DISSERTATION

Development of a method for wild boar (*Sus scrofa*) population size estimation by genotyping of non-invasive samples

▪

Entwicklung einer Methode zur Populationsschätzung von Wildschweinen (*Sus* scrofa) mittels Genotypisierung nicht-invasiv gewonnener Proben

Zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften
Fachbereich 7: Natur- und Umweltwissenschaften
Universität Koblenz-Landau
Campus Landau
Vorgelegt

am: 20.07.2012

Karolina Kolodziej
geboren am 31.12.1980 in Oppeln / Polen

Dean:	**Prof. Dr. Ralf Schulz**
First referee:	**Prof. Dr. Ralf Schulz**
Second referee:	**Prof. Dr. Thorsten Stoeck**
	(Technical University of Kaiserslautern)
Date of thesis` defence:	

Acknowledgments

After an exciting period of hard working on this thesis with a lot of ups and downs it came to an end. It would not be possible without the contribution, help, support and influence of many people, whom I like to express my very deep thanks.

First of all I would like to thank my first referee Prof. Dr. Ralf Schulz, when it came to the crunch I could always count on him.

I also thank Prof. Dr. Thorsten Stoeck for taking time to act as second referee.

I owe an enormous thank to my advisors Dr. Kathrin Theißinger and Dr. Holger Schulz. Kathrin was a big help by my manuscripts, I could learn a lot from her about the challenge of scientific writing. Holger attended me through this thesis with an abundance of patience and encouragment, especially during the downs.

A grateful thank also to Therese Bürgi for assisting me in the lab and for a lot of technical advice. I miss our time in Geilweilerhof with the cooking sessions during lunch, planted tomatoes, Swiss cheese fondue and chatting.

Many thanks go to my diploma students Christian Thometzek, Sebastian Eckert and Katja Seltmann as well as my case study students and trainees Susanne Gramlich, Christian Wolf, Stefanie Holzhäuser und Svenja Müßigbrodt for a great support and help with their work.

I thank Anne Schrimpf for being a perfect office mate during my writing time, for the coffee breaks and our friendship.

I am grateful to Jone Kammerer for the administrative support and all colleagues from the working group of Prof. Dr. Ralf Schulz and Prof. Dr. Klaus Schwenk for scientific discussions, meetings in Brennan´s, cooking events, barbecues, experiences like organizing a conference and mainly for the nice atmosphere during the last five years.

Big thanks go to the Lotto Foundation Rheinland-Palatinate for my financial support and to the „Rheinland-Pfalz für Innovation“, and the Ministry for Environment and Forestry for the financial founding of this project.

Furthermore I would like to thank my Co-operation partners of this project Cornelia Ebert, Dr. Ulf Hohmann, Ditmar Huckschlag and Dr. Jörg Brün for providing ideas, the monthly meetings, helpful discussions, the big fun during our conference trips and the collection of samples with the help of many students and forest workers.

Special thank goes to my family, they always supported me in every single decision taken in my life. Also to all my friends, especially Kristina and Jenny, thank you for the time behind the thesis, for the fun we´ve had together so far, for listening and loyalty.

And finally I would like to express my most precious thanks to Rajko Heydenreich, my loving spouse and best friend. Thank you for immersing me in the world of statistics, for correction reading, for supporting me, for being the love of my life, for comforting me, for building me up and in particular for the imperturbable belief in me.

Contents

Abstract

Wild boars belong to the most wide spread ungulates in the world. They are characterized by a well performed adaption to their environment mainly due to their omnivorous dietary. The wild boar population in Germany increased during the past three decades. Nowadays their high density leads to problems in agricultural areas due to damage of crops and plays a significant role as disease vector as the classical swine fever. For an effective population management population size information is of crucial importance.

Different traditional methods exist to estimate population sizes as direct sightnings, faecal drop counts or hunting harvest which provide only relative estimates and population trends. Absolute population sizes could be yielded by a Capture-Mark-Recapture (CMR) approach. However, capturing of wild boars is difficult to realize and costly in terms of personnel and field effort. Furthermore the capture probabilities are heterogeneous due to the variable behaviour of individuals influenced by age, sex, and experience of the animals. Non-invasive genetic methods are a promising complement to the traditional methods for population size estimation particularly for wild boar. These methods reduce stress and capture bias and increase the number of re-captures. Faeces proved to be a suitable DNA source for wild boar genotyping, due to almost equal capture probability. However working with faeces implicates difficulties such as low DNA quality and quantity, genotyping errors as dropout and false alleles.

The main aim of the present study was to develop a reliable, cost-efficient, reproducible and practicable method for wild boar genotyping. This method should provide a reliable dataset of genotypes obtained from the collected faeces samples.

Individual identification forms the basis for an improved mark-recapture approach. As there is no sound method for absolute population counts in free living wild boar, reference values for the validation of this new approach are missing. Therefore, different routines to reduce and to assess genotyping errors were compared within this thesis. For maximum amplification rate, the storage, the extraction methods and the PCR-procedure were optimised. A step by step procedure was evaluated in order to determine the minimum required microsatellite (MS) number for reliable individual identification including a test with family groups (female and embryo tissue) to distinguish even between close relatives. A multiple-tubes approach, post-amplification checking and different correction procedures were applied to reduce genotyping errors. In order to quantify real genotyping error rates (GER) of datasets derived from sampling in the Palatinate Forest in western Germany, different methods for GER determination were compared with each other, obtaining GERs between 0% and 57.5%. As a consequence, more strict criteria for the multi-tube approach and increased repetition number of homozygous samples were used. An additional method validation was the implementation of a blind test to achieve the reliability of the genotyping and error checking procedure. Finally a strict and practicable proposal for the lab procedure was developed, by beginning with faecal sample collection and ending with a reliable dataset with genotypes of each sample.

The results of the presented method were derived from two sampling periods in a 4000 ha area in the Palatinate Forest in Rhineland-Palatinate in December 2006 and 2007. Both provided high confidence intervals (CI) applying inaccurate estimates (eg. for 2006 population size amounted to 215 with CI 95% of 156-314 and for 2007 population size amounted to 415 with CI 95% of 318-561) due to low sampling sizes (for 2006 n = 141 and for 2007 n = 326), successfully analysed samples (for 2006 n = 89 and for 2007 n = 156) and recapture numbers (for 2006 n = 12 and for 2007 n = 24). Furthermore, the population estimates even for the lowest values were considerably higher than previously assumed by hunting statistics, which implicates an ineffective hunting regime in the study area. For the future prospect, to obtain more precise population size estimations the increase of sampling sizes is inevitable, because absolute and reliable estimates are highly desirable for wildlife management and the control of diseases transmission. Nevertheless, the method for individual genotyping of wild boars evaluated in this thesis could be successfully established resulting in reliable datasets for population estimation modelling with sufficiently low GER.

Zusammenfassung

Wildschweine gehören zu den meist verbreiteten Huftieren der Welt. Sie charakterisieren sich durch eine sehr gute Anpassung an ihre Umwelt, die hauptsächlich auf ihre omnivore Nahrungsaufnahme zurückzuführen ist. Die Wildschweinpopulation in Deutschland stieg in den letzten drei Jahrzehnten deutlich an. Heutzutage führt deren hohe Dichte zu erheblichen Schäden auf landwirtschaftlichen Flächen aufgrund von Zerstörung der Ernte und spielt eine wichtige Rolle bei der Ausbreitung von Krankheiten, wie der klassischen Schweinepest. Für ein effektives Wildschwein Management sind Informationen über absolute Populationszahlen daher von höchster Bedeutung. Es existieren verschiedene traditionelle Methoden wie z.B. die direkte Beobachtung der Tiere, das Zählen ihrer Losungen oder die Auswertung der Statistiken zu Jagderträgen, die nur relative Schätzungen oder Populationstrends liefern. Absolute Populationszahlen könnte der Fang-Markier-Wiederfang Ansatz hervorbringen. Nichtsdestotrotz ist das Fangen von Wildschweinen schwer zu realisieren und kostenaufwendig in Bezug auf die hierzu benötigten Arbeitskräfte und den Feldarbeitsaufwand. Weiterhin sind die Fangwahrscheinlichkeiten heterogen aufgrund der Verhaltensvariabilitaet der Individuen, die durch Alter, Geschlecht und Erfahrung der Tiere bedingt ist. Nicht-invasive genetische Verfahren sind vielversprechender gegenüber den traditionellen Methoden der Populationsgrößenschätzung speziell für Wildschweine. Denn zum Einen reduzieren diese Methoden den Stressfaktor der Tiere und senken Fangfehler und zum Anderen erhöhen sie die Zahl indirekter Beobachtungen. Kot eignet sich sehr

gut als DNS-Quelle für die Wildschwein Genotypisierung, da dieser eine nahezu gleiche Erfassungswahrscheinlichkeit für alle Tiere gewährleistet. Nichtsdestotrotz bringt das Arbeiten mit Kot Schwierigkeiten mit sich, die sich in in einer geringen DNS Qualität und Quantität oder Genotypisierungsfehlern äußern.

Das Hauptziel der vorliegenden Arbeit war die Entwicklung einer zuverlässigen, kostengünstigen, reproduzierbaren und praktikablen Methode zur Genotypisierung von Wildschweinen. Diese Methode sollte einen zuverlässigen Genotypendatensatz liefern, der aus der Genotypisierung gesammelter Kotproben stammte.

Individuelle Identifikation bildet die Basis für einen anzuwendenden Fang-Markierung-Wiederfang Ansatz. Da es bisher keine vergleichbaren Untersuchungen zu Populationsschätzungen frei lebender Wildscheine gibt, fehlen in dieser Hinsicht auch Referenzdaten zum nicht-invasiven Ansatz. Daher wurden verschiedene Versuche zur Reduzierung und Quantifizierung der Genotypisierungsfehlerraten (GFR) getestet, verglichen und evaluiert. Um die Amplifizierungsrate zu erhöhen wurden Hälterungs- und DNS-Isolationsverfahren sowie PCR Protokolle optimiert. Ein Schritt für Schritt Ansatz zur Bestimmung der minimal erforderlichen Anzahl von Mikrosatelliten Marker wurde entwickelt, welcher einen Test mit nahverwandten Individuen (Mütter und deren Föten) beinhaltete, um sogar diese voneinander zu unterscheiden. Ein so genanntes Multitube-Verfahren und diverse Korrekturverfahren wurden angewendet um die GFR zu reduzieren. Die Quantifizierung von GFR aus erhobenen Datensätzen von zwei Beprobungen im Pfälzerwald wurde evaluiert indem verschiedene Methoden zur GFR-Bestimmung getestet und miteinander verglichen wurden, hierbei

ergaben sich GFR zwischen 0% bis 54%. Als Konsequenz dessen wurden die Kriterien für das Multitube-Verfahren verschärft, indem die Anzahl der Wiederholungen von homozygoten Proben erhöht wurde. Eine zusätzliche Validierung in Form eines Blindtests wurde etabliert, um die Zuverlässigkeit der Genotypisierung und Fehlerkorrekturen zu bekräftigen. Abschließend wurde ein strikter und praktikabler Verfahrenvorschlag entwickelt, beginnend beim Sammeln der Kotproben und endend mit dem Erhalt eines zuverlässigen Datensatzes mit Genotypen einzelner Proben.

