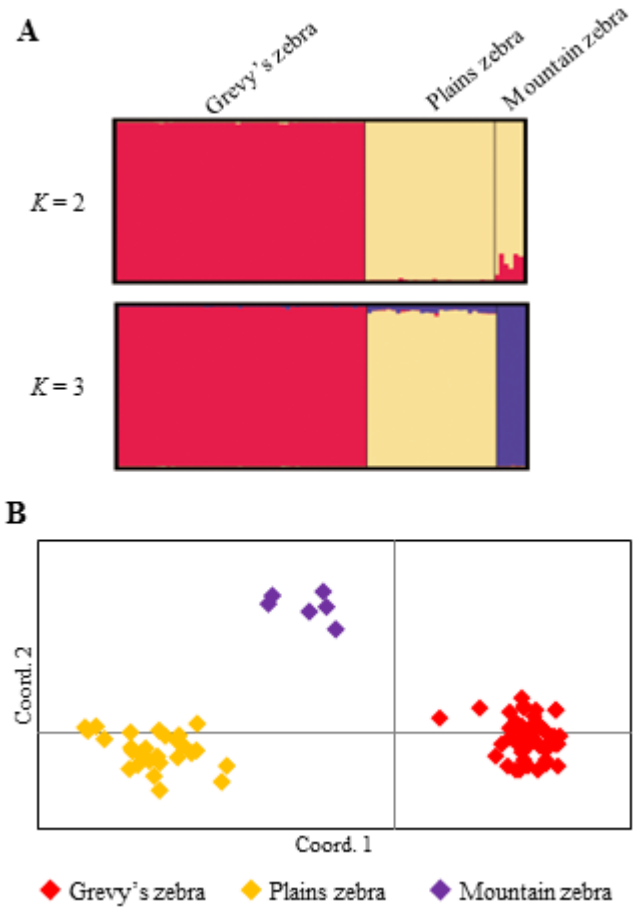


Fig. 2.2.1. (a) Bayesian analysis of the genetic structure showing differentiation of three zebra species based on 28 microsatellite loci. (b) First and second components of a principal coordinate analysis of 28 microsatellite loci in three zebra species.

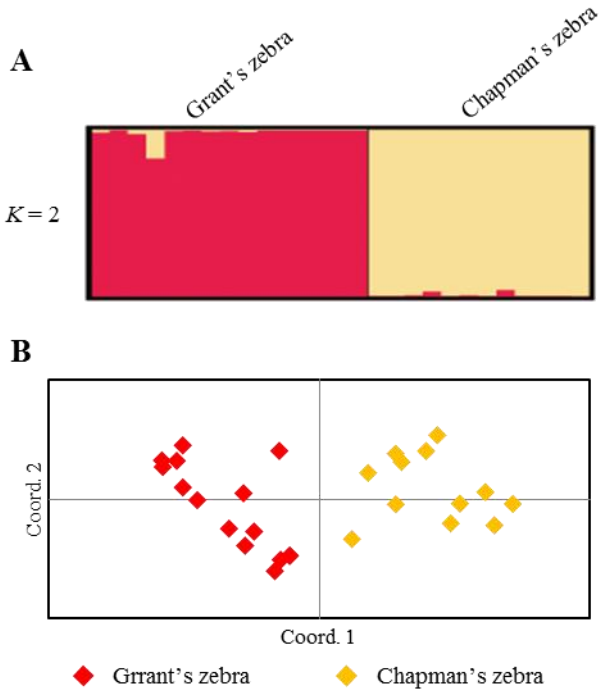


Fig. 2.2.2. (a) Bayesian analysis of the genetic structure showing differentiation of two Plains zebra subspecies (Grant's and Chapman's) based on 28 microsatellite loci. (b) First and second components of a principal coordinate analysis of 28 microsatellite loci in plains zebra explained 16.6% and 10.8% of total variance, respectively.

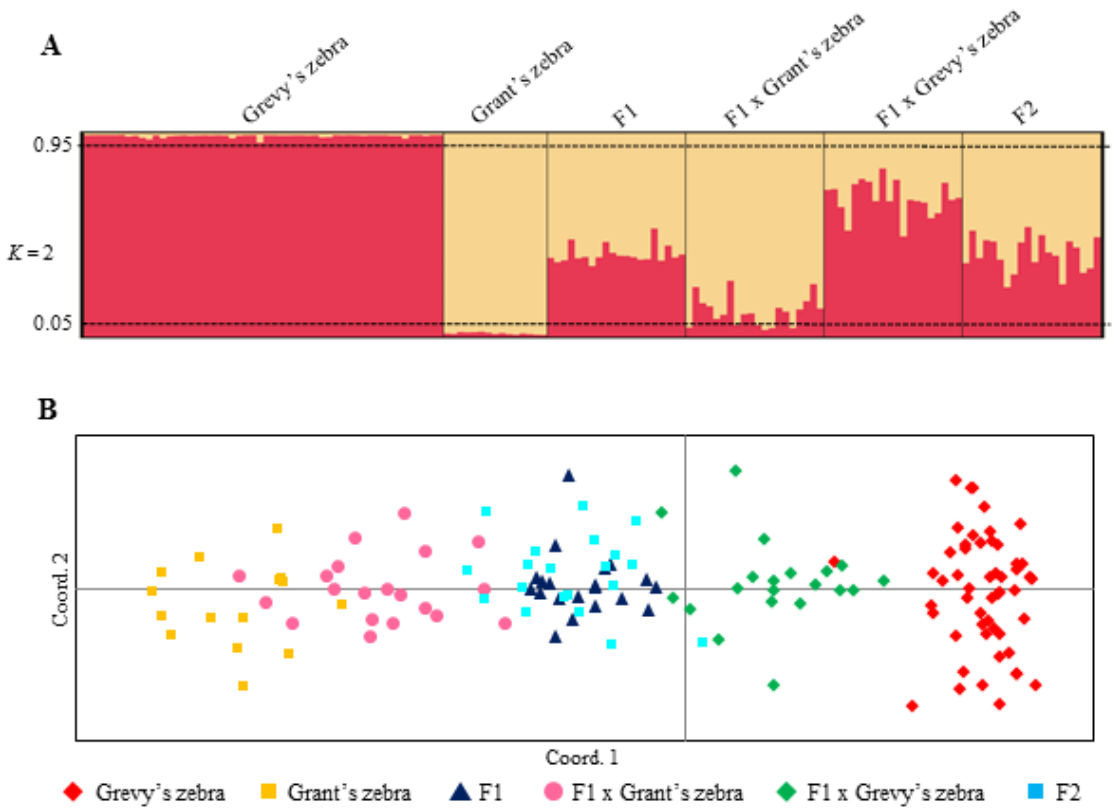


Fig. 2.2.3. (a) Bayesian analysis of the genetic structure of the Grevy's zebra, Grant's zebra and their hybrids (F1, F2, and back cross) based on 28 microsatellite loci. (b) First and second components of a principal coordinate analysis of 28 microsatellite loci in the Grevy's zebra, Grant's zebra and hybridization (F1, F2, and back cross). Percentages of variation explained by the first 2 axes were 18.4% and 3.8%, respectively.

Table 2.2.1 Results of cross-species amplification performed over the 28 microsatellite loci on the three zebra species and two subspecies of plains zebra (Grant's zebra and Chapman's zebra)

	n	*NumGI*	*NumPL*	*Ar*	*Na*	*Ho*	*He*	*PID*	*PID-sib*
Grevy's zebra	60	52	25	2.44 (1.00 - 4.23)	4.07 (1 - 9)	0.403 (0.000 - 0.885)	0.427 (0.000 - 0.782)	2.28×10^{-14}	1.06×10^{-6}
Plains zebra	53	27	27	3.21 (1.00 - 4.97)	5.32 (1 - 9)	0.496 (0.000 - 0.889)	0.573 (0.000 - 0.829)	7.63×10^{-21}	5.67×10^{-9}
Mountain zebra	10	6	23	2.47 (1.00 - 5.24)	2.75 (1 - 7)	0.305 (0.000 - 0.750)	0.368 (0.000 - 0.764)	7.18×10^{-20}	1.08×10^{-8}
Grant's zebra	33	15	27	3.15 (1.00 - 5.32)	4.64 (1 - 9)	0.550 (0.000 - 1.000)	0.559 (0.000 - 0.847)	3.53×10^{-18}	3.36×10^{-8}
Chapman's zebra	20	12	26	2.97 (1.00 - 4.79)	3.96 (1 - 7)	0.421 (0.000 - 0.833)	0.529 (0.000 - 0.799)	7.43×10^{-12}	9.77×10^{-6}

Abbreviations: *n*, number of tested individuals; *NumGI*, number of genotyped individuals; *NumPL*, number of polymorphic loci; *Ar*, allelic richness; *Na*, observed no. of alleles; *Ho*, observed heterozygosity; *He* expected heterozygosity; *PID*, Probability of Identity (unrelated); *PID-sib*, Probability of Identity of siblings.

Supplement Table 2.2.1 Profiles of 28 markers in each species and subspecies

Grevy's zebra (*Equus grevyi*)

Locus	Repeat unit	Allele range	*n*	*Ar*	*Na*	*Ho*	*He*		*PID*	*PID-sib*
EGR01	$(CAG)_5$	343-346	52	1.79	2	0.231	0.286		0.550	0.744
EGR02	$(AAT)_{11}$	163-180	50	2.35	4	0.440	0.521		0.321	0.570
EGR03	$(AGG)_8$	304-322	52	1.99	2	0.365	0.500		0.375	0.594
EGR05	$(CTT)_9$	154-175	52	2.70	5	0.558	0.595		0.247	0.514
EGR07	$(GGT)_8$	238-250	51	2.33	3	0.510	0.544		0.312	0.556
EGR08	$(AGAT)_{10}$	218-245	51	3.44	5	0.549	0.686		0.145	0.443
EGR09	$(AGAT)_{11}$	379-403	51	3.01	4	0.627	0.642		0.191	0.477
EGR10	$(AGGC)_8$	336-363	51	1.92	5	0.294	0.260		0.565	0.761
EGR11	$(ATCT)_{11}$	315-335	52	3.01	4	0.654	0.595		0.215	0.506
EGR12	$(ATCT)_9$	301-330	52	3.91	6	0.712	0.759		0.098	0.395
EGR13	$(ATCT)_9$	249-266	52	2.90	4	0.538	0.616		0.220	0.497
EGR14	$(CA)_{13}$	213-227	50	2.72	3	0.600	0.575		0.244	0.523
EGR15	$(CA)_{14}$	163-187	52	4.23	8	0.885	0.782		0.080	0.379
EGR16	$(CA)_{18}$	300-328	52	3.77	8	0.673	0.689		0.129	0.438
EGR18	$(CAG)_4$	295-304	48	1.93	4	0.229	0.245		0.581	0.773
EGR21	$(CAG)_4$	276-303	51	1.45	5	0.078	0.113		0.789	0.891
EGR22	$(CAG)_4$	171-188	52	1.15	3	0.038	0.038		0.926	0.963
EGR24	$(AAC)_{10}$	202-219	52	1.39	2	0.115	0.109		0.800	0.896
EGR25	$(AAT)_9$	378-393	52	1.00	1	0.000	0.000		1.000	1.000
EGR26	$(ACC)_9$	270-288	52	1.08	2	0.019	0.019		0.962	0.981
EGR27	$(CTT)_9$	142-163	52	1.08	2	0.019	0.019		0.962	0.981
EGR28	$(GAT)_9$	250-262	51	1.00	1	0.000	0.000		1.000	1.000
EGR29	$(GAT)_9$	280-301	51	1.00	1	0.000	0.000		1.000	1.000
EGR30	$(CA)_{13}$	136-172	51	4.12	9	0.784	0.727		0.102	0.412
EGR31	$(CA)_{14}$	367-383	46	3.33	4	0.543	0.666		0.159	0.457
EGR32	$(CA)_{15}$	342-360	52	3.58	5	0.692	0.727		0.121	0.417
EGR33	$(CA)_{16}$	326-348	52	3.41	7	0.692	0.672		0.154	0.453
EGR34	$(CA)_{16}$	196-216	52	2.68	5	0.442	0.585	*	0.257	0.522
Average				2.44	4.07	0.403	0.427		0.447	0.648

Significance of deviation from Hardy–Weinberg equilibrium at *P*-levels 0.05 (*) and 0.01 (**), and Bonferroni corrected

Abbreviations: *n*, number of tested individuals; *Ar*, allelic richness; *Na*, observed no. of alleles; *Ho*, observed heterozygosity; *He* expected heterozygosity; *PID*, Probability of Identity (unrelated); *PID-sib*, Probability of Identity of siblings.