Die Ergebnisse der hier präsentierten Methode aus zwei Beprobungen 2006 und 2007 in einem 4000 ha großen Areal im Pfälzer Wald führte zu ungenauen Schätzungen mit hohen Konfidenzintervallen. So lag die geschätzte Populationsgrösse in der Beprobung 2006 bei 215 Individuen, was bei einem Konfidenzintervall von 95% einer Variabilität zwischen 156-314 Individuen entspricht. Die Populationsschätzung 2007 brachte 415 Individuen hervor, was bei gleichem Konfidenzniveau wie 2006 einer Variabiltität zwischen 315-561 Individuen entspricht. Dies ließ auf zu niedrige Stichproben (2006 betrug n = 141, 2007 n = 326), zu wenig erfolgreich analysierte Proben (2006 n = 89, 2007 n = 156) und/ oder zu wenig Wiederfänge (2006 n = 12, 2007 n = 24) schließen. Zudem ergaben die Schätzungen sogar deutlich höhere Populationszahlen als zuvor durch Statistiken der Jagderträge vermutet wurde, was auf eine uneffektive Bejagungsstrategie in dem Studienareal hindeutet. Für den zukünftigen Ausblick ist es unabdingbar die Stichprobenzahl deutlich zu erhöhen um die Validität und Reliabiltät der Populationsschätzungen zu gewährleisten, da diese für das Wildmanagement und die epidemiologischen Lösungsstrategien von höchster

Bedeutung sind. Nichtsdestotrotz konnte die im Rahmen dieser Arbeit entwickelte Methode zur individuellen Wildschwein Genotypisierung erfolgreich etabliert werden. Die daraus resultierenden Datensets zur Modellierung von Populationschätzungen sind zuverlässig und weisen eine ausreichend geringe reale Genotypisierungsfehlerrate auf.

1 General introduction

1.1 Studied species wild boar (*Sus scrofa*)

The wild boar (*Sus scrofa*) is one of the most widely distributed wild ungulates in the world (Acevedo *et al.* 2007). The geographical distribution extends in the northern temperate zone from Western Europe to Southeast Asia, but also in parts of Oceania and North Africa (Heck & Raschke 1980). Moreover, the wild boar was introduced as game species by humans to other continents with exception of the Antarctica (Scandura *et al.* 2011), and spread enormously by mating with domestic pigs (Boback 1957). In Germany, the wild boar occurrs almost everywhere except high mountain areas. Wild boars are not evenly distributed throughout, they prefer swampy, deciduous wooded area with lakes and rivers and broad belts of reeds. However, they also prefer areas close to agricultural landscapes, where they can find protection in coniferous thickets (Heck & Raschke 1980). In recent years this species population size increased in colonized areas, e.g. Thuringia, Saxony, Schleswig-Holstein and Baden-Württemberg, which have long been known as wild boar free (Briedermann 1990). This wide distribution of wild boar and its 22 sub-species, which are distinguished by their lacrimal bones length and shapes, implies a remarkable ability to adapt, which only few species possess. The reason for the particular adaption and wide distribution of wild boars is often their dietary characteristics. Wild boars are omnivorous, their main food is composed of plant material such as leaves, shoots and fruits of many woody plants, herbs and grasses. Moreover, they burrow through the soil searching also for worms, grubs,

mice, snails, insects, fungi and even carrion (Briedermann 1990). In general wild boars confirm their diet to external circumstances, thus the ingestion can highly depend on the season, habitat and the human farming system in particular (Henning 1998).

Wild boars belong to animal groups which live in social groups called sounders. The social structures are clearly distributed among the boars. The adult females live together with their offspring and also with one-yearlings and other females under one dominant female direction. With the attainment of puberty the males begin a solitary life. In their sexual active time they come back to the sounders and often fight with rival males for dominance (Heck & Raschke 1980). Wild boars are strong connected to their living place, they return again and again to their selected resting and feeding places depending on season, available food and society structure. In winter the home ranges of wild boars are larger due to lower food availability. Generally, solitary individuals have bigger home range than family groups (Keuling *et al.* 2008).

1.2 Wild boar population dynamics and management

The wild boar population in Rhineland-Palatinate (see Figure 1.1) and also throughout Europe increased during the past three decades. The reasons for the increased wild boar populations are due to socio-economic changes such as abandonment of rural areas, changes in most common crops, lack of predators, reintroduction or restocking, insufficient hunting and significantly milder winters (Saez-Royela & Telleria 1986; Boitani *et al.* 1995). Furthermore, the reproduction

rate of wild boars is considerably higher than in other comparable ungulates (Servanty *et al.* 2009). The reasons are the early sexual maturity and high number of cubs. In Germany the number of piglets is six to eight whereas in other countries a number of four to five occur (Sodeikat & Pohlmeyer 2002). It is assumed that population regulation is strongly controlled by hunting. The population regulation is limited when harvest is focused on adult males or when hunting pressure on adult females and piglets is reduced (Toigo *et al.* 2008). However, the wild boars are often harvested with little control on game limits and the number of shooted boars still increase every year (Boitani *et al.* 1995), this implies that hunting does not have a sufficient effect on the population.

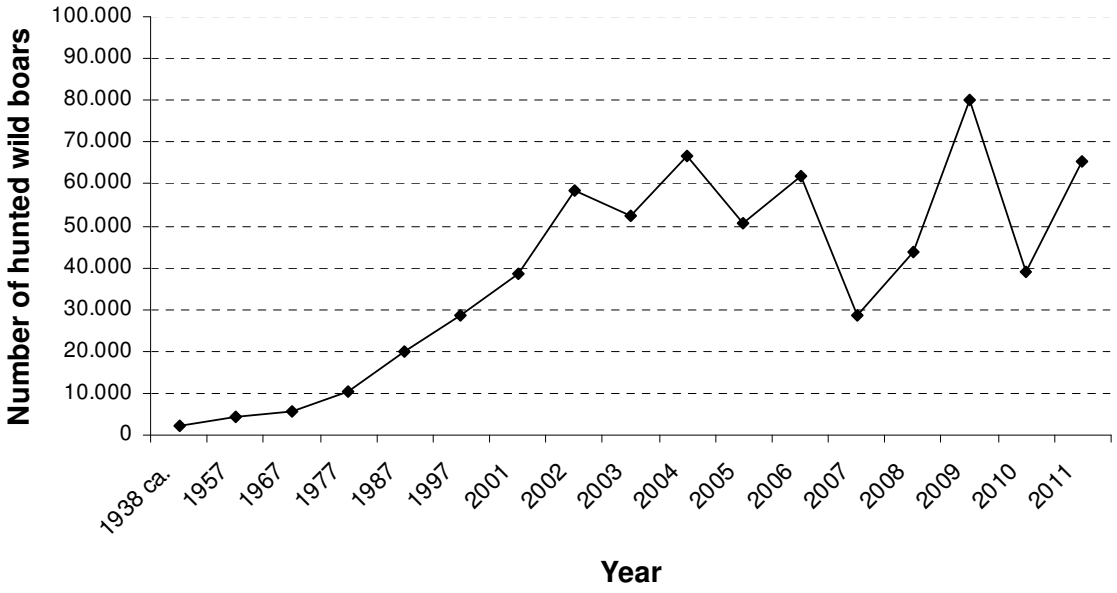

Figure 1.1: Huntingstatistics for wild boars in Rhineland-Palatinate since 1939. Numbers of hunted wild boars were published by "Ministerium für Umwelt, Landwirtschaft, Ernährung, Weinbau und Forsten" in Rhineland Palatinate (www.wald-rlp.de)

Meanwhile the high wild boar number lead to serious problems for agriculture, because they damage crops and rooting grassland by trampling and foraging (Schley *et al.* 2008; Scillitani *et al.* 2010). Therefore, costs of compensation to farmers have increased dramatically, i.e. 17 milion € in Western Europe in 2001 (Toigo *et al.* 2008). As possible solution for this problem, supplementary feeding besides intensive hunting was assumed. However, many studies advise against supplementary feeding and even point out a reverse effect as observed population increase resulting in even higher damage (Bieber & Ruf 2005; Schley *et al.* 2008). Schley *et al.* (2008) recommend to plant trichomatous cereals close to forests, which are negotiated by wild boars, and the preferred cereals should be planted further away from the forests. In this case the longer distance without protection would be too risky for the most individuals.

Another negative aspect of high wild boar numbers is the transmission of diseases to domestic livestock, pets and humans. They act as reservoir of diseases like the classical swine fever, Aujeszky´s Disease Virus (ADV), *Mycobacterium tuberculosis* complex (MTBC), *Trichinella* or Hepatitis E virus (Acevedo *et al.* 2007). In Germany the transmission of the swine fever to domestic pigs became a serious problem besides the agricultural damage in wildlife management. In order to stem the transmission of this disease it is necessary to reduce the wild boar population to less than two individuals per 100 ha (Kaden 1999). Absolute population size estimations are urgently needed for effectively studying the epidemiology of wildlife diseases for wildlife management of the regulation of overabundant populations and for conservation of endangered species. Estimations over a longer time period give important information about

demographic changes, as growth, migration and effectiveness of population regulation particularly. Nowadays population estimations are done in Rhineland Palatinate by taking the hunting statistics into account. This occurs also to Belgium, Italy, France and other European countries. In Poland observation methods with a high man power are used in some forest areas. Nevertheless these methods provide only an approximation to the real population sizes.