Supplement Table 2.2.1 Profiles of 28 markers in each species and subspecies (continue)

Plains zebra (*Equus quagga*)

Locus	Repeat unit	*n*	*Ar*	*Na*	*Ho*	*He*		*PID*	*PID-sib*
EGR01	$(CAG)_5$	26	1.29	2	0.000	0.074	**	0.860	0.928
EGR02	$(AAT)_{11}$	27	2.33	4	0.296	0.445		0.367	0.619
EGR03	$(AGG)_8$	27	3.10	5	0.667	0.647		0.191	0.474
EGR05	$(CTT)_9$	27	3.12	6	0.667	0.638		0.197	0.480
EGR07	$(GGT)_8$	26	3.40	5	0.423	0.618		0.181	0.486
EGR08	$(AGAT)_{10}$	27	3.78	7	0.704	0.718		0.125	0.422
EGR09	$(AGAT)_{11}$	26	4.14	6	0.654	0.755		0.094	0.396
EGR10	$(AGGC)_8$	26	2.71	6	0.538	0.442		0.337	0.613
EGR11	$(ATCT)_{11}$	26	3.74	6	0.538	0.705		0.126	0.429
EGR12	$(ATCT)_9$	27	4.15	8	0.556	0.757		0.094	0.395
EGR13	$(ATCT)_9$	27	2.30	3	0.593	0.444		0.370	0.621
EGR14	$(CA)_{13}$	27	3.18	6	0.630	0.643		0.189	0.476
EGR15	$(CA)_{14}$	24	3.70	6	0.333	0.720	**	0.124	0.421
EGR16	$(CA)_{18}$	27	4.81	7	0.889	0.829		0.051	0.348
EGR18	$(CAG)_4$	26	1.94	2	0.538	0.393		0.445	0.665
EGR21	$(CAG)_4$	26	2.15	4	0.038	0.371	**	0.441	0.675
EGR22	$(CAG)_4$	20	1.00	1	0.000	0.000		1.000	1.000
EGR24	$(AAC)_{10}$	26	2.53	4	0.462	0.544		0.291	0.551
EGR25	$(AAT)_9$	27	3.79	6	0.593	0.729		0.117	0.415
EGR26	$(ACC)_9$	27	2.88	5	0.556	0.514		0.274	0.562
EGR27	$(CTT)_9$	23	4.81	8	0.652	0.819		0.055	0.354
EGR28	$(GAT)_9$	26	1.71	2	0.269	0.233		0.615	0.787
EGR29	$(GAT)_9$	26	2.09	4	0.308	0.276		0.536	0.746
EGR30	$(CA)_{13}$	23	4.97	9	0.739	0.820		0.052	0.353
EGR31	$(CA)_{14}$	24	3.25	6	0.583	0.601		0.201	0.500
EGR32	$(CA)_{15}$	26	3.63	5	0.615	0.712		0.129	0.426
EGR33	$(CA)_{16}$	27	4.77	9	0.407	0.800	*	0.061	0.365
EGR34	$(CA)_{16}$	27	4.57	7	0.630	0.810		0.063	0.361
Average			3.21	5.32	0.496	0.573		0.271	0.531

Significance of deviation from Hardy–Weinberg equilibrium at *P*-levels 0.05 (*) and 0.01 (**), and Bonferroni corrected

Abbreviations: *n*, number of tested individuals; *Ar*, allelic richness; *Na*, observed no. of alleles; *Ho*, observed heterozygosity; *He* expected heterozygosity; *PID*, Probability of Identity (unrelated); *PID-sib*, Probability of Identity of siblings.

Supplement Table 2.2.1 Profiles of 28 markers in each species and subspecies (continue)

Mountain zebra (*Equus zebra*)

Locus	Repeat unit	*n*	*Ar*	*Na*	*Ho*	*He*	*PID*	*PID-sib*
EGR01	$(CAG)_5$	6	1.00	1	0.000	0.000	1.000	1.000
EGR02	$(AAT)_{11}$	6	1.67	2	0.167	0.153	0.729	0.856
EGR03	$(AGG)_8$	6	1.00	1	0.000	0.000	1.000	1.000
EGR05	$(CTT)_9$	6	3.00	4	0.500	0.417	0.364	0.633
EGR07	$(GGT)_8$	6	1.91	2	0.000	0.278	0.560	0.751
EGR08	$(AGAT)_{10}$	6	2.67	3	0.667	0.569	0.278	0.535
EGR09	$(AGAT)_{11}$	5	2.98	3	0.200	0.620	0.217	0.494
EGR10	$(AGGC)_8$	6	2.58	3	0.500	0.403	0.396	0.648
EGR11	$(ATCT)_{11}$	5	1.80	2	0.200	0.180	0.689	0.832
EGR12	$(ATCT)_9$	6	4.15	5	0.500	0.681	0.140	0.445
EGR13	$(ATCT)_9$	6	2.66	3	0.333	0.542	0.292	0.552
EGR14	$(CA)_{13}$	6	2.91	3	0.500	0.625	0.220	0.493
EGR15	$(CA)_{14}$	6	3.48	4	0.667	0.597	0.207	0.503
EGR16	$(CA)_{18}$	6	3.80	4	0.500	0.708	0.135	0.429
EGR18	$(CAG)_4$	5	1.00	1	0.000	0.000	1.000	1.000
EGR21	$(CAG)_4$	6	1.91	2	0.000	0.278	0.560	0.751
EGR22	$(CAG)_4$	5	2.60	3	0.200	0.340	0.461	0.695
EGR24	$(AAC)_{10}$	6	1.00	1	0.000	0.000	1.000	1.000
EGR25	$(AAT)_9$	4	2.00	2	0.000	0.375	0.461	0.678
EGR26	$(ACC)_9$	5	1.80	2	0.200	0.180	0.689	0.832
EGR27	$(CTT)_9$	6	1.91	2	0.333	0.278	0.560	0.751
EGR28	$(GAT)_9$	6	1.91	2	0.000	0.278	0.560	0.751
EGR29	$(GAT)_9$	5	1.00	1	0.000	0.000	1.000	1.000
EGR30	$(CA)_{13}$	4	4.00	4	0.750	0.656	0.169	0.464
EGR31	$(CA)_{14}$	6	2.65	3	0.500	0.486	0.327	0.589
EGR32	$(CA)_{15}$	4	4.00	4	0.750	0.563	0.229	0.526
EGR33	$(CA)_{16}$	6	5.24	7	0.667	0.764	0.080	0.388
EGR34	$(CA)_{16}$	5	2.60	3	0.400	0.340	0.461	0.695
Average			2.47	2.75	0.305	0.368	0.492	0.689

Significance of deviation from Hardy–Weinberg equilibrium at *P*-levels 0.05 (*) and 0.01 (**), and Bonferroni corrected

Abbreviations: *n*, number of tested individuals; *Ar*, allelic richness; *Na*, observed no. of alleles; *Ho*, observed heterozygosity; *He* expected heterozygosity; *PID*, Probability of Identity (unrelated); *PID-sib*, Probability of Identity of siblings.

Supplement Table 2.2.1 Profiles of 28 markers in each species and subspecies (continue)

Grant's zebra (*Equus quagga boehmi*)

Locus	Repeat unit	*n*	*Ar*	*Na*	*Ho*	*He*		*PID*	*PID-sib*
EGR01	$(CAG)_5$	15	1.47	2	0.000	0.124	*	0.774	0.881
EGR02	$(AAT)_{11}$	15	2.41	4	0.267	0.420		0.382	0.635
EGR03	$(AGG)_8$	15	2.71	4	0.600	0.562		0.268	0.536
EGR05	$(CTT)_9$	15	2.72	4	0.667	0.576		0.261	0.528
EGR07	$(GGT)_8$	15	2.55	4	0.267	0.429		0.362	0.626
EGR08	$(AGAT)_{10}$	15	3.31	5	0.867	0.662		0.172	0.462
EGR09	$(AGAT)_{11}$	15	3.73	6	0.600	0.693		0.137	0.438
EGR10	$(AGGC)_8$	15	2.89	5	0.600	0.480		0.300	0.585
EGR11	$(ATCT)_{11}$	15	3.71	5	0.667	0.698		0.133	0.434
EGR12	$(ATCT)_9$	15	4.28	6	0.600	0.773		0.086	0.385
EGR13	$(ATCT)_9$	15	1.92	2	0.467	0.358		0.476	0.690
EGR14	$(CA)_{13}$	15	3.69	6	0.667	0.676		0.146	0.449
EGR15	$(CA)_{14}$	15	4.00	6	0.533	0.724		0.113	0.416
EGR16	$(CA)_{18}$	15	4.91	7	0.933	0.829		0.053	0.349
EGR18	$(CAG)_4$	14	1.96	2	0.571	0.408		0.434	0.654
EGR21	$(CAG)_4$	14	1.99	2	0.000	0.459	*	0.398	0.620
EGR22	$(CAG)_4$	11	1.00	1	0.000	0.000		1.000	1.000
EGR24	$(AAC)_{10}$	15	2.72	4	0.600	0.576		0.261	0.528
EGR25	$(AAT)_9$	15	3.65	6	0.800	0.671		0.151	0.452
EGR26	$(ACC)_9$	15	2.61	4	0.600	0.464		0.332	0.601
EGR27	$(CTT)_9$	11	4.84	7	0.818	0.810		0.063	0.361
EGR28	$(GAT)_9$	15	1.87	2	0.400	0.320		0.514	0.718
EGR29	$(GAT)_9$	15	1.94	3	0.267	0.240		0.591	0.778
EGR30	$(CA)_{13}$	15	5.32	9	1.000	0.847		0.041	0.337
EGR31	$(CA)_{14}$	15	3.51	6	0.533	0.600		0.190	0.497
EGR32	$(CA)_{15}$	15	3.49	4	0.600	0.696		0.144	0.438
EGR33	$(CA)_{16}$	15	4.72	8	0.667	0.796		0.068	0.369
EGR34	$(CA)_{16}$	15	4.24	6	0.800	0.771		0.088	0.386
Average			3.15	4.64	0.550	0.559		0.283	0.541

Significance of deviation from Hardy–Weinberg equilibrium at *P*-levels 0.05 (*) and 0.01 (**), and Bonferroni corrected

Abbreviations: *n*, number of tested individuals; *Ar*, allelic richness; *Na*, observed no. of alleles; *Ho*, observed heterozygosity; *He* expected heterozygosity; *PID*, Probability of Identity (unrelated); *PID-sib*, Probability of Identity of siblings.

Supplement Table 2.2.1 Profiles of 28 markers in each species and subspecies (continue)

Chapman's zebra (*Equus quagga chapmani*)

Locus	Repeat unit	*n*	*Ar*	*Na*	*Ho*	*He*		*PID*	*PID-sib*
EGR01	$(CAG)_5$	11	1.00	1	0.000	0.000		1.000	1.000
EGR02	$(AAT)_{11}$	12	2.30	3	0.333	0.469		0.360	0.606
EGR03	$(AGG)_8$	12	3.24	4	0.750	0.677		0.167	0.453
EGR05	$(CTT)_9$	12	3.21	4	0.667	0.674		0.172	0.456
EGR07	$(GGT)_8$	11	4.21	5	0.636	0.760		0.094	0.393
EGR08	$(AGAT)_{10}$	12	3.78	5	0.500	0.701		0.131	0.432
EGR09	$(AGAT)_{11}$	11	4.31	5	0.727	0.781		0.083	0.380
EGR10	$(AGGC)_8$	11	2.49	4	0.455	0.380		0.411	0.663
EGR11	$(ATCT)_{11}$	11	3.88	6	0.364	0.702		0.131	0.431
EGR12	$(ATCT)_9$	12	3.78	6	0.500	0.698		0.140	0.436
EGR13	$(ATCT)_9$	12	2.66	3	0.750	0.531		0.283	0.555
EGR14	$(CA)_{13}$	12	2.00	2	0.583	0.497		0.377	0.596
EGR15	$(CA)_{14}$	9	1.99	2	0.000	0.444		0.407	0.630
EGR16	$(CA)_{18}$	12	4.44	6	0.833	0.781		0.080	0.379
EGR18	$(CAG)_4$	12	1.94	2	0.500	0.375		0.461	0.678
EGR21	$(CAG)_4$	12	1.67	3	0.083	0.156	**	0.718	0.851
EGR22	$(CAG)_4$	9	1.00	1	0.000	0.000		1.000	1.000
EGR24	$(AAC)_{10}$	11	2.34	3	0.273	0.492		0.342	0.590
EGR25	$(AAT)_9$	12	3.72	5	0.333	0.694		0.138	0.437
EGR26	$(ACC)_9$	12	3.12	4	0.500	0.559		0.235	0.529
EGR27	$(CTT)_9$	12	3.79	6	0.500	0.698		0.135	0.435
EGR28	$(GAT)_9$	11	1.36	2	0.091	0.087		0.838	0.916
EGR29	$(GAT)_9$	11	2.13	3	0.364	0.310		0.504	0.721
EGR30	$(CA)_{13}$	8	3.77	6	0.250	0.578		0.203	0.512
EGR31	$(CA)_{14}$	9	2.44	3	0.667	0.537		0.311	0.559
EGR32	$(CA)_{15}$	11	3.45	4	0.636	0.690		0.153	0.443
EGR33	$(CA)_{16}$	12	4.31	6	0.083	0.747	**	0.096	0.401
EGR34	$(CA)_{16}$	12	4.79	7	0.417	0.799		0.066	0.367
Average			2.97	3.96	0.421	0.529		0.323	0.566

Significance of deviation from Hardy–Weinberg equilibrium at *P*-levels 0.05 (*) and 0.01 (**), and Bonferroni corrected

Abbreviations: *n*, number of tested individuals; *Ar*, allelic richness; *Na*, observed no. of alleles; *Ho*, observed heterozygosity; *He* expected heterozygosity; *PID*, Probability of Identity (unrelated); *PID-sib*, Probability of Identity of siblings.