1.3 Population size estimation

There are several methods to determine a population sizes, such as direct observations by day at feeding stations or using line transects (Ickes 2001; Focardi *et al.* 2002). By night there are possibilities to count mammals with spotlights or using infrared cameras to take aerial pictures with helicopters (Acevedo *et al.* 2007). Further methods are indirect observations, indices as faeces counts (Focardi *et al.* 2002) and using hunting statistics (Boitani *et al.* 1995). Methods including direct observations are often limited due to limited number of observer and the dependence on the observation effort. Especially for wild boar counting methods by night should be neglected due to the failed reflecting *tapetum lucidum* in their eyes (Acevedo *et al.* 2007) what makes the observation of wild boars with spotlight almost impossible. Problematic is also the use of infrared cameras in helicopters because wild boars often stay in thickets so that they cannot be detected (Acevedo *et al.* 2007). Indirect observations or indices as faeces counts are biased by decomposition rates among seasons and habitats and by lack of individual assignment (Focardi *et al.* 2002).

The most widely distributed method in wildlife management to estimate wild boar population size is based on hunting effort and it´s harvest (Boitani *et al.* 1995; Acevedo *et al.* 2007). But this method depends on seasons where protected areas are not available, while large sampling areas are necessary to reduce the bias (Siren *et al.* 2004). Nevertheless, all above presented methods provide relative population size estimations or population trends, which could be sufficient for questions concerning e.g. population dynamics approximations, but they are not sufficient enough for precise and absolute censuses which are inevitable for epidemiological cases.

A common method to obtain absolute population size estimations for wildlife animals is the capture mark recapture (CMR) approach (Otis *et al.* 1978), where random samples from a population are taken by catching individuals followed by marking them (e.g. with ear tags or color). Afterwards the marked individuals are released into the wild. A second random sample from the same population allows the population size estimation based on the proportion of recaptured and marked individuals (Andrzejewski *et al.* 1978, Otis *et al.* 1978, Pollock *et al.* 1990). This method is based on the following assumptions which have to be fulfilled: 1) a closed population where population additions as births and / or immigrants or deletions as deaths and / or emigrants do not occur or are too low to bias the model calculations; 2) the marked individuals do not lose their marks; 3) the marked individuals mix completely with the unmarked individuals after their release and before the next capture; 4) the capture has no impact on the remaining individuals; 5) the probability to be captured varies not between individuals; 6) the capture probability does not change between the sampling

occasions; 7) marked and unmarked individuals have the same survival probability (Otis et al. 1978; Pollock et al. 1990). However for animals like wild boar, this approach carries a high potential for systematic error, since the basic requirements as explained above under 1), 4), 5) and 6) are often not given. The reason for 4), 5) and 6) is a group-specific capture heterogeneity. Younger and inexperienced individuals are much more frequently captured (Baubet 1998). Therefore the capture and recapture probabilities are variable and not equal for each individual. Captured animals are shyer than inexperienced animals. To ensure the requirement of a closed population (see point 1) a short experimentation time is necessary and for precise estimation and the statistical CMR models the capture recapture experiments in a study should be repeated more than twice (Pollock *et al.* 1990). The realization of enough captures of wild boar individuals in a short time period requires a high effort regarding personal as well material and is almost unrealistic. Furthermore it leads to considerable disturbance and stress for the animals.

Most of the above mentioned disadvantages of CMR can be circumvented by using individual identification with non-invasive genotyping (Gagneux *et al.* 1997; Mowat & Strobeck 2000), an approach that has increased within the last five years (Adams & Waits 2007). Genotyping for population size estimation has already been successfully used for different species e.g. as estimations based on genetic hair analyses by chimpanzee (*Pan troglodyte*; Gagneux *et al.* 1997), pine martens (*Martes americana*; Mowat & Paetkau 2002) or bears (*Ursus spp*; Mowat *et al.* 2005). The population size estimations based on genotyping of faeces are similarly to that based on hairs. Arrendal *et al.* (2007) successfully estimated

population sizes of otters (*Lutra lutra*), Solberg *et al.* (2006) of brown bears (*Ursus arctos*) or Prugh *et al.* (2005) of coyotes (*Canis latrans*). A comparable approach for wild boar population size estimation is still lacking. Therefore a modified CMR approach using non-invasive genotyping is a promising tool for absolute population size estimations of wild boars (see Figure 1.2).

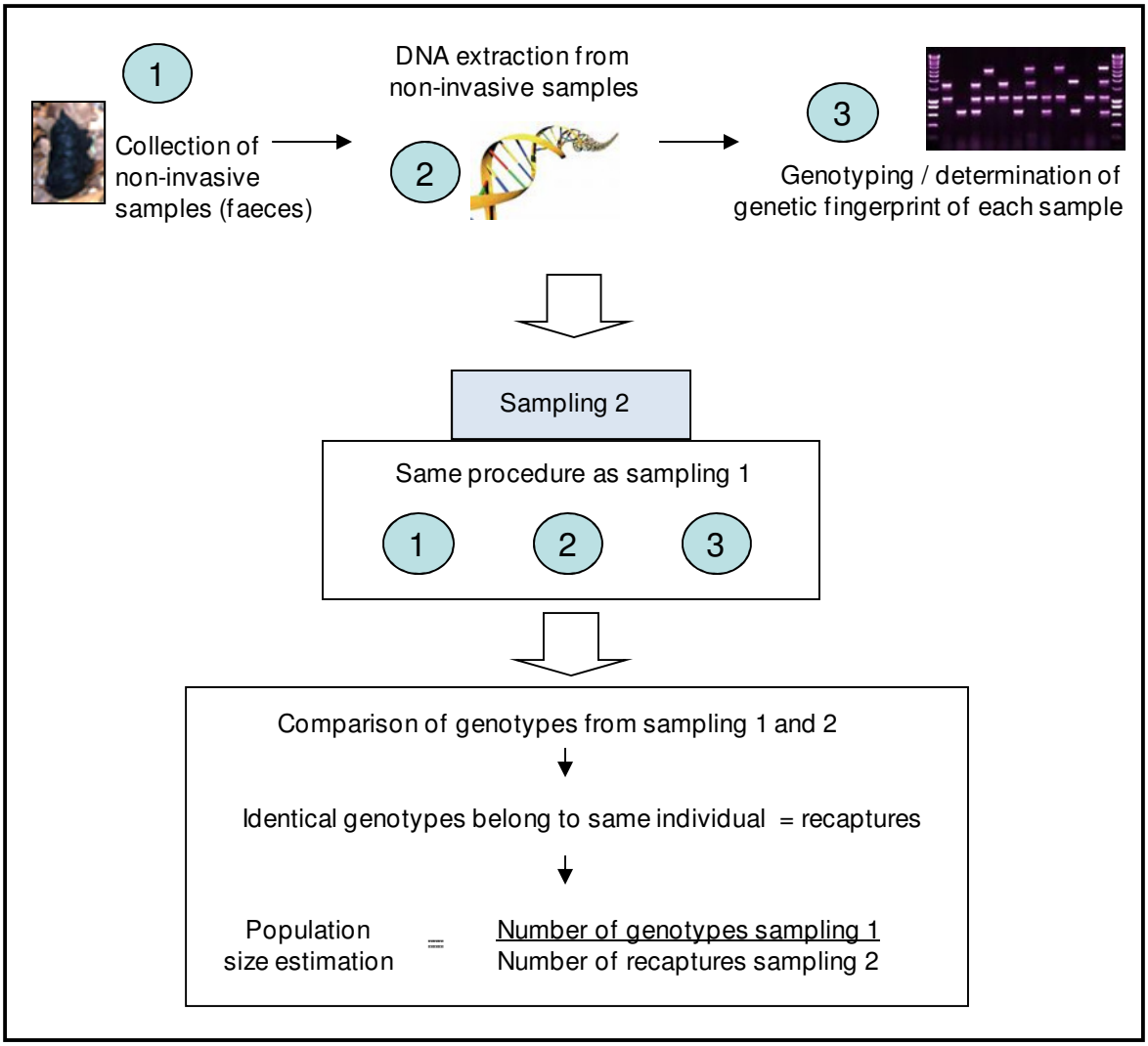

Figure 1.2: Flowchart for a modified capture-mark-recapture (CMR) approach using non-invasive genotyping. Sampling 1 and 2 are collection days with intervals at least of 24 hours. An increased number of collection days provided more accurate estimates.

1.4 Individual identification by genotyping

Genotyping is based on microsatellite (MS) analysis, which is used to determine genetic fingerprints. MS are short and non-coding DNA fragments consisting of several tandem repeats of two to four nucleotides (Ellegren 2004). They have high mutation rates and thus length polymorphisms; hence they can vary among individuals in their sequence length. After DNA isolation the amplification of specific DNA fragments is carried out with polymerase chain reaction (PCR) and finally the fragment length of the MS is determined. The combination of several MS markers provides a reliable identification of an individual.

For this thesis faeces were used as DNA source instead of hairs. The pilot study explained in Ebert *et al.* (2010) showed that sampling with baited hair traps is not suitable due to heterogeneous individual sampling probabilities, indicating that adult and subadult animals differ in their behaviour dependent on their group status. Collecting of faeces does also not require any improvements such as hair catcher and feeding sites, which are necessary to attract the animals. Furthermore, the number of hair samples to be analysed per individual is very high compared to the number of faeces and would increase the costs. However, the use of faecal samples implicates a big challenge for the laboratory methodology caused by low DNA quality and quantity. Low DNA quality occurs due to contamination with PCR inhibitors as herbal recycled fibre stock (Monteiro *et al.* 1997; Reed *et al.* 1997) or alien DNA from bacterial and diet (Murphy *et al.* 2000; Lampa *et al.* 2008). Low DNA quantity occurs due to low target DNA concentrations (Hajkova *et al.* 2006; Hedmark & Ellegren 2006) and degradation (Kohn *et al.* 1995; Frantzen *et al.* 1998; Idaghdour *et al.* 2003) by UV-light,

endogenous endonuclease activity or oxidative damage (Deagle *et al.* 2006). These disadvantages provide low amplification and genotyping successes and furthermore incorrect genotypes due to genotyping errors as allelic dropouts and false alleles (Taberlet *et al.* 1996; Huber *et al.* 2003; Wehausen *et al.* 2004). An allelic dropout occurs when one allele of a heterozygous individual is not amplified during a successful PCR, whereas false allele is a wrong allele which is generated by PCR (Pompanon *et al.* 2005; Broquet *et al.* 2007) resulting from contaminations with foreign DNA, from slippage artefacts during the first cycles of PCR, or cross-contaminations. Genotyping errors have a big impact on population size estimations, as they can lead to either underestimations when genetic information is not sufficient or to overestimations when genotyping errors lead to false genotypes not present in the study population (Creel *et al.* 2003). Several studies were carried out to avoid or reduce this kind of bias, by e.g. 1) sampling of fresh faeces in winter to reduce DNA degradation (Maudet *et al.* 2004), 2) using optimized storage, extraction methods and PCR conditions to reduce PCR inhibitors (Wasser *et al.* 1997; Flagstad *et al.* 1999; Murphy *et al.* 2002; Piggott & Taylor 2003; Murphy *et al.* 2007), 3) using specific primers to avoid amplification from alien DNA (Bradley *et al.* 2001; Broquet *et al.* 2007), 4) reducing of genotyping errors by elaborative selection of genetic markers and application of multitubes approach (multiple repeated PCR per locus and per sample) (Goossens *et al.* 2000; Waits *et al.* 2001; Hedmark & Ellegren 2006; Adams & Waits 2007; Broquet *et al.* 2007), 5) preselecting low quality samples with quantitative PCR (Morin *et al.* 2001; Deagle *et al.* 2006), 6) checking the presence of null alleles, which are non-amplifying alleles due to a mutation in the primer

target sequence (Pompanon *et al.* 2005). The huge variety of recommended methodological tools established a labour-intensive demand for this thesis.