2.3 Mitochondrial DNA sequence diversity of Grevy's zebra for management in captivity

2.3.1 Introduction

The Grevy's zebra inhabits the arid to semi-arid region in northern Kenya and Ethiopia. It is the largest species among wild equids and is characterized by the narrow width of its stripes and large ears. During recent decades, this species has suffered severe population declines due to habitat destruction, overhunting and competition with livestock. The current population size is estimated at about 2,800 individuals (KWS, 2012), and the species is listed under CITES Appendix I and classified as 'Endangered' in the IUCN's red list (Moehlman et al., 2013).

Ex situ conservation programmes in many endangered species are established, mainly in zoos. The primary goal of captive breeding programmes is to maintain the genetic diversity of populations to avoid the negative effect of inbreeding and preserve their adaptive potential (Frankham et al., 2010). The current goal of most captive breeding programmes is to retain 90% of genetic diversity observed in the founding population for 100 years (Frankham et al., 2010). Because of the potential for relatedness among individual founders, the true level of genetic diversity is likely to lower than in any theoretical founding population. This situation causes higher levels of inbreeding and lower genetic diversity than estimated from the pedigree data. In addition, pedigree records may contain errors or data gaps. To try to address these issues, several captive breeding programmes have recently begun to combine molecular genetic analysis with traditional pedigree analysis (Ferrie et al., 2013; Gautschi et al., 2003; Henkel et al., 2012; Ivy et al., 2009; McGreevy et al., 2011; Ogden et al., 2007). These combined analyses can allow us to compare genetic diversity expected by pedigree analysis with that of genetic variation observed in the molecular data, and to develop more appropriate genetic management.

The breeding programme for the conservation of Grevy's zebra is performed in zoos around the world. A total of 508 Grevy's zebras are currently kept in captivity around the world (Langenhorst, 2014a). Most individuals are managed by the EAZA and the AZA and small numbers are kept by the JAZA as the third population. The captive population represents 15% of the total global population (wild and captivity), and is a very important resource for conservation of this species. The Grevy's zebra is listed in "Top Ten Mammal Species Reliant on Zoos" in the British and Irish Association of Zoos and Aquariums (BIAZA) (http://www.biaza.org.uk/home/). While the management of populations runs smoothly in EAZA and AZA, the population in JAZA has decreased from 50 individual in 1982 to 19 individuals at the end of 2013 (Langenhorst, 2014a). Therefore a captive breeding programme that includes genetic management is required, however there is little information about the genetic diversity of the Grevy's zebra.

In this study, I examined the genetic diversity of the captive population of Grevy's zebra in Japan and the United Kingdom by analyzing the cytochrome b (*cytb*) and control region sequences of mitochondrial DNA (mtDNA). In addition, for comparative purposes, I investigated the same genetic markers in the congeneric species of plains zebra and Hartmann's mountain zebra. This molecular information will be applied to future captive population management and implementation of breeding programmes.

Methods

Sample collection

Samples were the same with chapter 2.2. Sampling and DNA extraction method was explained in chapter 2.2.

PCR Amplification and Sequencing for mtDNA

A 382 base pair region of the cytochrome b (*cytb*) gene was amplified by PCR using universal primers; L15146 (5'- CATGAGGACAAATATCATTCTGAG-3') (Ducroz et al., 1998) and H15553 (5'- TAGGCAAATAGGAAATATCATTCTGGT-3') (Kocher et al., 1989). Primers MDLF (5'-CCAAAGCTGAAATTCTACTTA-3') and MDLR (5'-CATCGAGATGTCTTATTTAAGG-3') were designed from a consensus of obtained sequences to amplify 445 bp of the 5' end of the mountain zebra control region (Moodley and Harley, 2006). For amplification of both markers the PCR mixture contained 0.75 U of LA-Taq DNA polymerase (TaKaRa, Japan), PCR buffer, 400 µM of each dNTP, 0.4 µM each of forward and reverse primers, 0.1 µg of T4 Gene 32 Protein (Nippon Gene, Japan) and 20 ng of template DNA in a total volume of 15 µl. PCR cycling condition consisted of initial denaturation of 95˚C for 2 min, followed by 35 cycles of 95˚C for 30 sec, 55˚C for 30 sec, 74˚C for 1 min and a final extension of 74˚C for 10 min. In faecal samples, instead of 20 ng DNA, 2 µl of extracted DNA solution and 0.1 µg of T4 Gene 32 Protein (Nippon Gene) were added. In faecal samples, the PCR conditions consisted of an initial denaturation at 95 °C for 15 min, followed by 15 cycles of denaturation at 94 °C for 30 s, annealing at 57 °C for 1 min 30 s, extension at 72 °C for 1 min, and followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 52 °C for 1 min 30 s, extension at 72 °C for 1 min, and final extension at 60 °C for 30 min. PCR products were purified using High Pure PCR purification kit (Roche) and sequenced using Big Dye Terminator ver. 3.1 cycle sequencing kit (Applied Biosystems) according to the manufacturers protocol and electrophoresed on an ABI PRISM 3130xl sequencer (Applied Biosystems). Both forward and reverse complements of the reverse sequences were aligned to get a consensus sequence using MEGA ver. 6.0 (Tamura et al., 2013).

DNASP ver. 5.10.01 (Librado and Rozas, 2009) was used to determine the number of polymorphic sites, number of haplotypes and haplotype diversity (*Hd*). Nucleotide diversity

(*Nd*) and haplotype richness (*Hr*) in species were calculated with Arlequin ver. 3.5 (Excoffier and Lischer, 2010) and CONTRIB ver. 1.4 (Petit et al., 1998), respectively. Haplotype diversity indicates genetic diversity within populations and nucleotide diversity is estimated as the function of the number of polymorphic sites and the frequency of transitions or transversions and insertions or deletions within the population. A haplotype network was constructed based on statistical parsimony methods (TCS approach) (Posada and Crandall, 2001; Templeton et al., 1992) using software POPART (http://popart.otago.ac.nz/documentation.shtml).

Pedigree Analysis

The number of individuals, living descendants, effective number of individuals (Ne), Ne/N, mean kinship (MK), Gene Value, the numbers of founders, founder genome equivalents (FGE) and genetic diversity (GD) in the Grevy's zebra and mountain zebra were calculated using PMx software (Lacy et al., 2012) based on the International studbook of Grevy's zebra (Langenhorst, 2014a) and Hartmann's mountain zebra (Langenhorst, 2014b). No studbook data is available for Plains zebra.

Results

Nucleotide sequences of 382 bp in *cytb* gene were determined for a total of 117 individuals across the three zebra species (Fig. 2.3.1). The sequences were monomorphic in Grevy's zebras (accession number LC097287) while multiple sequence alignments in plains zebra showed 24 polymorphic sites (21 transitions, three transversions), generating a total of nine different haplotypes (accession numbers LC097292-LC097300). In mountain zebra four haplotypes were observed derived from 11 nucleotide transitions (accession numbers LC097288-LC097291).

A total of 107 control region nucleotide sequences were generated. Sequence length was determined to be 445 bp in Grevy's zebra, 444 bp in plains zebra and mountain zebra (Fig. 2.3.2), with the length difference due to two indel at 95 bp (Grevy's: T; Mountain: C; Plains: deletion) and 324 bp (Grevy's: T; Plains: T; Mountain: deletion) from the 5' end of the sequence. Grevy's zebra had only one variable site (transition) (accession numbers LC097301-LC097302). Plains zebra showed 32 polymorphic sites (30 transitions and two transversions), generating a total of 12 different haplotypes (accession numbers LC097306-LC097317). Mountain zebra displayed 13 polymorphic sites (12 transitions and one transversion), generating a total of three different haplotypes (accession numbers LC097303-LC097306).

The number of haplotypes in *cytb* and control region indicated Table 2.3.1a and 2.3.1b. Haplotype networks based on the *cytb* and control region sequences show a distinct lack of diversity at both markers in Grevy's zebra, despite having the largest sample size (Fig. 2.3.3a and 2.3.3b, respectively). All diversity indices (haplotype diversities, nucleotide diversity, and haplotype richness) calculated among the three zebra species were lowest in Grevy's zebra (Table 2.3.2).

Pedigree Analysis

The number of individuals, living descendants, Ne, Ne/N, MK, Gene Value, FGE and GD in the Grevy's zebra and mountain zebra are indicated in Table 2.3.3. All of these that indicators of genetic diversity, based on calculations from studbook data, clearly show that theoretically, Grevy's zebra should have significantly more genetic diversity than mountain zebra. This is particularly evident from estimates of Ne, where the value for Grevy's zebra is almost ten times larger than for mountain zebra, and for MK, where mountain zebra are estimated to have more than three times greater mean pairwise relatedness than Grevys's

zebra. These results are in stark contrast to those for mitochondrial haplotype diversity.

Discussion

Even though there were more samples from the Grevy's zebra, the Grevy's zebra was found to have the lowest index of molecular genetic diversity (the number of haplotypes, the haplotype diversity, the nucleotide diversity and haplotype richness) compared to the plains zebra and mountain zebra at *cytb* and control region sequences. This pattern has also been observed in other genera that include endangered species (Moodley and Harley, 2006). Although the plains zebra population is the largest of the three zebra species in both captivity and the wild, Grevy's zebra has a larger population size (n = 508) and number of founders (n = 123) than mountain zebra (population size; n = 209, number of founders; n = 48), and the overall captive breeding programme for Grevy's zebra is considered to be better established with greater historical data than that of mountain zebra (Table 2.3.3). Considering this situation, it is unlikely that Grevy's zebra have experienced greater loss of genetic diversity than mountain zebra since establishment of the captive population. Since the numbers of Grevy's zebra (estimate = 2,800) (Moehlman et al., 2013) are smaller than that of mountains zebra (25,000 individuals) (Novellie, 2008) in the wild, the low genetic diversity of Grevy's zebra in captivity is likely to be derived from low genetic diversity of the founding population. Low genetic diversity of founding population is thought two possibilities. One is founders were introduced from limited regions, the other is wild population had low genetic diversity. Although the detailed origin of many founder is unknown (the origins of almost individuals were registered as East Africa, not country), at least in the individuals for whom detailed areas are recorded, three countries (Kenya, Ethiopia, and Somalia) are recorded as the captured areas. Therefore it is thought that the founding population was introduced from the plural areas. Because the periods that founders were incorporated into the captive

population were from 1950's to 1970's (Langenhorst, 2014a), when there were estimated to be more than 10,000 individuals in the wild (KWS, 2012), these possibilities may indicate that a reduction in genetic diversity in the wild populations occurred before the 1970's, that is, prior before drastic population declines.

Mitochondrial DNA is used extensively to study population structure and to trace maternal lineages in both the wild and captivity. Due to the presence of multiple copies within a cell, success rate of PCR amplification/genotyping of mtDNA is significantly higher than nuclear DNA markers (such as microsatellites), which is particularly advantageous when using non-invasive faecal material. However, in this study, no evidence of population structure was found between the two Grevy's zebra populations (UK and Japan) and the mtDNA regions analyzed had extremely low genetic diversity, limiting their usefulness as tools for genetic management of captive Grevy's zebra (for example, in evaluating genetic diversity in local populations or analyzing maternal lineage). To complement the studbook data in managing Grevy's zebra within the breeding programmes, the analysis of additional nuclear DNA markers (such as single nucleotide polymorphisms or microsatellites) would therefore be required. Recent publication of whole genome sequence data for the three zebra species in this study (Jonsson et al., 2014) has created the possibility for much more detailed analysis of nuclear diversity, raising the possibility to understand functional genetic variation relevant to conservation management of these charismatic equids.

In this study, I showed that the genetic diversity of mtDNA in Grevy's zebra was low compared to congeneric zebras in captivity. Analysis of studbook data indicates that this low genetic diversity is likely to be derived from founder populations, rather than drift in captivity, and that low diversity in the wild may pre-date recorded population declines in the late 20th century. Importantly, the comparative molecular data for Grevy's and mountain zebra contradict the expected relative genetic diversity in these species estimated from

pedigree analysis of studbook records. This finding emphasizes the need to validate theoretical estimates of genetic diversity in captive breeding programmes with empirical molecular genetic data.