2 Objectives

The present thesis is part of a collaboartive project between the Research Institute of Forest Ecology and Forestry (FAWF) in Trippstadt and the Institute for Environmental Sciences at the University of Koblenz-Landau aiming to develop a non-invasive method for the detection of spatial-temporal dynamics and density of wild boar populations in order to control the spreading of classical swine fever in Rhineland-Palatinate. This project contained two PhD theses, the here presented one and the other written by Cornelia Ebert, dealing with the development of sampling design, radio-telemetry data and population size modeling. The present thesis has the following objectives:

- Development of a laboratory method for individual identifications of wild boar faecal samples by genotyping, which has to fulfill the following criteria:
    a) Reliability
    b) Practicability
    c) Reproducibility
    d) Cost-efficiency
- Proposal of a strict procedure and cost calculation for individual identifications of wild boar faecal samples.
- Application of the developed method on an open wild boar population in order to estimate the population size and sex proportion in a 4000 ha area in the Palatinate forest.

3 Thesis structure

The present PhD thesis is written as cumulative thesis including three manuscripts (see Appendix) which are published or submitted as scientific manuscripts in peer reviewed journals:

Appendix I: Evaluation of faecal storage and DNA extraction methods in wild boar *(Sus scrofa)*

In this manuscript the main focus lies on the evaluation of eight storage and DNA extraction method combinations for wild boar faecal samples by detecting the amplification and genotyping success and by determining the wild boar DNA amount using quantitative PCR.

Appendix II: Determination of the minimum number of microsatellite markers for individual genotyping in wild boar (*Sus scrofa*) using a test with close relatives

This manuscript describes a step by step procedure to determinate the minimum of MS markers for individual genotyping of wild boar faecal samples. It deals with the choice and test of a MS marker set with three different wild boar populations and finally with a test with close relatives.

Appendix III: Comparison of established methods for quantifying genotyping error rates in wildlife forensics

This manuscript deals with the quantification of real genotyping error rates (GER) in a data set obtained from wild boar MS analysis. Different methods for determination of GER are conducted and compared for the same dataset. Finally a strict procedure for the determination of real GER is proposed with an additional blind test.

4 Main thesis insights

4.1 Method development for individual identification and fulfilment of the required criteria

a) Reliability

Reliability is one of the most important criteria for individual identification. As there are no reference data for absolute wild boar population size estimations, a rigorous and conservative method evaluation is of crucial importance. In the following sections the approaches for obtaining reliability are explained.

In order to increase the DNA quality, eight combinations of storage and extraction methods were tested by determining amplification and genotyping successes for three randomly chosen MS markers out of eight possible MS markers (see Appendix II) and by determining the target DNA amount using a quantitative PCR (qPCR) assay (see Appendix I), which is useful for preselecting of reliable faecal DNA extracts (Morin *et al.* 2001, Hausknecht *et al.* 2010). Furthermore the efficiency of five different Taq polymerases was compared. Successful PCR-runs in the range of 27.75% to 77.73% were obtained, whereas the rate of genotyping success ranged between 22.2% and 66.62%. The mean wild boar DNA amount was between 0.15 ng/µl and 2.07 ng/µl (see Figure 4.1). Testing different Taq polymerases provided PCR rates between 33% and 70%, and genotyping success rates between 30% and 63%. The main outcome was that PCR and genotyping success could be considerably increased by optimised storage / extraction and PCR conditions for non-invasive samples. An increased

DNA quality leads to a reduction of genotyping errors and thus to an increase of the required reliability.

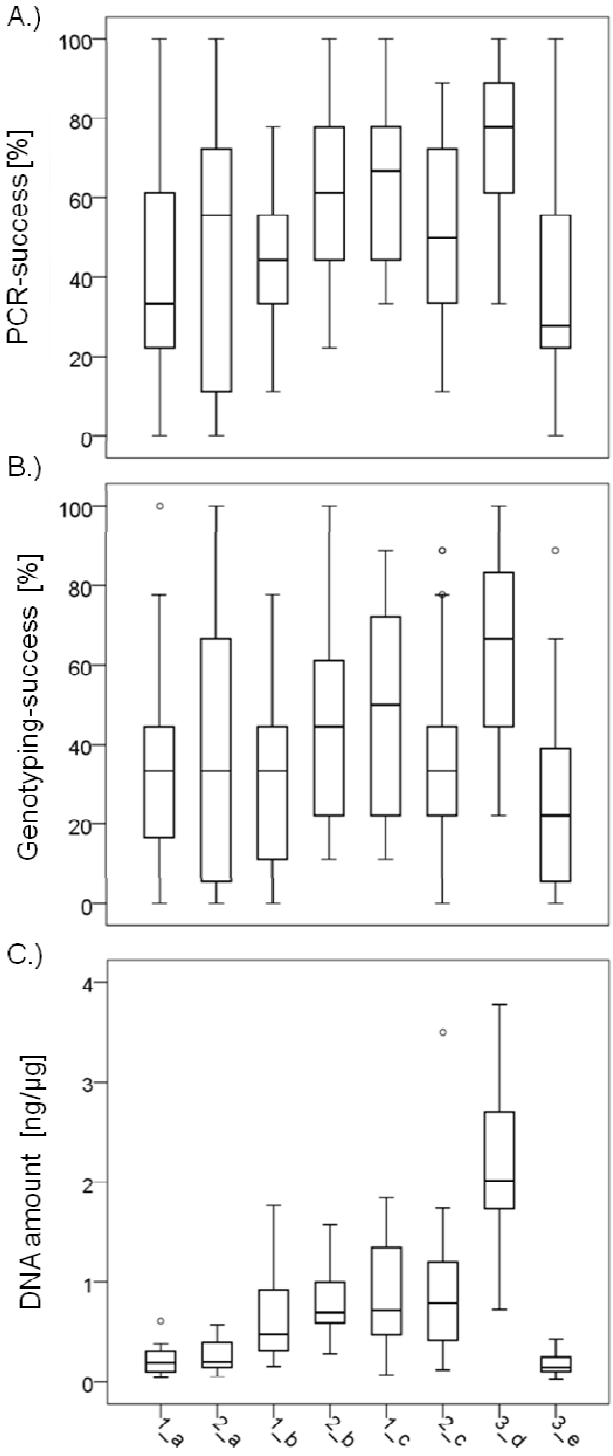

Figure 4.1: Comparison of eight combinations of storage / extraction protocols using 20 wild boar faecal samples. A) PCR success and B) genotyping success was determined using three microsatellite-markers (CGA, Sw742 and Sw2496) with three repeats per sample and locus. C) Wild boar DNA amount was quantified using quantitative PCR with primer TAGLN-Sus and two repeats per sample (Appendix I).
— Median; □ 25% - 75%; Error bars indicate standard deviation

In most other studies on wildlife forensics six to ten MS markers are commonly used (e.g. Wilson *et al*. 2003; Hajkova *et al*. 2009; Marucco *et al.* 2009). However, a higher number of MS markers increase the potential GER. To ensure the reliability of the used MS a stepwise procedure to reduce the number of MS loci for individual genotyping in wild boar was developed (see Appendix II). Step1: An initial marker set was tested for species specificity with non-target DNA. Step 2: A variability test regarding heterozygosity and deviations from Hardy Weinberg equilibrium was carried out. Step 3: Test for transferability across populations with three separate wild boar sample sets. Step 4: Calculation of probability of identity (P_{ID}), which predicates the ability of molecular markers to distinguish between different individuals (Taberlet & Luikart 1999). Step 5: A novel test using tissue samples from female wild boars and their embryos provided evidence that four variable MS markers and one sex-marker are sufficient for individual identification of close relatives (see Table 4.1). Step 6: Faeces samples were finally used to estimate PCR and genotyping success. This step by step procedure allowed using four MS markers and an additional sex marker to obtain reliable dissolution of individuals.

Table 4.1: Thirty-two four-loci combinations and additional sex marker (PigSRY) for reliable resolution of closely related wild boar individuals (Appendix II). Combinations are arranged by increasing product P_{ID}sib calculated per each combination by allele frequencies from the closely related individuals (n=23)

	Sw742	CGA	S0068	S0005	Sw461	Sw2496	Sw2021	TNFB	PigSRY	P_{ID} sib
1	x	x		x	x				x	0.010
2		x	x	x	x				x	0.011
3		x	x		x			x	x	0.011
4		x		x	x			x	x	0.011
5	x	x	x		x				x	0.012
6	x		x	x	x				x	0.012
7	x	x	x	x					x	0.013
8		x		x	x		x		x	0.013
9		x	x		x		x		x	0.014
10	x	x			x		x		x	0.014
11	x	x	x				x		x	0.015
12	x	x		x			x		x	0.015
13	x			x	x		x		x	0.015
14		x		x	x	x				0.016
15		x	x		x	x			x	0.016
16	x		x	x			x		x	0.016
17	x	x			x	x				0.017
18	x	x			x			x	x	0.017
19	x	x		x		x				0.018
20	x			x	x	x			x	0.018
21	x		x		x	x				0.018
22	x	x		x				x	x	0.019
23	x			x	x			x	x	0.019
24		x			x	x	x			0.020
25		x			x		x	x	x	0.020
26	x	x				x	x			0.022
27		x		x			x	x	x	0.022
28	x	x					x	x	x	0.023
29	x		x			x	x			0.024
30		x			x	x		x		0.024
31	x			x			x	x	x	0.025
32*	x					x	x	x	x	0.026

* Four-loci combination recommended for faeces samples.

Furthermore to achieve reliable MS datasets for population size estimations in wildlife forensics with a realistic GER of less than 5%, ensuring reliable population size estimations (Taberlet & Luikart 1999), three methods for determination of

GER within one study were compared and a blind-test for quantifying a realistic GER was presented (Figure 4.2). It is important to consider the real GER instead of the theoretical one because it reflects the accuracy of the individual assignment (Frantz *et al.* 2003). The error rates differed widely between these three methods (0 to 57.5%) and underline the need of a consensus approach. The blind-test resulted in a GER of 4.3% (see Appendix III) which confirmed the required GER less than 5% and thus the reliability of the obtained dataset.

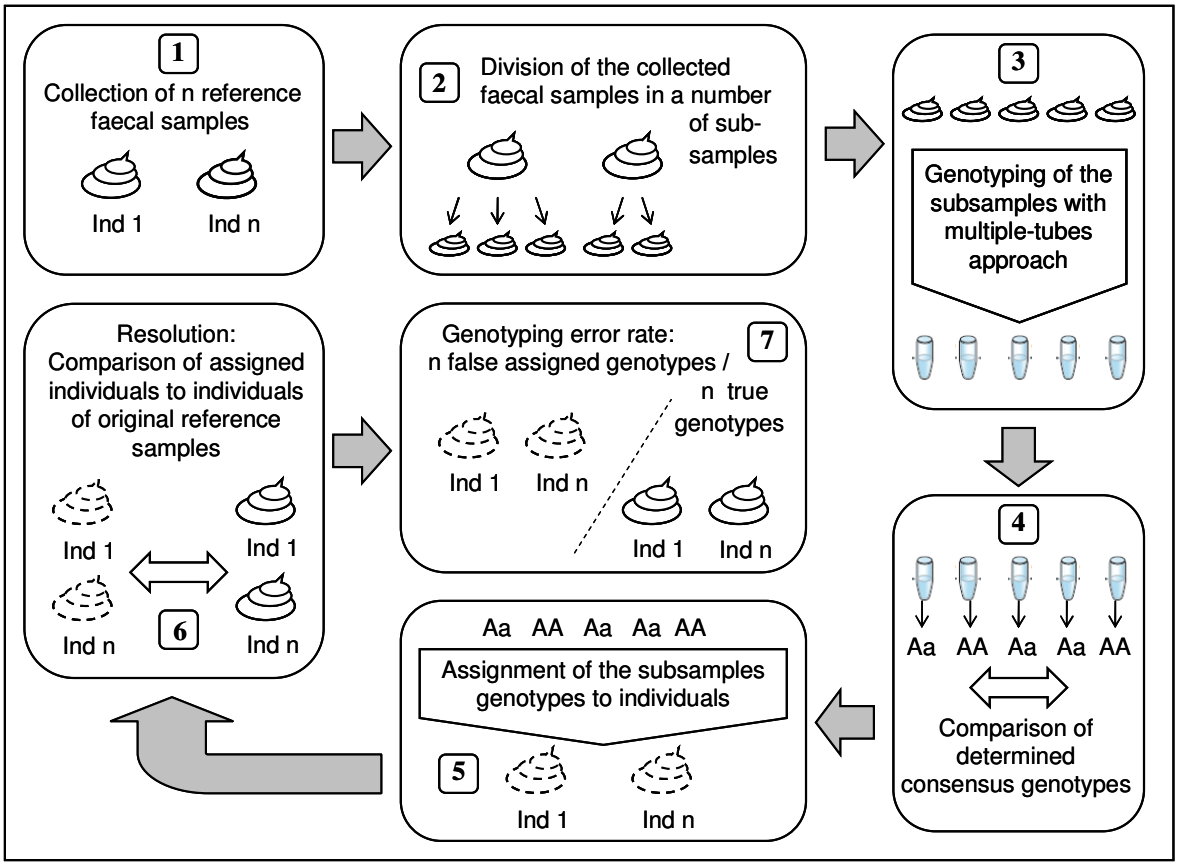

Figure 4.2: Flowchart of a blind-test to estimate realistic genotyping error rates. The number of collected reference faecal samples should not be lower than 20. We recommend a minimum number of 10% of collected samples in one collection period and to provide a higher number of samples from different individuals than subsamples. We suggest starting the test with a comparative multiple-tubes approach (fewer repetitions) in respect of reduced costs. In case of high error rates, the multiple-tubes approach could be modified by increased repetitions (see Appendix III).

Reproducibility

In order to ensure the reproducibility of the developed method a multiple tube approach was initially established. The procedure included dividing the DNA extract among several tubes, then amplifying and genotyping the contents of each tube separately (Navidi *et al.* 1992). Samples with ambiguous or different genotyping results were discarded from analysis (see Appendix II and III). Furthermore a transferability of the used marker set for geographical disconnected wild boar populations and for work in other laboratories was carried out (see Appendix II). The genetic variation for populations from Rhineland Palatinate Forest, Lower Saxony and Mecklenburg–Western Pomerania was compared by calculating the observed H_o and expected Heterozygosity H_e which gives information about the polymorphism of a marker, and the conformance to Hardy Weinberg equilibrium of each marker (Table 4.2). H_e values ranged in the optimum for MS from 0.60 to 0.86 and did not differ significantly across populations. Moreover, between the two subsamples from the same population (RP.1, RP.2), which were analysed in different laboratories, the mean difference of H_e and H_o was below 0.03, indicating high transferability of the markers even across different laboratory systems. This transferability supports the reproducibility of the developed method.

Table 4.2: Comparison of expected (H_e) and observed (H_o) heterozygosities and Hardy Weinberg Equilibrium (HWE) for eight microsatellite loci, as well as inbreeding coefficient (F_{IS}) among different wild boar populations (Appendix II).

Pop	RP.1			RP.2			LS			MWP		
F_{is}	0.0350			0.0131			0.2213			0.0971		
Locus	H_e	H_o	HWE	H_e	H_o	HWE	H_e	H_o	HWE	H_e	H_o	HWE
Sw742	0.83	0.75	ns	0.85	0.86	ns	0.79	0.64	ns	0.78	0.78	Ns
CGA	0.82	0.84	ns	0.84	0.78	ns	0.85	0.85	ns	0.85	0.90	Ns
S0068	0.82	0.77	ns	0.83	0.83	ns	0.84	0.67	***	0.66	0.65	Ns
S0005	0.80	0.82	ns	0.87	0.74	***	0.88	0.80	ns	0.82	0.86	Ns
Sw461	0.80	0.82	ns	0.86	0.85	ns	0.75	0.61	ns	0.60	0.67	Ns
Sw2496	0.73	0.61	ns	0.76	0.74	ns	0.76	0.55	***	0.80	0.76	Ns
Sw2021	0.71	0.61	ns	0.73	0.68	ns	0.76	0.35	***	0.71	0.10	***
TNFB	0.70	0.80	ns	0.72	0.70	ns	0.68	0.48	***	0.77	0.72	Ns
mean	0.78	0.75		0.81	0.77		0.79	0.62		0.75	0.68	

ns not significant, * p<0.01, *** p<0.0001

c) Cost-efficiency

The reduction of the required number of MS marker to the minimum (see Appendix II) is an important tool for the reduction of costs. For example in our case the cost of molecular analyses (without personnel costs) using four MS markers is about 29€ per sample and increases by 13€ for every additional marker (explained in detail in section 4.3). Thus the use of four MS marker leads to a cost reduction of around 31% compared to a study using six markers. Compared to a study with eight markers the cost reduction would be around 48% and in comparison with 10 markers study the cost reduction would be around 58%.

d) Practicability

The present thesis succeeds in an elaborated stringent laboratory procedure (see Figure 4.3) for individual genotyping of wild boar faecal samples, which provides datasets with a realistic GER lower than 5%. Taberlet *et al.* (1999) Paetkau (2003) and Lukacs & Burham (2005) reported that laboratory protocols can obtain GER within a range of 5% without influencing the estimation of population sizes by performing misidentification models. Stringent recommendations for faecal storage, DNA extraction, PCR conditions; MS marker set, preselection of low quality samples, multiple tube approach and quantification of real GER (see Appendix I, II and III) were elaborated.