Table 2.3.1a Number of haplotypes in cytochrome b gene

	Haplotype	Number of individuals	Accession number
Grevy's zebra	*cytb-G1*	58	LC097287
Plains zebra	*cytb-PG1*	8	LC097292
	cytb-PG2	1	LC097293
	cytb-PG3	1	LC097294
	cytb-PG4	4	LC097295
	cytb-PG5	1	LC097296
	cytb-PCG1	22	LC097297
	cytb-PC1	2	LC097298
	cytb-PC2	9	LC097299
	cytb-PC3	1	LC097300
Mountain zebra	*cytb-M1*	2	LC097288
	cytb-M2	1	LC097289
	cytb-M3	2	LC097290
	cytb-M4	4	LC097291

Table 2.3.1b Number of haplotypes in control region

	Haplotype	Number of individuals	Accession number
Grevy's zebra	*CR-G1*	39	LC097301
	CR-G2	14	LC097302
Plains zebra	*CR-PG1*	4	LC097306
	CR-PG2	7	LC097307
	CR-PG3	17	LC097308
	CR-PG4	1	LC097309
	CR-PG5	1	LC097310
	CR-PG6	1	LC097311
	CR-PG7	1	LC097312
	CR-PC1	7	LC097313
	CR-PC2	1	LC097314
	CR-PC3	2	LC097315
	CR-PC4	2	LC097316
	CR-PC5	2	LC097317
Mountain zebra	*CR-M1*	5	LC097303
	CR-M2	1	LC097304
	CR-M3	2	LC097305

Table 2.3.2 Genetic diversity indices of three zebra species in cytochrome b gene and control region

		amplified length	*n*	*NumPs*	*NumH*	*Hd*	*Nd*	*Pb*
cytb	Grevy's zebra	382	58	0	1	-	-	0
	Mountain zebra	382	10	11	4	0.778 +/- 0.012	0.012 +/- 0.007	2.889
	Plains zebra	382	49	24	10	0.761 +/- 0.022	0.012 +/- 0.007	3.252
control region	Grevy's zebra	445	53	1	2	0.357 +/- 0.004	0.001 +/- 0.001	0.931
	Mountain zebra	444	9	13	3	0.607 +/- 0.027	0.011 +/- 0.007	2
	Plains zebra	444	45	32	12	0.827 +/- 0.002	0.017 +/- 0.009	3.892

n: number of individuals, *NumPs*: number of polymorphic sites, *NumH*: number of haplotypes, *Hd*: haplotype diversity, *Nd*: nucleotide diversity, *Pb*: Haplotype richness with rarefaction

Table 2.3.3 Genetic diversity indexes from the pedigree (studbook) data

								Number of founders		
	n	LD	Ne	Ne/N	MK	Gene Value	FGE	GD	Total	Males	Females
Grevy's zebra											
World	508	468.42	164.21	0.3506	0.0138	0.9855	36.43	0.9863	123	42	81
JAZA	20	17.73	5.17	0.2916	0.099	0.8937	5.05	0.901	47	20	27
UK	48	45.61	9.83	0.2155	0.0419	0.9581	17.45	0.9581	58	19	39
Samples	48	43.75	22.18	0.5070	0.0377	0.9617	13.26	0.9623	79	29	50
Hartmann's mountain zebra											
World	209	117.9	53.46	0.4534	0.0638	0.929	7.83	0.9362	48	19	29
Samples	10	7.62	2.57	0.3371	0.1153	0.7648	4.33	0.8484	26	11	15

LD: Living Descendants, Ne: effective population size, MK: mean kinship, FGE: founders genome equivalent, GD: genetic diversity

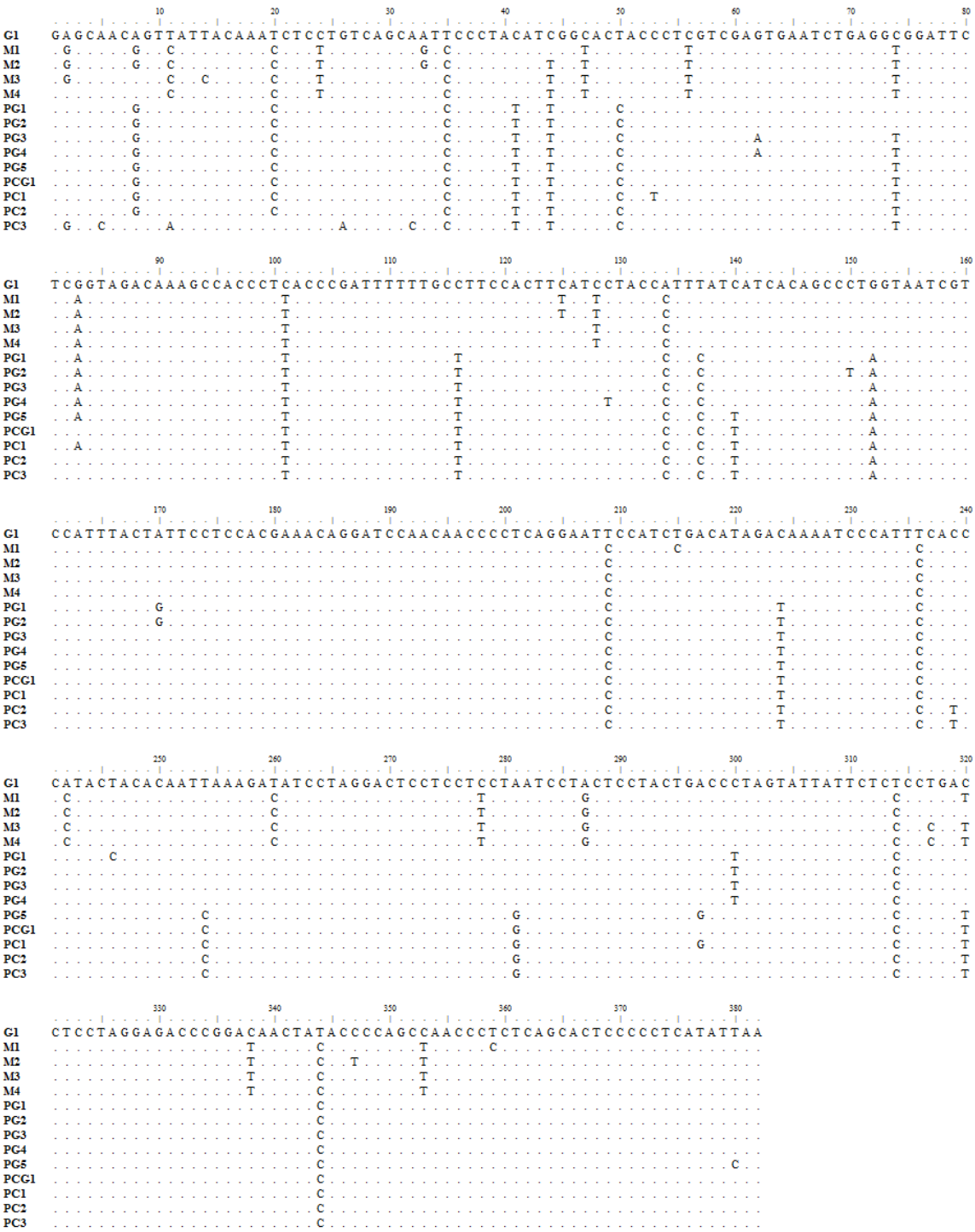


Fig. 2.3.1 Sequences of haplotype in *cytochrome b* gene
G1 indicate Grevy's zebra haplotype, P1-P9 are plains zebra haplotypes (PG1-PG5, grant's zebra haplotypes; PC1-PC3, Chapman's zebra haplotypes; PCG1, shared both subspecies haplotype), M1-M4 are mountain zebra haplotypes.

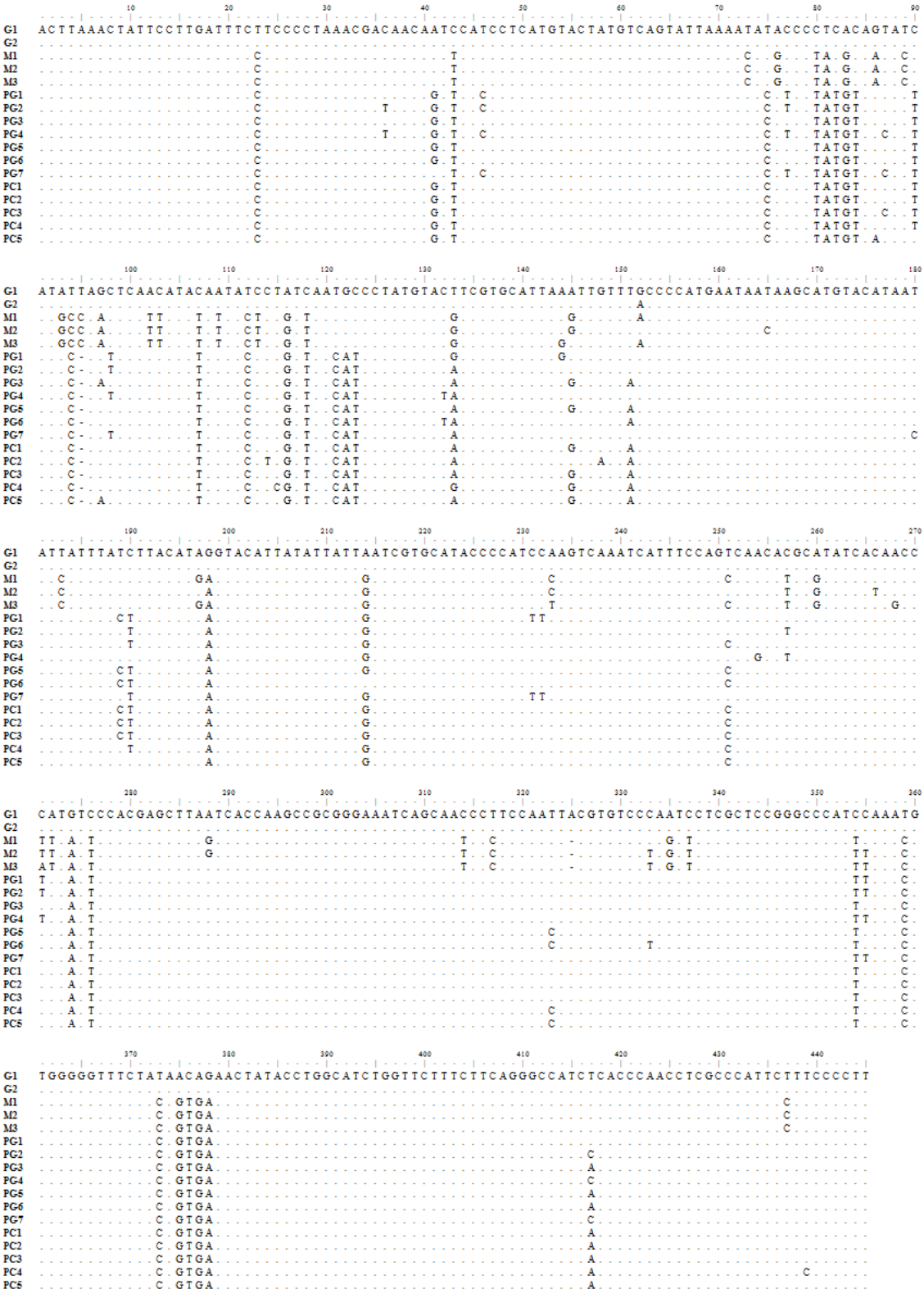


Fig. 2.3.2 Sequences of haplotype in control region
G1-G2 indicate Grevy's zebra haplotypes, P1-P12 are plains zebra haplotypes (PG1-PG7, grant's zebra haplotypes; PC1-PC5, Chapman's zebra haplotypes), M1-M3 are mountain zebra haplotypes

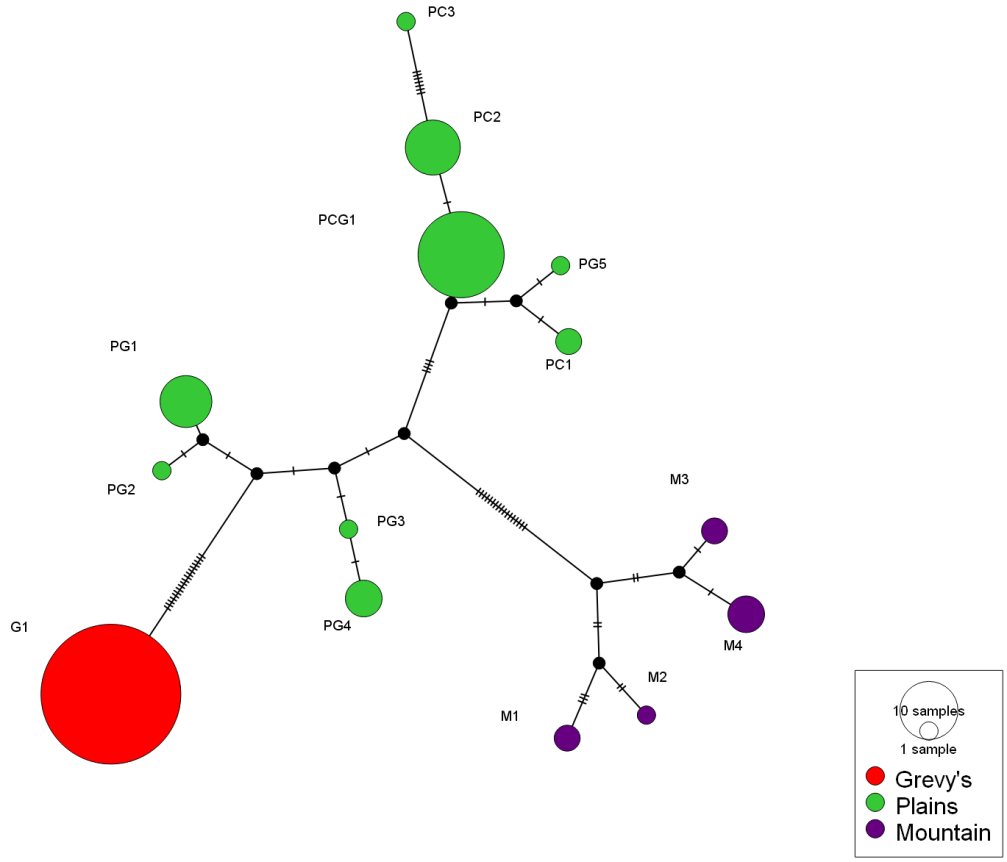


Fig. 2.3.3a Network among haplotypes of *cytb* gene. G1 indicate Grevy's zebra haplotype (red circle), P1-P9 are plains zebra haplotypes (PG1-PG5, grant's zebra haplotypes; PC1-PC3, Chapman's zebra haplotypes; PCG1, shared both subspecies haplotype) (green circles), M1-M4 are mountain zebra haplotypes (purple circles). The size of the circle is proportional to the total number of individuals of that haplotype. The crossbars show the number of substitution between haplotypes whilst the black nodes indicate missing haplotypes.

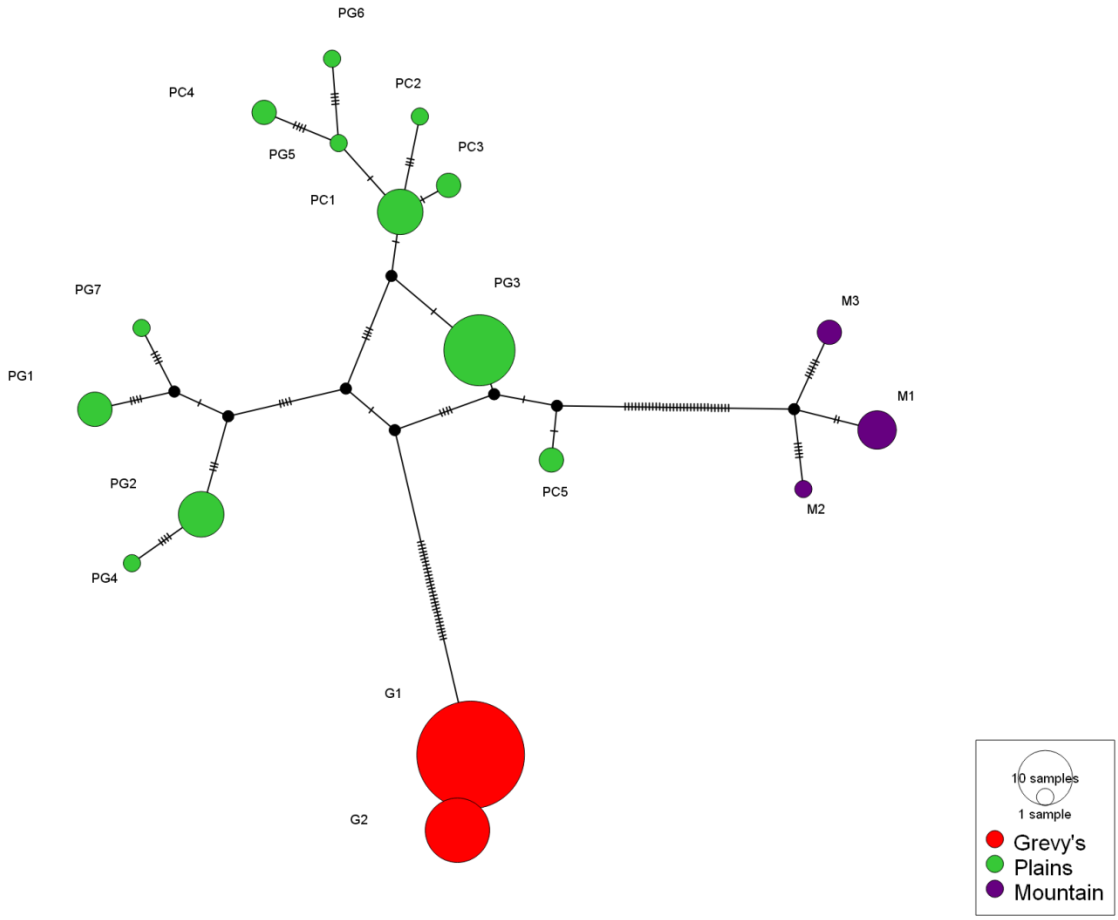


Fig. 2.3.3b Network among haplotypes of control region gene. G1-G2 indicate Grevy's zebra haplotypes (red circles), P1-P12 are plains zebra haplotypes (PG1-PG7, grant's zebra haplotypes; PC1-PC5, Chapman's zebra haplotypes) (green circles), M1-M3 are mountain zebra haplotypes (purple circles). The size of the circle is proportional to the total number of individuals of that haplotype. The crossbars show the number of substitution between haplotypes whilst the black nodes indicate missing haplotypes.

Chapter 3 Androgen receptor gene polymorphism in zebra species

3.1. Introduction

Androgen receptors mediate the effects of androgens that are responsible for development and maintenance of the male reproductive system. Deleterious mutations in *AR* can result in syndromes ranging from mild abnormalities to total failure of normal male phenotypic development (McPhaul, 2002a; McPhaul, 2002b). Androgen receptors also have associations with disease states (such as prostate cancer) and behaviour in humans (Collaer and Hines, 1995; Wyce et al., 2010; Zitzmann and Nieschlag, 2003). Androgen receptor is DNA-binding transcription factors, the main regulators of androgen signalling in the cell, activated mostly by testosterone and 5α-dihydrotestosterone. In humans the *AR* is located on the X chromosome and comprised of three important structural domains: the N-terminal transactivation domain, the central DNA binding domain, and the C-terminal ligand binding domain (Mangelsdorf et al., 1995). The N-terminal domain in *AR* contains two polymorphic trinucleotide repeat regions, CAG encoding glutamine and GGN encoding glycine (*ARQ* and *ARG*, respectively) in humans (Chang et al., 1988). *In vitro* studies have shown that a relatively short CAG repeat sequence enhances the transcriptional activity of the androgen receptor by promoting interactions between the receptor and coactivators (Chamberlain et al., 1994). The CAG repeat polymorphisms have been shown to affect reproductive capability, various disease risks and personality not only in men and but also in women (Choong and Wilson, 1998; Giovannucci et al., 1997; Mouritsen et al., 2013; Olsen et al., 2014; Robeva et al., 2013; Tut et al., 1997). Association between polymorphism of the *AR* and various traits (infertility, disease, personality) has also been documented in a number of animal species (Konno et al., 2011; Lai et al., 2008; Lyons et al., 2014; Revay et al., 2012; Yasui et al., 2013).

Equidae are comprised of eight species, which all belong to a single genus, *Equus*. *Equus* contain four sub-genera, *Equus*, *Asinus*, *Dolichohippus*, and *Hippotigris* (Fig.1.1). Social structures could be reflected in genetic differences, and members of the genus *Equus* show two patterns of social organization (Rubenstein, 2011) (Fig. 1.1). In the sub-genus *Equus* and *Hippotigris*, a single breeding male is found in constant association with a fixed group of 1-6 unrelated females and their offspring; this is the forming of a harem. On the other hand, in the sub-genus *Asinus* and *Dolichohippus*, breeding males have territories, and females do not form stable groups. The diversity of these social systems is not consistent with genetic phylogeny (Fig. 1.1). In addition, humans succeeded only in the domestication of the horse in sub-genus *Equus* and donkey in sub-genus *Asinus*, which display different social systems. The other species, including wild asses and zebras, have been considered not suitable for domestication because of their untameable nature.

As genus *Equus* displays different social systems, it is a suitable system to investigate the function of personality- and sociality-related genes in different sub-genera. Because AR variation may influence aggressiveness along with other traits, they are among the ideal candidate genes to study the basis for social system and behavioural differences in equids, which in turn can help us to understand the functions and evolution of these genes. As a first step toward this approach, I analysed the locus *ARQ* in three zebra species for comparison with those of horses.

3.2 Material and methods

Samples

Samples were the same with chapter 2.2. Sampling and DNA extraction method was explained in chapter 2.2. All individuals were kept in captivity at zoos in Japan or the UK. Captive populations of Grevy's and Hartmann's zebra are closely managed under

endangered species breeding programmes therefore it is highly unlikely that the samples I obtained were from hybridised individuals.

PCR and Sequence

I amplified a 220 base pair region of exon1 in equine *AR* by polymerase chain reaction (PCR). I designed the primer pairs based on horse *AR* sequence information (GenBank: JN187443) (Revay et al., 2012) using Primer3 software (Forward primer: GAACAGCAGCCTTCACAACA, Reverse primer: CTGCCTCCCTCGCTCTCC). A 10 μl PCR reaction contained 20 ng DNA, 0.5 μM of each of the primers, 0.5 U LA Taq polymerase, GC buffer I (TaKaRa), and 400 μM of each dNTP. For fecal samples, instead of 20 ng DNA, 2 μl of extracted DNA solution and 0.1 μg of T4 Gene 32 Protein (Nippon Gene) were added. The PCR conditions consisted of an initial denaturation at 95 °C for 2 min, followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 62 °C for 30 s, extension at 72 °C for 1 min, and final extension at 72 °C for 10 min. In fecal samples, the PCR conditions consisted of an initial denaturation at 95 °C for 2 min, followed by 40-45 cycles of denaturation at 95 °C for 30 s, annealing at 62 °C for 30 s, extension at 72 °C for 1 min, and final extension at 72 °C for 10 min. PCR products were purified using a High Pure PCR Product Purification Kit (Roche), and their nucleotide sequences were determined by cycle sequencing using a Big Dye Terminator v3.1 Cycle Sequencing Kit (Roche) and an Applied Biosystems 3130xl Genetic Analyzer (Applied Biosystems). All determined sequences were checked visually using BioEdit software, ver.7.0.9.0. Polymorphisms in the sequences (SNPs and VNTR) were detected by the alignment using MEGA software, ver.6.0 (Tamura et al., 2013).

3. 3 Results

The nucleotide and amino acid sequence alignments are shown in Fig. 3.2a and 3.2b, respectively. In three zebra species, I found 5 lengths in CAG repeat regions, 8, 9, 10, 12, and 15 (Fig. 3.1). Allele frequencies in three zebra species are presented in Table 3.1. The data for these sequences were deposited in the DDBJ Genbank database under accession numbers LC030245-LC030253.

Grevy's zebra had two alleles; 8-repeat allele and 10-repeat allele. Plains zebra had three alleles; 9-repeat, 10-repeat, and 12-repeat allele. Mountain zebra had three alleles: 9-repeat, 10-repeat, and 15-repeat allele. In each zebra species, VNTRs were observed, where none have been reported in horses (Revay et al., 2012) (Figure 3.1). All alleles detected were longer than that of the horse (5 repeat). The longest allele (15 CAG repeats) had a deletion of AAA (Lysin) at 156bp (Figure 3.1).

There was only one SNP (G6C) observed among the three zebra species: this SNP was a non-synonymous substitution (V2L), and was found only in the 10-repeat allele in Grevy's zebra (Table 3.1). There was one fixed non-synonymous substitution observed between zebra and horse (nucleotide G36A; amino acid G12S) (Fig. 3.1).

3. 4. Discussion

In this study, I surveyed the allele frequencies and nucleotide sequence differences in CAG repeat regions of the *AR* in three zebra species and found significant differences to the existing horse reference data. In horses, VNTRs at the *ARQ* locus have not been reported and only a single allele including 5 CAG repeats has been previously described (Revay et al., 2012). My study provides the first information about *AR* in other equids.