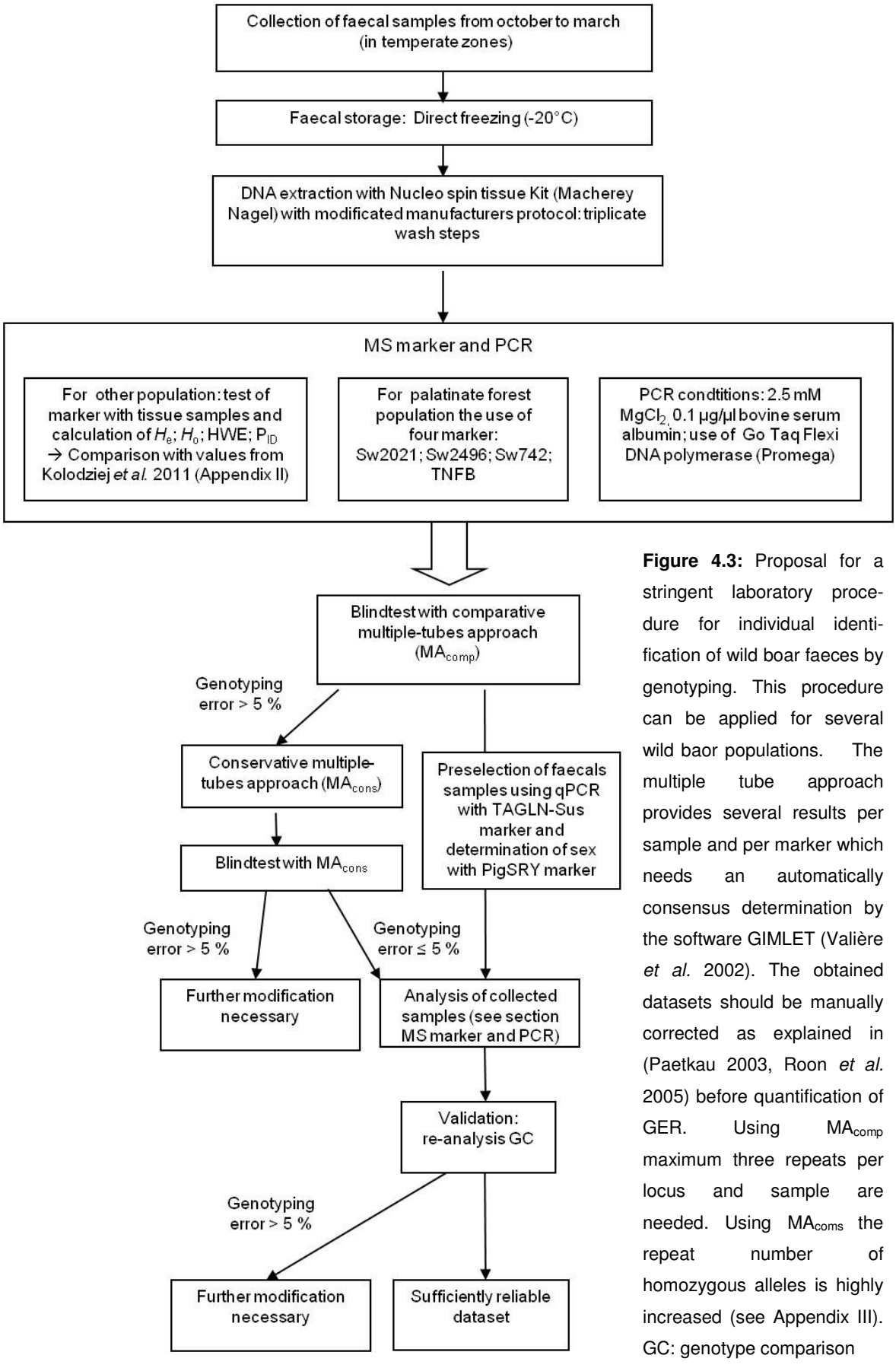

Figure 4.3: Proposal for a stringent laboratory procedure for individual identification of wild boar faeces by genotyping. This procedure can be applied for several wild baor populations. The multiple tube approach provides several results per sample and per marker which needs an automatically consensus determination by the software GIMLET (Valière *et al.* 2002). The obtained datasets should be manually corrected as explained in (Paetkau 2003, Roon *et al.* 2005) before quantification of GER. Using MA_{comp} maximum three repeats per locus and sample are needed. Using MA_{coms} the repeat number of homozygous alleles is highly increased (see Appendix III). GC: genotype comparison

4.2 Estimation of populations sizes in wild boar: application of the developed method

The following section summarizes the results of the first two population size estimations using the method developed in this thesis (Appendix I, II and III). The results are part of Cornelia Ebert's PhD thesis and are important for the presented thesis to demonstrate the application of this method in a real wild boar population. Faeces sampling was carried out in December 2006 and 2007 in an area of 4000 ha situated in the Palatinate Forest (see Ebert 2011). The samples were collected along 16 transects (see Figure 4.4) every 48 hours and during 12 days in each trial.

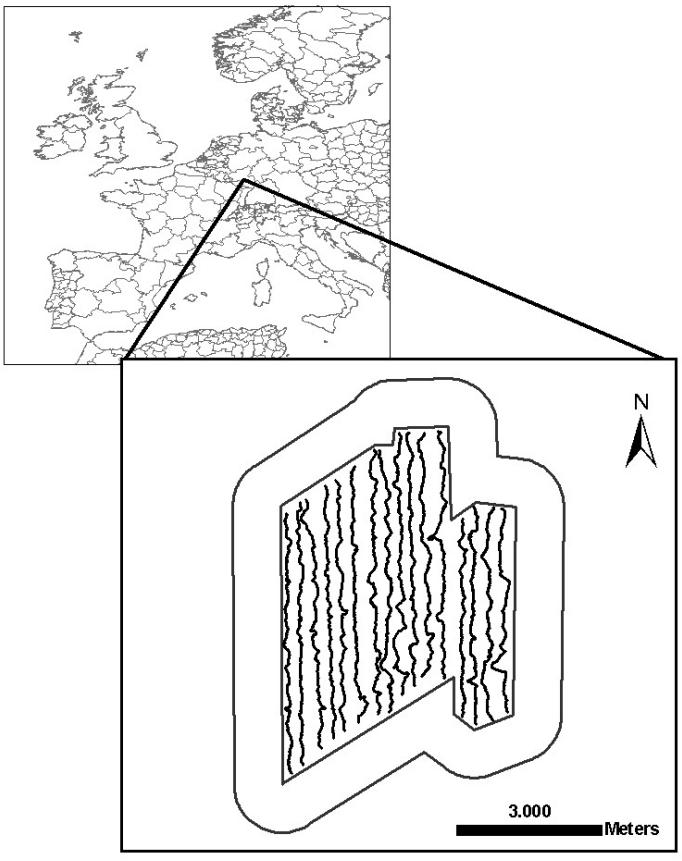

Figure 4.4: Transect design for collection of wild boar faeces for use in non-invasive genetic population estimation (Ebert 2011). The transects are orientated in N-S direction. The area covered by transects together with the buffer represents the effectively samples area. The study area is situated in the federal state of Rhineland-Palatinate in south western Germany.

In 2006 141 wild boar faeces were collected. After analysis 89 (63%) could be successfully genotyped. From these 75 different individuals and 12 recaptures could be identified. In the following year 2007 the double amount of faecal samples (n = 326) was collected. A number of 156 samples (47.8%) yielded complete consensus genotypes. From these, 132 individuals were identified with 24 recaptures. This genotyping success is in the same range as reported in previous non-invasive genotyping studies (Reed *et al.* 1997; Kohn *et al.* 1999; Ernest *et al.* 2000; Lucchini *et al.* 2002; Wilson *et al.* 2003; Bellemain *et al.* 2004; Hedmark *et al.* 2004 and Arrendal *et al.* 2007). The sex ratio was 1.14 : 1 male to female in 2006 and 1.03 : 1 in 2007 respectively. However it is not sure, if the ratio reflects the reality or should be regarded as an artifact of the small sample size. The resulting population size estimations calculated with four different models, each with and without the misidentification of 5% (based on real GER of 4.3%), are given in Table 4.3. The main outcomes were: the heterogeneity models (M h) differed widely from the remaining three models in their estimates and the confidence intervals are in all models too high implying inaccuracy of the estimates. Furthermore, the estimated population sizes between models of the same type with and without misidentification differed marginally from each other (mean 10%). The resulting inaccuracies of the estimations were due to low recapture numbers and particularly due to low sampling sizes. For example Solberg *et al.* (2005) obtained accurate estimations of brown bear population size with low confidence intervals by collecting 2.5 to 3 times as many samples as the estimated population sizes. Wheras Otis *et al.* 1978 recommend to achieve 30% of recapture numbers out of the capture numbers per collection event.

The estimates generated by M Null and M t models appeared to be the most reliable models because they correspond much more to the relation of sample size, hunting bag and population size between the two study years.

Table 4.3: Population estimates derived from wild boar faeces sampling in December 2006 and 2007 using different models in program MARK. Pop size N is the estimated population size including 95% confidence interval (CI).

	Sampling December 2006		Sampling December 2007	
Model	Pop size N	CI 95%	Pop size N	CI 95%
M h 5%	534	179 - 1204	1630	413 - 6544
M h	602	236 - 1793	1842	414 - 6529
M t 5%	215	156 - 314	415	318 - 561
M t	235	169 - 346	457	350 - 616
M th 5%	283	196 - 1766	583	413 - 854
M th	312	267 - 684	542	398 - 765
M 0 5%	219	159 - 321	433	330 - 589
M 0	242	176 - 352	479	365 - 648

M basic model; M x 5% misidentification model due to genotyping error Lukacs and Burham 2005); h heterogeneity; a mixture model incorporating two groups of animals with differing probability (p); t p varying over time; th heterogeneity and p varying over time; 0 Null as the most parsimonious model with capture p being constant over time among individuals

4.3 Costs calculations

The costs for personnel and transport during the field work reached about 8000€ (Ebert *et al.* 2009). This is carried out by up of four persons working on twelve sampling days and on the proceeding of data. The analysis in the laboratory reached costs between 14.31€ and 40.80€ per sample (Table 4.4) depending on the required repeats of genotyping. In the case of both samplings for 2006 and

2007 the costs for analysis were composed as followed: 40% of the analysed samples were in the range of 40.8€ per sample, 45% were in the range of 21.61€ and 15% in the range of 14.31€. This ratio resulted on average in a cost of 29€ per sample. The personnel effort consisted of one lab technician, who analysed 100 samples in six weeks including the proceeding of data. Assuming that the technician obtain the level TV-L 9 with a middle grade of the civil service with a monthly salary of 2666€, the costs for the processing of 100 samples would be amounts about 4000€ or 40€ per sample. In sum, the laboratory costs including whole analysis and personnel effort are about 69€ per faeces sample. In comparison to other studies the costs are quiet low. Solberg *et al.* (2005) amounted costs about 116€ per faeces sample whereas Wasser *et al.* (2004) obtained cost of approximately 500$ per analysed sample.

Table 4.4: Costs calculation for genotyping analysis of one faecal sample. The costs depend on number of required repeats per sample and per locus. Minimum number of repeats (min No. repeats) occurs when a sample obtain three repeats per locus. Maximum number of repeats (max No. repeats) occurs when ten repeats per locus and per sample are necessary. Preselected low quality samples are listed by without analysis.