Association between *ARQ* variation and aggressiveness has been demonstrated in humans (Vermeersch et al., 2010) and previous reports have shown that individuals with shorter CAG repeat in humans and dogs tend to be more aggressive (Jonsson et al., 2001;

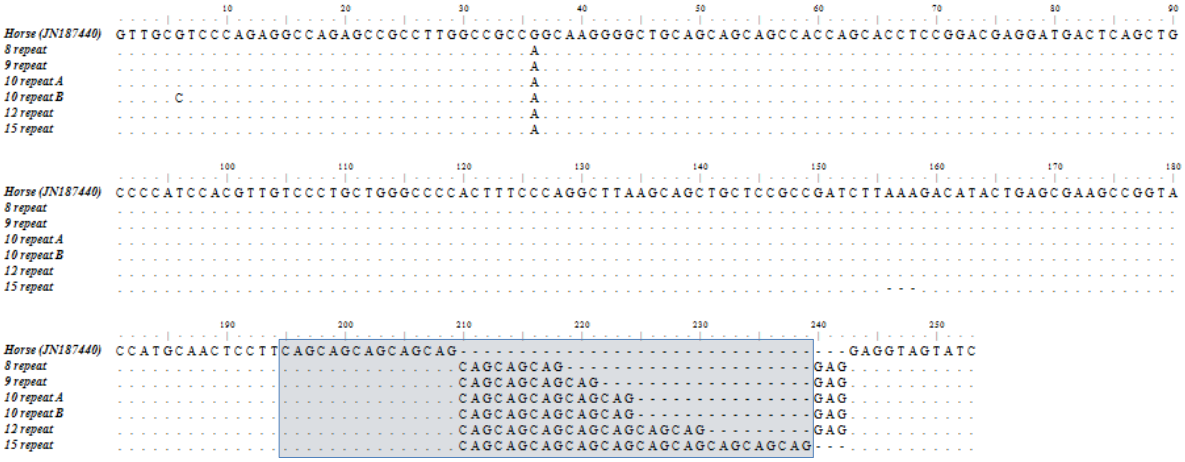
Konno et al., 2011). I found that locus *ARQ* in three zebra species had longer alleles than that in horses. As zebras have not been domesticated it might be expected that they are more aggressive than horses. The increased allele length observed in zebras in this study therefore does not correspond to the pattern of allele length and aggression observed in humans and dogs. Although this unexpected result remains to be explained, and there is a need for analysis of more samples in each species to increase confidence in the observed patterns, *in vitro* studies have shown that medium-length CAG repeats had high transcriptional activity of *AR* compared with both shorter and longer CAG repeats (Nenonen et al., 2010). Indeed, in the study that compared between two species (chimpanzee and bonobo), chimpanzees that had more aggressive traits had generally shorter alleles than bonobos (Garai et al., 2014). It would be interesting to investigate the association between androgen receptor activity and gene polymorphisms. Furthermore, it will be useful to conduct a standardised assessment of aggressiveness in zebra species and horse to compare personality in equids. About the relation to social system, the species which develop harem have shortest and longest alleles, and Grevy's zebra that develop territory have medium length allele, so the relation between genotypes and social systems is unclear. However, as *in vitro* studies (Nenonen et al., 2010), there is the case that medium-length repeat had highest activity, so it is required to investigate more samples in various equid species.

All samples used in this study were derived from captive populations, so might not reflect the number of alleles and allele frequencies of native populations, due to the founder effect and genetic drift. It is interesting to note that polymorphism was maintained in each zebra species, despite these populations being derived from a small number of founders, suggesting that diversity in natural populations could be higher. In contrast the lack of diversity previously observed in horses may relate to its long history of domestication.

This study could not identify if these polymorphisms occurred in the common ancestor of horse and zebra or were only generated in zebra species after they diverged. In the first case, the *AR* might have undergone positive selection in horse domestication. To clarify the derivation of these polymorphisms, I must investigate other *Equus* species. Recently, the gene groups that are associated with the domestication of the horse were reported (Schubert et al., 2014). One gene group is related to muscular and limb development, articular junctions, and cardiac systems. The second group relates to cognitive functions including social behaviour, learning capabilities, fear response, and agreeableness. Genes involved in tameness and social behaviour are important in domestication, so it is thought that *AR* contributing to the activity of testosterone might be significant in domestication. Moreover, the analysis of the relationship between the *AR* and reproductive ability may be useful to the conservation of endangered *Equus* species. Intra species polymorphisms that I found in zebra species will be valuable to investigate the relationship between genetic polymorphism and function (responsiveness and metabolism).

In this study, I found evidence for variation between zebra species and horse at locus *ARQ* that may relate to differences in aggression and social systems, although the number of samples in this study does limit interpretation of these findings. To understand the evolution of *AR* in equids, it would be useful to conduct further studies on the association of these traits with the polymorphisms that I found in zebra species and to compare the results with those of other equids.

a)



b)

Fig. 3.1. a) Sequences of the androgen receptor exon1 CAG repeat region in equid. Locations of VNTR are marked by squares. b) Sequences of amino acid of the androgen receptor exon1 CAG repeat region in equid. Locations of VNTR are marked by squares.

Table. 3.1 Allele frequencies of androgen receptor in three zebra species

		Grevy's zebra		Plains zebra		Mountain zebra		Horse	
		N	frequency	*N*	frequency	*N*	frequency	*N*	frequency
VNTR	*5*	0	0.000	0	0.000	0	0.000	25	1.000
(CAG)n	*8*	16	0.198	0	0.000	0	0.000	0	0.000
	9	0	0.000	12	0.364	1	0.333	0	0.000
	10	65	0.802	16	0.485	1	0.333	0	0.000
	12	0	0.000	5	0.152	0	0.000	0	0.000
	15	0	0.000	0	0.000	1	0.333	0	0.000
SNP	*G*	68	0.840	33	1.000	3	1.000	25	1.000
G325C	*C*	13	0.160	0	0.000	0	0.000	0	0.000

N: the number of X chromosomes

Chapter 4 General discussion and final remarks

Genetic diversity in population is required to adapt environmental changes to survive. The loss of genetic diversity leads to decreasing of adaptation of populations by reduction of reproductive rate and accumulating of deleterious gene. Especially in small population, genetic diversity tends to reduce more rapidly by bottleneck and genetic drift, and small population tends to be trapped by extinction vortex (Blomqvist et al., 2010; Fagan and Holmes, 2006; Frankham et al., 2010; Reed and Frankham, 2003). Therefore, to conserve the endangered species, developing appropriate genetic management plan is very important. Also, *ex situ* conservation (i.e. breeding in captivity) is important in the conservation of endangered species as well as *in situ* conservation, and the quantifying genetic variation is a basic element of endangered species breeding programmes in captivity. It is also necessary to plan a conservation programme that includes genetic management in endangered Grevy's zebra. However, there is little information on Grevy's zebra genetics before this study. Therefore the development of genetic markers that can evaluate the genetic diversity is essential.

In the present study, partial genome sequence data of Grevy's zebra that were obtained using NGS were screened for microsatellite repeats. I developed 28 novel microsatellite markers, and quantified the genetic diversity in captive Grevy's zebra populations which can be used as basis for genetic management plan using *de novo* microsatellite markers and mtDNA (*cytb* and control region). To understand the genetic diversity of a population or species, a combined analysis of nuclear DNA and mtDNA is more powerful (Kolleck, Yang et al. 2013). Moreover, I applied cross-species amplification in other two zebra species, and investigated the genetic structure of three zebra species in the captivity. Furthermore, to develop suitable keeping methods for each individual, I analyzed the polymorphism of *AR*e,

which is a candidate gene for personality trait, in three zebra species.

In the analysis of genetic diversity using *de novo* microsatellite markers, comparative diversity estimates indicated greater genetic variation in plains zebra and its subspecies than in Grevy's zebra, despite potential ascertainment bias. Also genetic diversity of mtDNA analyses was extremely low in Grevy's zebra. In view of the situations of these species in the wild (Grevy's zebra is endangered and plains zebra is abundant), this result may not be surprising. Genetic diversity of Grevy's zebra was lower than that of mountain zebra. Considering that the captive population of Grevy's zebra is kept in better condition than mountain zebra (Table 2.3.2), these results indicated the possibility that the low genetic diversity derived from founding population, and that wild population had low genetic diversity before 1970's. Recently, Kebede et al. (2014) reported that wild populations of Grevy's zebra in Ethiopia had only three haplotype in control region, and very low genetic diversity. This report supports my hypothesis. The goal of captive breeding is maintaining populations that are both demographically self-sustaining and genetically healthy, and most generally target is to retain 90% of genetic diversity during 100 years (Frankham et al., 2010; Lacy, 2013). There is a greater chance of accomplishing this goal in Grevy's zebra based on recent pedigree data (Langenhorst, 2014a). However, the understanding of kinship among individuals of founding population is important, because it greatly affects the genetic diversity in the population of descendants (Frankham et al., 2010). Although the relatedness among founders is imprecise as with other many species, this study showed living population have low genetic diversity and inferred founder population had little genetic diversity. Also, this study indicated the differences between genetic diversity in pedigree analysis and genetic diversity in molecular analysis genetic diversity, and it is important to monitor the possessing genetic diversity of population. Moreover given importance of captive population, more moderate breeding plan will be required. Despite the numbers of

founders of captive Grevy's zebra is larger (n = 123) than recommended numbers (at least n = 20-30) (Frankham et al., 2010), genetic diversity is very low. Discordancy between pedigree analysis and molecular analysis may derive from large numbers of founders and low genetic diversity of founders, the molecular genetic analysis is required to develop captive breeding plan for the other species with these characters. In the comparison of the genetic diversity of Grevy's zebra and the mountain zebra, although Grevy's zebra have low diversity in mtDNA, the genetic diversity in the microsatellite analysis in little differences between both species. These differences may derived from bias in development of microsatellite markers or social systems.

Novel microsatellite markers developed in this study will be useful to conserve for wild populations. Cumulative *PID* for all loci was very low (*PID* = 2.28 x 10^{-14} for unrelated individuals and *PID-sib* = 1.06 x 10^{-6} when siblings are involved), indicating the usefulness of these markers for individual discrimination. Recently, various investigations have been performed to conserve in the habitat (KWS, 2012; Langenhorst, 2014a). In these investigations, there are Digital Stripe Pattern Identification Database (Stripe ID) and tracking by collaring. Stripe ID is the method for the identification of individuals using the differences of stripe patterns among individuals, and has been used to estimate population size, determine inter-birth interval, and understand resource hot spots and to determine movement patterns. The collaring can provide the information of Grevy's zebra about use of the environment, selection of habitats and resource needs, although there is disadvantage of that analyzed samples are limited. These developed microsatellite loci have ability of identification in all individuals, they may be useful as complementary method. These markers may also serve as effective tools to understand the phylogeography, dispersal patterns and genetic structure of wild Grevy's zebra. Understanding of the population dynamics may help to monitor them for future conservation, microsatellite analysis of

groups in the wild will help us to understand the social structure. This information will be useful for the management of the Grevy's zebra in the captivity.

Hybrid populations between Grevy's and plains zebra were simulated to investigate subspecies and to detect hybridized individuals. Polymorphisms of these microsatellite markers were conserved across species (plains zebra and mountain zebra) with sufficient variation to enable individual identification in all populations. Species and subspecies differentiation were clearly demonstrated, and simulated F1 and F2 hybrids were correctly identified. Also, there are differences in mtDNA between both species, if using mtDNA analysis simultaneously, hybridized individuals are able to detect more exactly. These findings provide insights into captive population genetic diversity in zebras and support the use of these markers for identifying hybrids, including the known hybrid issue in the endangered Grevy's zebra (Chiyo and Alberts, 2009; Cordingley et al., 2009). Also, as management by the studbook has not done in the plains zebra, exact relatedness among individuals is unclear. In addition, hybrid between subspecies may exist because management as unit of subspecies is insufficient. Because these markers are useful for the identification of individual and the subspecies of the plains zebra, this study can contribute to resolve this problem.

AR has been found to have associations with reproductive development, behavioural traits, and disorders in humans. However, the influence of similar genetic effects on the behaviour of other animals is uncertain. The comparison of genes and their functions will provide information for evolution and welfare of zoo animals, but the investigation of relation between gene and functions is difficult for zoo animal. However, genus *Equus* may be better model, because there are species have various characters in the same genus and include domesticated animals. In the horse, which is related species, correlations between genes and personality have been documented (Hori et al., 2015; Momozawa et al., 2005). I

examined the loci the repeat of glutamin in *AR* in three zebra species, and compared them with those of domesticated horses. I observed polymorphism among zebra species and between zebra and horse. This is the first finding of the polymorphism of *AR* in zebra species, and in non-domesticated equid as far as I know. As androgens such as testosterone influence aggressiveness, *AR* polymorphism among equid species may be associated with differences in levels of aggression and tameness. My findings indicate that it would be useful to conduct further studies focusing on the potential association between *AR* and personality traits, and to understand domestication of equid species. Moreover, discovering relation between *AR* and phenotype may lead to develop keeping methods that moderate for each individual, and population management plans (such as developing bachelor group).

I developed the first microsatellite marker derived from the Grevy's zebra genome. These developed microsatellite markers have abilities enough to evaluate the genetic diversity and to infer the relatedness between individuals in the Grevy's zebra, but also in the other two zebra species (plains zebra and mountain zebra). Moreover, I found the inconsistency between pedigree analysis and molecular analysis, and indicated the importance of integrating both analysis in captive breeding plan for conservation of Grevy's zebra and other endangered species. However, I failed genotyping for many non-invasive samples. The improvement of genotyping in the samples by non-invasive methods is important to apply to the wild populations and more individuals in the captivity. Recently, it will be able to get massive sequence information due to the spread of using next generation sequencing and to develop the markers from massive sequence data. Among equid animals, only horse that is an economically important animal had been analyzed whole-genome (Wade et al., 2009). Although whole genome sequences have not been obtained in other equid, massive sequence information had been obtained using next generation sequencing (Jonsson et al., 2014; Orlando et al., 2013). In the Grevy's zebra, about 42 Gb

(Depth-of-coverage 17.05 x) has been analyzed (Jonsson et al., 2014). If methods can be developed, various genetic analysis for conservation can be performed. In genotyping of microsatellite markers, miss-typing ratio is lower in longer repeat unit than in short unit, but longer repeats unit tend to have wider size range (Gardner et al., 2011; Guichoux et al., 2011), so Guichoux et al. (2011) recommended that tetra-repeats unit is the first choice. Also, the success rate of PCR is in the short length than longer length amplification, therefore it is desirable to apply in the samples by non-invasive methods. It is most desirable to develop microsatellite markers that have 3-4 base repeated unit and short length amplified regions. Also, single nucleotide polymorphisms (SNP) markers can be developed, and restriction site associated DNA (RAD) sequencing can type and score tens to hundreds of thousands of SNP markers in hundreds of individuals, simultaneously (Etter et al., 2011; Hohenlohe et al., 2012). The development and analysis of massive markers make it possible to understand relation between gene and important traits for conservation such as reproductive ability, resistance/susceptibility of infectious disease, and longevity. Although genome wide analysis is difficult in endangered species due to small population size, detailed pedigree data with many remarks (such as history of disease, fertility and personality) can help to analyze. In the future, integration of these data will be useful for conservation of endangered Grevy's zebra.

5. References

Abdelkrim J, Robertson B, Stanton JA, Gemmell N. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. Biotechniques 46(3):185-192.

Avise J. 1987. Intraspecific Phylogeography: The Mitochondrial-DNA Bridge Between Population Genetics And Systematics. Annual Review of Ecology and Systematics 18(1):489-522.

Barilani M, Bernard-Laurent A, Mucci N, Tabarroni C, Kark S, Perez Garrido JA, Randi E. 2007. Hybridisation with introduced chukars (*Alectoris chukar*) threatens the gene pool integrity of native rock (*A. graeca*) and red-legged (*A. rufa*) partridge populations. Biological Conservation 137(1):57-69.

Barilani M, Sfougaris A, Giannakopoulos A, Mucci N, Tabarroni C, Randi E. 2006. Detecting introgressive hybridisation in rock partridge populations (*Alectoris graeca*) in Greece through Bayesian admixture analyses of multilocus genotypes. Conservation Genetics 8(2):343-354.

Blomqvist D, Pauliny A, Larsson M, Flodin LA. 2010. Trapped in the extinction vortex? Strong genetic effects in a declining vertebrate population. BMC Evolutionary Biology 10:33.

Carlstead K, Mellen J, Kleiman DG. 1999. Black rhinoceros (*Diceros bicornis*) in U.S. zoos: I. individual behavior profiles and their relationship to breeding success. Zoo Biology 18(1):17-34.

Chamberlain NL, Driver ED, Miesfeld RL. 1994. The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. Nucleic Acids Research 22(15):3181-3186.

Chang CS, Kokontis J, Liao ST. 1988. Structural analysis of complementary DNA and

amino acid sequences of human and rat androgen receptors. Proceedings of the National Academy of Sciences of the United States of America 85(19):7211-7215.

Chen JJZ, Hebert PDN. 1999. Intraindividual sequence diversity and a hierarchical approach to the study of mitochondrial DNA mutations. Mutation research DNA repair 434(3):205-217.

Chiyo PI, Alberts SC. 2009. An assessment of the threat to Grevy's zebra from hybridization. Animal Conservation 12(6):514-515.

Choong CS, Wilson EM. 1998. Trinucleotide repeats in the human androgen receptor: a molecular basis for disease. Journal of Molecular Endocrinology 21(3):235-257.

Collaer ML, Hines M. 1995. Human behavioral sex differences: a role for gonadal hormones during early development? Psychological Bulletin 118(1):55-107.

Cordingley JE, Sundaresan SR, Fischhoff IR, Shapiro B, Ruskey J, Rubenstein DI. 2009. Is the endangered Grevy's zebra threatened by hybridization? Animal Conservation 12(6):505-513.

Dieringer D, Schlotterer C. 2003. Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. Genome Research 13(10):2242-2251.

Ducroz JF, Volobouev V, Granjon L. 1998. A molecular perspective on the systematics and evolution of the genus *Arvicanthis* (Rodentia, Muridae): inferences from complete cytochrome b gene sequences. Molecular Phylogenetics and Evolution 10(1):104-117.

Earl D, vonHoldt B. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4(2):359-361.

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA. 2011. SNP discovery and

genotyping for evolutionary genetics using RAD sequencing. Methods in Molecular Biology 772:157-178.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14(8):2611-2620.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10(3):564-567.

Fagan WF, Holmes EE. 2006. Quantifying the extinction vortex. Ecology Letters 9(1):51-60.

Faircloth BC. 2008. msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. Molecular Ecology Resources 8(1):92-94.

Ferrie GM, Cohen OR, Schutz P, Leighty KA, Plasse C, Bettinger TL, Hoffman EA. 2013. Identifying parentage using molecular markers: improving accuracy of studbook records for a captive flock of marabou storks (*Leptoptilos crumeniferus*). Zoo Biology 32(5):556-564.

Frankham R, Ballou JD, Briscoe DA. 2010. Introduction to conservation genetics. Cambridge, New York: Cambridge University Press.

Garai C, Furuichi T, Kawamoto Y, Ryu H, Inoue-Murayama M. 2014. Androgen receptor and monoamine oxidase polymorphism in wild bonobos. Meta Gene 2(0):831-843.

Gardner MG, Fitch AJ, Bertozzi T, Lowe AJ. 2011. Rise of the machines--recommendations for ecologists when using next generation sequencing for microsatellite development. Molecular Ecology Resources 11(6):1093-1101.

Gautschi B, Muller JP, Schmid B, Shykoff JA. 2003. Effective number of breeders and maintenance of genetic diversity in the captive bearded vulture population. Heredity 91(1):9-16.

Giovannucci E, Stampfer MJ, Krithivas K, Brown M, Brufsky A, Talcott J, Hennekens CH,

Kantoff PW. 1997. The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. Proceedings of the National Academy of Sciences of the United States of America 94(7):3320-3323.

Guichoux E, Lagache L, Wagner S, Chaumeil P, LÉGer P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, Petit RJ. 2011. Current trends in microsatellite genotyping. Molecular Ecology Resources 11(4):591-611.

Hack MA, Lorenzen E. 2008. *Equus quagga*. The IUCN Red List of Threatened Species. Version 2015.2 ed: International Union for Conservation of Nature.

Henkel JR, Jones KL, Hereford SG, Savoie ML, Leibo SP, Howard JJ. 2012. Integrating microsatellite and pedigree analyses to facilitate the captive management of the endangered Mississippi sandhill crane (*Grus canadensis pulla*). Zoo Biology 31(3):322-335.

Hohenlohe PA, Catchen J, Cresko WA. 2012. Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. Methods in Molecular Biology 888:235-260.

Hori Y, Tozaki T, Nambo Y, Sato F, Ishimaru M, Inoue-Murayama M, Fujita K. 2015. Evidence for the effect of serotonin receptor 1A gene (*HTR1A*) polymorphism on tractability in Thoroughbred horses. Animal Genetics: in press

Ito H, Hayano A, Langenhorst T, Sakamoto H, Inoue-Murayama M. 2013. Using next generation sequencing to develop microsatellite markers for the endangered Grevy's zebra (*Equus grevyi*). Conservation Genetics Resources 5(2):507-510.

Ivy JA, Miller A, Lacy RC, Dewoody JA. 2009. Methods and prospects for using molecular data in captive breeding programs: an empirical example using parma wallabies (*Macropus parma*). Journal of Heredity 100(4):441-454.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program

for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14):1801-1806.

Jonsson EG, von Gertten C, Gustavsson JP, Yuan QP, Lindblad-Toh K, Forslund K, Rylander G, Mattila-Evenden M, Asberg M, Schalling M. 2001. Androgen receptor trinucleotide repeat polymorphism and personality traits. Psychiatric Genetics 11(1):19-23.

Jonsson H, Schubert M, Seguin-Orlando A, Ginolhac A, Petersen L, Fumagalli M, Albrechtsen A, Petersen B, Korneliussen TS, Vilstrup JT, Lear T, Myka JL, Lundquist J, Miller DC, Alfarhan AH, Alquraishi SA, Al-Rasheid KA, Stagegaard J, Strauss G, Bertelsen MF, Sicheritz-Ponten T, Antczak DF, Bailey E, Nielsen R, Willerslev E, Orlando L. 2014. Speciation with gene flow in equids despite extensive chromosomal plasticity. Proceedings of the National Academy of Sciences of the United States of America 111(52):18655-18660.

Kalinowski ST. 2005. hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes 5(1):187-189.

Kebede F, Rosenbom S, Khalatbari L, Moehlman PD, Beja-Pereira A, Bekele A. 2014. Genetic diversity of the Ethiopian Grevy's zebra (*Equus grevyi*) populations that includes a unique population of the Alledeghi Plain. Mitochondrial DNA.

Khosravi R, Rezaei HR, Kaboli M. 2013. Detecting hybridization between Iranian wild wolf (*Canis lupus pallipes*) and free-ranging domestic dog (*Canis familiaris*) by analysis of microsatellite markers. Zoological Science 30(1):27-34.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proceedings of the National Academy of Sciences of the United States of America 86(16):6196-6200.

Konno A, Inoue-Murayama M, Hasegawa T. 2011. Androgen receptor gene polymorphisms are associated with aggression in Japanese Akita Inu. Biology Letters 7(5):658-660.

Kuhar CW, Stoinski TS, Lukas KE, Maple TL. 2006. Gorilla Behavior Index revisited: Age, housing and behavior. Applied Animal Behaviour Science 96(3):315-326.

KWS. 2012. Conservation and Management Strategy for Grevy's zebra (*Equus grevyi*) in Kenya (2012-2016). 2 ed. Nairobi, Kenya: Kenya Wildlife Service. 40 p.

Lacy RC. 2013. Achieving True Sustainability of Zoo Populations. Zoo Biology 32(1):19-26.

Lacy RC, Ballou JD, Pollak JP. 2012. PMx: software package for demographic and genetic analysis and management of pedigreed populations. Methods in Ecology and Evolution 3(2):433-437.

Lai CL, L'Eplattenier H, van den Ham R, Verseijden F, Jagtenberg A, Mol JA, Teske E. 2008. Androgen receptor CAG repeat polymorphisms in canine prostate cancer. Journal of Veterinary Internal Medicine 22(6):1380-1384.

Langenhorst T. 2014a. International studbook for Grevy's zebra *Equus grevyi* 2013 until 31. 12. 2013. Hampshire, UK: Marwell Wildlife.

Langenhorst T. 2014b. International studbook for Hartman's mountain zebra *Equus zebra hartmannae* 2013 until 31. 12. 2013. Hampshire, UK: Marwell Wildlife.

Larizza A, Pesole G, Reyes A, Sbisa E, Saccone C. 2002. Lineage specificity of the evolutionary dynamics of the mtDNA D-loop region in rodents. Journal of Molecular Evolution 54(2):145-155.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25(11):1451-1452.

Lorenzen ED, Arctander P, Siegismund HR. 2008. High variation and very low differentiation in wide ranging plains zebra (*Equus quagga*): insights from mtDNA and microsatellites. Molecular Ecology 17(12):2812-2824.

Lyons RE, Loan NT, Dierens L, Fortes MR, Kelly M, McWilliam SS, Li Y, Bunch RJ, Harrison BE, Barendse W, Lehnert SA, Moore SS. 2014. Evidence for positive selection of taurine genes within a QTL region on chromosome X associated with testicular size in Australian Brahman cattle. BMC Genetics 15:6.

Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schutz G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P, Evans RM. 1995. The nuclear receptor superfamily: the second decade. Cell 83(6):835-839.

McGreevy TJ, Jr., Dabek L, Husband TP. 2011. Genetic evaluation of the Association of Zoos and Aquariums Matschie's tree kangaroo (*Dendrolagus matschiei*) captive breeding program. Zoo Biology 30(6):636-646.

McPhaul MJ. 2002a. Androgen receptor mutations and androgen insensitivity. Molecular and Cellular Endocrinology 198(1-2):61-67.

McPhaul MJ. 2002b. Molecular defects of the androgen receptor. Recent Progress in Hormone Research 57:181-194.

Meglecz E, Neve G, Biffin E, Gardner MG. 2012. Breakdown of phylogenetic signal: a survey of microsatellite densities in 454 shotgun sequences from 154 non model eukaryote species. PloS One 7(7):e40861.

Moehlman PD, Rubenstein DI, Kebede F. 2013. *Equus grevyi*. The IUCN Red List of Threatened Species. Version 2015.2 ed: International Union for Conservation of Nature.