Working steps	min No. repeats	middle No. repeats	max No. repeats	without analysis
DNA Extraction	1,84€	1,84€	1,84€	1,84€
Sex-PCR	0,21€	0,21€	0,21€	0,21€
qPCR	0,41€	0,41€	0,41€	0,41€
PCR	1,56€	2,18€	4,68€	
Frag.-Analysis	10€	16,68€	33,36€	
Lab-materials	ca. 0,30€	ca. 0,30€	ca. 0,30€	ca. 0,30€
Sum	**14,31€**	**21,61€**	**40,80€**	**2,76€**

Sex-PCR determination of gender with sex-marker; qPCR quantitative PCR; Frag.-analysis determination of fragment lengths on automatic sequencer

5 Conclusions

The laboratory method presented in this thesis (see Figure 4.3) for individual genotyping of faecal wild boar samples has been supplying reliable datasets for population size estimation with sufficiently low real GER. Furthermore, proceedings as step by step procedure for determination of minimum required MS number (see Appendix II) or determination of real GER (see Appendix III) were established. These proceedings could be also successfully applied to individual genotyping in other species to obtain datasets for population size estimation or kinship analyses. For the future prospect, additional validation of the reproducibility of this method should be conducted. I recommend applying this method from extraction to the dataset determination (see Figure 4.3) on other populations and on other laboratory systems. Moreover, a validation of this method in a population of free living wild boars in a closed areal with known population size would be beneficial caused by the possibility comparing the results to a known reference population size. In order to apply the genotyping method as described in this thesis in an appropriate way for wild boar population size estimation, it is crucial to increase the sample size. Solberg *et al.* (2006) recommend to ensure sampling size, ideally 2.5 to 3 times higher than the expected population size. On the other hand Miller *et al.* (2005) recommend to achieve 2.5 collected samples per detected individual in average. Possible approaches for increasing the sample size are the addition of hair samples after evaluating a collection method which should be homogenic and / or searching more intensively along wild boar passes, at wallows or feeding sites. Finally, even though the estimated population sizes are

inaccurate due to high confidence intervals, the numbers do not correspond in any way to assumed population sizes derived from hunting bags which were communicated by local foresters. The obtained estimates, especially even the lower values in the confidence intervals, are 3-4 times higher as assumed. Thus, the current hunting regime in the study area does not seem to regulate effectively the wild boar population. This implies that the main regulatory mechanisms could be natural factors as food availability and disease occurrence. Other population regulatory mechanisms should be evaluated as changing the hunting regime or using contraceptives as reported in Massei *et al.* (2008).

6 Author's contributions

Table 6.1: Authors contribution for the three manuscripts (MS) included in PhD thesis of Karolina Kolodziej. Sum showed the percentage contribution of whole manuscript work of present thesis author.

	Appendix I	**Appendix II**	**Appendix III**
Original idea	HS 5%; IN 2%	HS 6%; **KK 2%**	UH 2%; HS 6%; CE 1%; **KK 1%**
Project coordination	**RS 5%**	**RS 5%**	**RS 5%**
Lab work	**KK 30%;** IN 6%	**KK 30%;** JB 4%	**KK 30%**
Data analysis	**KK 20%**	**KK 15%;** JB 2%	**KK 20%**
MS preparation	**KK 20%**; KT 6%; RS 2%	**KK 20%**; KT 9%; RS 3%	**KK 20%**; KT 7%; RS 1%
Proofreading	HS 2%; IN 2%	HS 2%; JB 2%	HS 2%; UH 1% ; CE 1%
Field work	/	/	CE 3%
Sum	KK 70%;	KK 67%	KK 71%

KK Karolina Kolodziej; HS Holger Schulz; RS Ralf Schulz; KT Kathrin Theissinger; CE Cornelia Ebert; JB Jörg Brün; UH Ulf Hohmann; IN Ivan Nikolov; MS Manuscript

7 References

Acevedo P., Vicente J., Hofle U., Cassinello J., Ruiz-Fons F. and Gortazar C. (2007) Estimation of European wild boar relative abundance and aggregation: a novel method in epidemiological risk assessment. *Epidemiology and Infection*, 135: 519-527

Adams J.R. and Waits L.P. (2007) An efficient method for screening faecal DNA genotypes and detecting new individuals and hybrids in the red wolf (Canis rufus) experimental population area. *Conservation Genetics*, 8: 123-131

Andrzejewski D.D. and Jezierski W. (1978) Management of a wild boar population and its effects on commercial land. *Acta Theriologica*, 23: 309-339

Arrendal J., Vila C. and Bjorklund M. (2007) Reliability of noninvasive genetic census of otters compared to field censuses. *Conservation Genetics*, 8: 1097-1107

Baubet E. (1998) Biologie du sanglier en montagne: biodemographie, occupation de léspace et regime alimentaire. PhD-thesis, Université Claude Bernard, Lyon

Bellemain E. and Taberlet P. (2004). Improved noninvasive genotyping method: application to brown bear (*Ursus arctos*) faeces. *Molecular Ecology Notes,* 4: 519-522

Bieber C. and Ruf T. (2005) Population dynamics in wild boar Sus scrofa: ecology, elasticity of growth rate and implications for the management of pulsed resource consumers. *Journal of Applied Ecology*, 42: 1203-1213

Boback A.W. (1957) Das Schwarzwild. Biologie und Jagd. *Neumann Verlag,* Leipzig

Boitani L., Trapanese P. and Mattei L. (1995) Demographic patterns of a wild boar (*Sus scrofa* L.) population in Tuscany. *IBEX Journal Mountain Ecology*, 3: 197-201

Bradley B.J., Chambers K.E. and Vigilant L. (2001) Accurate DNA-based sex identification of apes using non-invasive samples. *Conservation Genetics*, 2: 179-181

Briedermann L. (1990) Schwarzwild. *Deutscher Landwirtschaftsverlag,* Berlin, Germany

Broquet T., Menard N. and Petit E. (2007) Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics*, 8: 249-260

Creel S., Spong G., Sands J.L., Rotella J., Zeigle J., Joe L., Murphy K.M. and Smith D. (2003) Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Molecular Ecology*, 12: 2003-2009

Deagle B.E., Eveson J.P. and Jarman S.N. (2006) Quantification of damage in DNA recovered from highly degraded samples--a case study on DNA in faeces. *Frontiers in Zoology*, 3: 11

Ebert C., Kolodziej K., Schikora T.F., Schulz H.K. and Hohmann U. (2009) Is non-invasive genetic population estimation via faeces sampling feasible for abundant mammals with low defecation rates? A pilot study on free ranging wild boar (*Sus scrofa*) in South-West Germany. *Acta Silvatica Lignaria Hungary*, 5: 167-177

Ebert C., Huckschlag D., Schulz H. and Hohmann U. (2010) Can hair traps sample wild boar (*Sus scrofa*) randomly for the purpose of non-invasive population estimation? *European Journal of Wildlife Research*, 56: 583-590

Ebert C. (2011) Development of a method for non-invasive and representative tissue sampling for use in population estimation of wild boar. PhD-thesis, Albert-Ludwigs-University, Freiburg

Ellegren H. (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews*, 5: 435-445

Ernest H.B., Penedo M.C., May B.P. Syvanen M. and Boyce W.M. (2000) Molecular tracking of mountain lions in the Yosemite valley region in California: genetic analysis using microsatellites and faecal DNA. *Molecular Ecology,* 9: 433-441

Flagstad O., Roed K., Stacy J.E. and Jakobsen K.S. (1999) Reliable noninvasive genotyping based on excremental PCR of nuclear DNA purified with a magnetic bead protocol. *Molecular Ecology*, 8: 879-883

Focardi S., Isotti R. and Tinelli A. (2002) Line transect estimates of ungulate populations in a Mediterranean forest. *Journal of Wildlife Management*, 66: 48-58

Frantz A.C., Pope L.C., Carpenter P.J., Roper T.J., Wilson G.J., Delahay R.J. and Burke T. (2003) Reliable microsatellite genotyping of the Eurasian badger (Meles meles) using faecal DNA. *Molecular Ecology*, 12: 1649-1661

Frantzen M.A., Silk J.B., Ferguson J.W., Wayne R.K. and Kohn M.H. (1998) Empirical evaluation of preservation methods for faecal DNA. *Molecular Ecology*, 7: 1423-1428

Gagneux P., Boesch C. and Woodruff D.S. (1997) Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology*, 6: 861-868

Goossens B., Chikhi L., Utami S., de Ruiter J. and Bruford M.W. (2000) A multi-samples, multi-extracts approach for microsatellite analysis of faecal samples in an arboreal ape. *Conservation Genetics*, 1: 157-162

Hajkova P., Zemanova B., Bryja J., Hajek B., Roche K., Tkadlec E. and Zima J. (2006) Factors affecting success of PCR amplification of microsatellite loci from otter faeces. *Molecular Ecology Notes*, 6: 559-562

Hajkova P., Zemanova B., Roche K. and Hajek B. (2009) An evaluation of field and noninvasive genetic methods for estimating Eurasian otter population size. *Conservation Genetics*, 10: 1667-1681

Hansen A.J., Willerslev E., Wiuf C., Mourier T. and Arctander P. (2001) Statistical evidence for miscoding lesions in ancient DNA templates. *Molecular Biology and Evolution*, 18: 262-265

Hausknecht R., Bayerl H., Gula R. and Kuehn R. (2010) Application of Quantitative Real- Time Polymerase Chain Reaction for Noninvasive Genetic Monitoring. *Journal of Wildlife Management*, 74: 1904-1910

Heck L. and Raschke G. (1980) Die Wildsauen. Naturgeschichte-Ökologie-Hege und Jagd. *Verlag Paul Parey,* Hamburg, Berlin

Hedmark E., Flagstad Ø., Segerstroem P., Persson J., Landa A. and Ellegren H. (2004) DNA-based individual and sex identification from wolverine (*Gulo gulo)* faeces and urine. *Conservation Genetics*, 5: 405–410

Hedmark E. and Ellegren H. (2006) A test of the multiplex pre-amplification approach in microsatellite genotyping of wolverine faecal DNA. *Conservation Genetics*, 7: 289-293

Henning R. (1998) Schwarzwild. Biologie, Verhalten, Hege und Jagd. *Blv Buchverlag,* München

Huber S., Bruns T. and Arnold R. (2003) Genotyping herbivore feces facilitating their further analyses. *Wildlife Society Bulletin*, 31: 692-697

Ickes K. (2001) Hyper-abundance of native wild pigs (*Sus scrofa*) in a lowland dipterocarp rain forest of peninsular Malaysia. *Biotropica*, 33: 682-690

Idaghdour Y., Broderick D. and Korrida A. (2003) Faeces as a source of DNA for molecular studies in a threatened population of great bustards. *Conservation Genetics*, 4: 789-792

Kaden V. (1999) Bekämpfung der klassischen Schweinepest beim Schwarzwild. *Zeitschrift der Jagdwissenschaften*, 45: 45-59