Momozawa Y, Takeuchi Y, Kusunose R, Kikusui T, Mori Y. 2005. Association between equine temperament and polymorphisms in dopamine D4 receptor gene. Mammalian Genome 16(7):538-544.

Moodley Y, Harley EH. 2006. Population structuring in mountain zebras (*Equus zebra*): The molecular consequences of divergent demographic histories. Conservation Genetics

6(6):953-968.

Mouritsen A, Hagen CP, Sorensen K, Aksglaede L, Mieritz MG, Main KM, Almstrup K, Rajpert-De Meyts E, Juul A. 2013. Androgen receptor CAG repeat length is associated with body fat and serum SHBG in boys: a prospective cohort study. Journal of Clinical Endocrinology and Metabolism 98(3):E605-609.

Negri A, Pellegrino I, Mucci N, Randi E, Tizzani P, Meneguz PG, Malacarne G. 2012. Mitochondrial DNA and microsatellite markers evidence a different pattern of hybridization in red-legged partridge (*Alectoris rufa*) populations from NW Italy. European Journal of Wildlife Research 59(3):407-419.

Nenonen H, Bjork C, Skjaerpe PA, Giwercman A, Rylander L, Svartberg J, Giwercman YL. 2010. CAG repeat number is not inversely associated with androgen receptor activity in vitro. Molecular Human Reproduction 16(3):153-157.

Nielsen EE, Bach LA, Kotlicki P. 2006. hybridlab (version 1.0): a program for generating simulated hybrids from population samples. Molecular Ecology Notes 6(4):971-973.

Novellie P. 2008. *Equus zebra*. The IUCN Red List of Threatened Species. Version 2015.2 ed: International Union for Conservation of Nature.

Ogden R, Langenhorst T, McEwing R, Woodfine T. 2007. Genetic markers and sample types for pedigree reconstruction in Grevy's zebra (*Equus grevyi*). Der Zoologische Garten 77(1):29-35.

Olsen NJ, Benko AL, Kovacs WJ. 2014. Variation in the androgen receptor gene exon 1 CAG repeat correlates with manifestations of autoimmunity in women with lupus. Endocrine Connections 3(2):99-109.

Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, Johnson PL, Fumagalli M, Vilstrup JT, Raghavan M, Korneliussen T, Malaspinas AS, Vogt J, Szklarczyk D, Kelstrup CD, Vinther J,

Dolocan A, Stenderup J, Velazquez AM, Cahill J, Rasmussen M, Wang X, Min J, Zazula GD, Seguin-Orlando A, Mortensen C, Magnussen K, Thompson JF, Weinstock J, Gregersen K, Roed KH, Eisenmann V, Rubin CJ, Miller DC, Antczak DF, Bertelsen MF, Brunak S, Al-Rasheid KA, Ryder O, Andersson L, Mundy J, Krogh A, Gilbert MT, Kjaer K, Sicheritz-Ponten T, Jensen LJ, Olsen JV, Hofreiter M, Nielsen R, Shapiro B, Wang J, Willerslev E. 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature 499(7456):74-78.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics 28(19):2537-2539.

Petit RJ, El Mousadik A, Pons O. 1998. Identifying Populations for Conservation on the Basis of Genetic Markers. Conservation Biology 12(4):844-855.

Posada D, Crandall KA. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends in Ecology & Evolution 16(1):37-45.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155(2):945-959.

Ramadan S. 2012. Evaluation of genetic diversity and conservation priorities for Egyptian chickens. Open Journal of Animal Sciences 02(03):183-190.

Randi E. 2008. Detecting hybridization between wild species and their domesticated relatives. Molecular Ecology 17(1):285-293.

Reed DH, Frankham R. 2003. Correlation between Fitness and Genetic Diversity. Conservation Biology 17(1):230-237.

Revay T, Villagomez DA, Brewer D, Chenier T, King WA. 2012. GTG mutation in the start codon of the androgen receptor gene in a family of horses with 64,XY disorder of sex development. Sexual Development 6(1-3):108-116.

Robeva R, Tanev D, Andonova S, Kirilov G, Savov A, Stoycheva M, Tomova A, Kumanov P, Rashkov R, Kolarov Z. 2013. Androgen receptor (CAG)n polymorphism and androgen levels in women with systemic lupus erythematosus and healthy controls. Rheumatology International 33(8):2031-2038.

Rosenberg NA. 2004. distruct: a program for the graphical display of population structure. Molecular Ecology Notes 4(1):137-138.

Rousset F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources 8(1):103-106.

Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology 132:365-386.

Rubenstein DI. 2011. Family equidae (Horses and relatives). In: Wilson DE, Mittermeier RA, editors. Handbook of Mammals of the World. Spain: Lynx Edicions. p 106-143.

Santana Q, Coetzee M, Steenkamp E, Mlonyeni O, Hammond G, Wingfield M, Wingfield B. 2009. Microsatellite discovery by deep sequencing of enriched genomic libraries. Biotechniques 46(3):217-223.

Sanz N, Araguas RM, Fernández R, Vera M, García-Marín J-L. 2008. Efficiency of markers and methods for detecting hybrids and introgression in stocked populations. Conservation Genetics 10(1):225-236.

Scandura M, Iacolina L, Apollonio M, Dessì-Fulgheri F, Baratti M. 2009. Current status of the Sardinian partridge (*Alectoris barbara*) assessed by molecular markers. European Journal of Wildlife Research 56(1):33-42.

Schubert M, Jonsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A, Albrechtsen A, Dupanloup I, Foucal A, Petersen B, Fumagalli M, Raghavan M, Seguin-Orlando A, Korneliussen TS, Velazquez AM, Stenderup J, Hoover CA, Rubin CJ, Alfarhan AH, Alquraishi SA, Al-Rasheid KA, MacHugh DE, Kalbfleisch T, MacLeod JN, Rubin

EM, Sicheritz-Ponten T, Andersson L, Hofreiter M, Marques-Bonet T, Gilbert MT, Nielsen R, Excoffier L, Willerslev E, Shapiro B, Orlando L. 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proceedings of the National Academy of Sciences of the United States of America.

Steiner CC, Ryder OA. 2011. Molecular phylogeny and evolution of the Perissodactyla. Zoological Journal of the Linnean Society 163(4):1289-1303.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30(12):2725-2729.

Templeton AR, Crandall KA, Sing CF. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132(2):619-633.

Toth G, Gaspari Z, Jurka J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Research 10(7):967-981.

Tut TG, Ghadessy FJ, Trifiro MA, Pinsky L, Yong EL. 1997. Long polyglutamine tracts in the androgen receptor are associated with reduced trans-activation, impaired sperm production, and male infertility. Journal of Clinical Endocrinology and Metabolism 82(11):3777-3782.

van Oers K, Mueller JC. 2010. Evolutionary genomics of animal personality. Philosophical transactions Biological sciences 365(1560):3991-4000.

Vermeersch H, T'Sjoen G, Kaufman JM, Vincke J, Van Houtte M. 2010. Testosterone, androgen receptor gene CAG repeat length, mood and behaviour in adolescent males. European Journal of Endocrinology of the European Federation of Endocrine Societies 163(2):319-328.

Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL,

Bailey E, Bellone RR, Blocker H, Distl O, Edgar RC, Garber M, Leeb T, Mauceli E, MacLeod JN, Penedo MC, Raison JM, Sharpe T, Vogel J, Andersson L, Antczak DF, Biagi T, Binns MM, Chowdhary BP, Coleman SJ, Della Valle G, Fryc S, Guerin G, Hasegawa T, Hill EW, Jurka J, Kiialainen A, Lindgren G, Liu J, Magnani E, Mickelson JR, Murray J, Nergadze SG, Onofrio R, Pedroni S, Piras MF, Raudsepp T, Rocchi M, Roed KH, Ryder OA, Searle S, Skow L, Swinburne JE, Syvanen AC, Tozaki T, Valberg SJ, Vaudin M, White JR, Zody MC, Broad Institute Genome Sequencing P, Broad Institute Whole Genome Assembly T, Lander ES, Lindblad-Toh K. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 326(5954):865-867.

Wielebnowski NC. 1999. Behavioral differences as predictors of breeding status in captive cheetahs. Zoo Biology 18(4):335-349.

Witzenberger KA, Hochkirch A. 2011. Ex situ conservation genetics: a review of molecular studies on the genetic consequences of captive breeding programmes for endangered animal species. Biodiversity and Conservation 20(9):1843-1861.

Wyce A, Bai Y, Nagpal S, Thompson CC. 2010. Research Resource: The Androgen Receptor Modulates Expression of Genes with Critical Roles in Muscle Development and Function. Molecular Endocrinology 24(8):1665-1674.

Yasui S, Konno A, Tanaka M, Idani G, Ludwig A, Lieckfeldt D, Inoue-Murayama M. 2013. Personality assessment and its association with genetic factors in captive Asian and African elephants. Zoo Biology 32(1):70-78.

Zitzmann M, Nieschlag E. 2003. The CAG repeat polymorphism within the androgen receptor gene and maleness. International Journal of Andrology 26(2):76-83.

6. ACKNOWLEDGMENTS

My heartfelt thanks go to my supervisor Miho Inoue-Murayama for her tremendous support. Her concern and support included not only academics but my general wellbeing of which I am profoundly grateful. I appreciate to Tanya Langenhorst (Marwell Wildlife) and Rob Ogden for their kindly supports, collaborations and suggestions. I would also like to appreciate the following teaching staff of Wildlife Research Center for their guidance; Shiro Koshima, Gen'ichi Idani, Satoshi Hirata, Hideki Sugiura, Michio Nakamura and Hideaki Abe. I thank the administrative staff members including Tomoko Kurihara, Hiraku Yamamoto, Tokunori Nanjo, Shigeki Fukugaki, Sawako Takahashi and Miki Sugitani for their administrative support.

I would like to sincerely acknowledge the Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT), Asia and Africa Science Platform Program under the Japan Society for the Promotion of Science (JSPS), the Global Center of Excellence of Kyoto University and Environment Research and Technology Development Fund for providing funds and resources for this study.

Furthermore, I would like to express my appreciation to Azusa Hayano, Miyuki Sato, Hiromi Kobayashi and Yuki Makimura (Kyoto University) for their dedicated work on maintenance of laboratory equipment and data management. My sincere thanks also go to all the graduate students and postdoctoral fellows in the Wildlife Research Center for the helpful suggestions and discussion. I would like to thank all members of onigashima including Eiji Inoue, Azusa Hayano, Akitsugu Konno, Hideaki Abe, Coline Arnaud, Kodzue Kinoshita, Sherif Ramadan, Christopher Adenyo, Etienne-François Okoue, Augustin Basabose, Chihiro Hiramatsu, Yusuke Hori, Minori Arahori, Akane Ishihara, Yoshinori Endo and Yu Sato for their useful comments on my study during our weekly meetings.

My research for this thesis was not complete without cooperation of staffs, especially Mr.

Hidefusa Sakamoto in Kyoto City Zoo. I thank following zoos for the providing samples; Kyoto City Zoo, Ishikawa Zoo, Himeji Central Park, Toyama Family Park, Hamamatsu City Zoo, Morioka Zoological Park, Noichi Zoo, Wanpa-ku Zoo, Nogeyama Zoo, Kushiro Zoo, Hirakawa Zoo, Izu Animal Kingdom, Kawasaki Zoo, Toyohashi Zoo, Tama Zoological Park, Tobe Zoo, Hirakawa Zoo in Japan, Edingburg Zoo, Marwell Wildlife, Whipsnade Zoo, Port Lympne Zoo, Woburn Safari Park in United Kingdom.

.

7. PUBLICATIONS ASSOCIATED WITH THIS THESIS

This thesis consists of three parts of which two have been published in peer-reviewed scientific journals and the latter part is now in preparation to be submitted soon. Although I was assisted by co-authors, I was the coordinator and main contributor to all parts of this scientific research, including laboratory works, data analysis and manuscript writing. The publications associated with this thesis are the following:

Chapter 2

Hideyuki Ito, Azusa Hayano, Tanya Langenhorst, Hidefusa Sakamoto, Miho Inoue-Murayama: Using next generation sequencing to develop microsatellite markers for the endangered Grevy's zebra (*Equus grevyi*). Conservation Genetics Resources 5(2), 507－510, 2013

Hideyuki Ito, Tanya Langenhorst, Rob Ogden, Miho Inoue-Murayama: Population genetic diversity and hybrid detection in captive zebras. Scientific Reports 5, Article number: 13171, 2015

Hideyuki Ito, Tanya Langenhorst, Rob Ogden, Miho Inoue-Murayama: Mitochondrial DNA sequence diversity of Grevy's zebra (Equus grevyi) for management in captivity. Manuscript in preparation

Chapter 3

Hideyuki Ito, Tanya Langenhorst, Rob Ogden, Miho Inoue-Murayama: Androgen receptor gene polymorphism in zebra species. Meta Gene 5, 120－123, 2015

8. LIST OF FIGURES

9. LIST OF TABLES