Keuling O., Stier N. and Roth M. (2008) Annual and seasonal space use of different age classes of female wild boar *Sus scrofa L. European Journal of Wildlife Research*, 54: 403-412

Kohn M., Knauer F., Stoffella A., Schroder W. and Paabo S. (1995) Conservation genetics of the European brown bear--a study using excremental PCR of nuclear and mitochondrial sequences. *Molecular Ecology*, 4: 95-103

Kohn M.H., York E.C., Kamradt D.A., Haught G., Sauvajot R.M. and Wayne R.K. (1999) Estimating population size by genotyping faeces. *Proceedings. Biological Sciences*, 266: 657-663

Lampa S., Gruber B., Henle K. and Hoehn M. (2008) An optimisation approach to increase DNA amplification success of otter faeces. *Conservation Genetics*, 9: 201-210

Lucchini V., Fabbri E., Marucco F., Ricci S., Boitani L. and Randi E. (2002) Noninvasive molecular tracking of colonizing wolf (*Canis lupus*) packs in the western Italian Alps. *Molecular Ecology*, 11: 857-868

Lukacs P.M. and Burnham K.P. (2005) Review of capture-recapture methods applicable to noninvasive genetic sampling. *Molecular Ecology*, 14: 3909-3919

Marucco F., Pletscher D.H., Boitani L., Schwartz M.K., Pilgrim K.L. and Lebreton J.D. (2009) Wolf survival and population trend using non-invasive capture-recapture techniques in the Western Alps. *Journal of Applied Ecology*, 46: 1003-1010

Massei G., Cowan D.P., Coats J., Gladwell F., Lane J.E. and Miller L.A. (2008) Effect of the GnRH vaccine GonaCon on the fertility, physiology and behaviour of wild boar. *Wildlife Research* 35:540-547

Maudet C., Luikart G., Dubray D., von Hardenberg A. and Taberlet P. (2004) Low genotyping error rates in wild ungulate faeces sampled in winter. *Molecular Ecology Notes*, 4: 772-775

Miller C.R., Joyce P. and Waits L.P. (2005) A new method for estimating the size of small populations from genetic mark–recapture data. *Molecular Ecology,* 14: 1991-2005

Monteiro L., Bonnemaison D., Vekris A., Petry K.G., Bonnet J., Vidal R., Cabrita J. and Megraud F. (1997) Complex polysaccharides as PCR inhibitors in feces: Helicobacter pylori model. *Journal of Clinical Microbiology*, 35: 995-998

Morin P.A., Chambers K.E., Boesch C. and Vigilant L. (2001) Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology*, 10: 1835-1844

Mowat G., Heard D., Seip D., Poole K., Stenhouse G. and Paetkau D. (2005) Grizzly *Ursus arctos* and black bear *U. americanus* densities in the interior mountains of North America. *Wildlife Biology*, 11: 31-48

Mowat G. and Paetkau D. (2002) Estimating marten *Martes americana* population size using hair capture and genetic tagging. *Wildlife Biology*, 8: 201-209

Mowat G. and Strobeck C. (2000) Estimating population size of grizzly bears using hair capture, DNA profiling, and mark-recapture analysis. *Journal of Wildlife Management*, 64: 183-193

Murphy M.A., Kendall K.C., Robinson A., and Waits L.P. (2007) The impact of time and field conditions on brown bear (*Ursus arctos*) faecal DNA amplification. *Conservation Genetics*, 8: 1219-1224

Murphy M.A., Waits L.P., Kendall K.C., Wasser S.K., Higbee J.A. and Bogden R. (2002) An evaluation of long-term preservation methods for brown bear (*Ursus arctos*) faecal DNA samples. *Conservation Genetics*, 3: 435-440

Murphy M.A., Waits L.P., and Kendall K.C. (2000) Quantitative evaluation of fecal drying methods for brown bear DNA analysis. *WIildlife Society Bulletin*, 28: 951-957

Navidi W., Arnheim N. and Waterman M.S. (1992) A multiple-tubes approach for accurate genotyping of very samll DNA samples by using PCR – statistical considerations. *American Jouranl of Human Genetics*, 50: 347-359

Otis D.L., Burnham K.P., White G.C. and Anderson D.R. (1978) Statistical inference for capture- recapture experiments. *Wildlife Monographs*, 62

Paetkau D. (2003) An empirical exploration of data quality in DNA-based population inventories. *Molecular Ecology*, 12: 1375-1387

Piggott M.P. and Taylor A.C. (2003) Remote collection of animal DNA and its applications in conservation management and understanding the population biology of rare and cryptic species. *Wildlife Research*, 30: 1-13

Pollock K.H., Nichols J.D., Hines J.E. and Brownie C. (1990) Statistical inference for capture–recapture experiments. *Wildlife Monographs,* 107

Pompanon F., Bonin A., Bellemain E. and Taberlet P. (2005) Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics*, 6: 847-859

Prugh L.R., Ritland C.E., Arthur S.M. and Krebs C.J. (2005) Monitoring coyote population dynamics by genotyping faeces. *Molecular Ecology*, 14: 1585-1596

Reed J.Z., Tollit D.J., Thompson P.M. and Amos W. (1997) Molecular scatology: the use of molecular genetic analysis to assign species, sex and individual identity to seal faeces. *Molecular Ecology*, 6: 225-234

Roon D.A., Waits L.P. and Kendall K.C. (2005) A simulation test of the effectiveness of several methods for error-checking non-invasive genetic data*. Animal Conservation,* 8: 203-215

Saez-Royela C. and Telleria J.L. (1986) Increased population of wild boar in Europe. *Mammal Review*, 16: 97-101

Scandura M., Iacolina L. and Apollonio M. (2011) Genetic diversity in the European wild boar *Sus scrofa*: phylogeography, population structure and wild x domestic hybridization. *Mammal Review*, 41: 125-137

Schley L., Dufrene M., Krier A. and Frantz A.C. (2008) Patterns of crop damage by wild boar (*Sus scrofa*) in Luxembourg over a 10-year period. *European Journal of Wildlife Research*, 54: 589-599

Scillitani L., Monaco A. and Toso S. (2010) Do intensive drive hunts affect wild boar (*Sus scrofa*) spatial behaviour in Italy? Some evidences and management implications. *European Journal of Wildlife Research*, 56: 307-318

Servanty S., Gaillard J.-M.Toigo C., and Brandt, S. and Baubet, E. (2009) Pulsed resources and climate-induced variation in the reproductive traits of wild boar under high hunting pressure. *Journal of Animal Ecology,* 78: 1278-1290

Siren A., Hamback P. and Machoa E. (2004) Including spatial heterogeneity and animal dispersal when evaluating hunting: a model analysis and an empirical assessment in an Amazonian community. *Conservation Genetics,* 18: 1315-1329

Sodeikat G. and Pohlmeyer K. (2002) Temporary home range modifications of wild boar family groups (*Sus scrofa L.*) caused by drive hunts in Lower Saxony (Germany). *Zeitschrift der Jagdwissenschaften*, 48: 161-166

Solberg K.,H. Bellemain E., Drageset O.M., Taberlet P. and Swenson J.E. (2006) An evaluation of field and non-invasive genetic methods to estimate brown bear (*Ursus arctos*) population size. *Biology Conservation*, 128: 158-168

Taberlet P., Griffin S., Goossens B., Questiau S., Manceau V., Escaravage N., Waits L.P. and Bouvet J. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, 24: 3189-3194

Taberlet P. and Luikart G. (1999) Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*, 68: 41-55

Taberlet P., Waits L.P. and Luikart G. (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology & Evolution*, 14: 323-327

Toigo C., Servanty S., Gaillard J.M., Brandt S. and Baubet E. (2008) Disentangling natural from hunting mortality in an intensively hunted wild boar population. *Journal of Wildlife Management*, 72: 1532-1539

Valière N. (2002) GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes*, 2: 377-379

Van Horn R.C., Altmann J. and Alberts S.C. (2008) Can't get there from here: inferring kinship from pairwise genetic relatedness. *Animal Behaviour* 75 (3):1173-1180

Waits L.P., Luikart G. and Taberlet P. (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology*, 10: 249-256

Wasser S.K., Houston C.S., Koehler G.M., Cadd G.G. and Fain S.R. (1997) Techniques for application of faecal DNA methods to field studies of Ursids. *Molecular Ecology*, 6: 1091-1097

Wasser S.K., Davenport B., Ramage E.R., Hunt K.E., Parker M., Clarke C. and Stenhouse G. (2004) Faeces detection dogs in wildlife research and management: application to grizzly and black bears in the Yellowhead Ecosystem, Alberta, Canada. *Canadian Journal of Zoology* 82: 475-492

Wehausen J.D., Ramey R.R. 2nd and Epps C.W. (2004) Experiments in DNA extraction and PCR amplification from bighorn sheep feces: the importance of DNA extraction method. *Journal of Heredity*, 95: 503-509

Wilson G.J., Frantz A.C., Pope L.C., Roper T.J., Burke T.A., Cheeseman C.L. and Delahay R.J. (2003) Estimation of badger abundance using faecal DNA typing. *Journal of Applied Ecology*, 40: 658-666

Appendix I

Evaluation of faecal storage and DNA extraction methods in wild boar (*Sus scrofa*)

Karolina Kolodziej, Ivan Nikolov, Holger K. Schulz, Kathrin Theissinger, Ralf Schulz

Submitted for second review in *Biochemical Genetics*

Introduction

The wild boar (*Sus scrofa*) is one of the most common and widely distributed ungulates in Europe. Population sizes have been growing rapidly in recent years, leading to agricultural damage and farmer compensation costs (Toigo *et al.* 2008). Furthermore, wild boars play an important role in the transmission of diseases (Fickel & Hohmann 2005). Reliable information on absolute population sizes is of crucial importance for effective wildlife management. However, conventional methods based on hunting harvests, direct sightings or faecal drop counts yield only relative estimates or predictions about population trends. A useful alternative, with great potential as a feasible census method, may be non-invasive genetic sampling without the need of individual capture (Sloane *et al.* 2000; Fickel & Hohmann 2005). Faeces as a DNA source is attractive because of easy sampling and the possibility for an almost equal capture probability (Wehausen *et al.* 2004).

However, faecal genotyping has some pitfalls, including low amplification and genotyping success due to the following possible reasons: 1) low target DNA concentrations (Hajkova *et al.* 2006), 2) contamination with diet and bacterial DNA (Lampa *et al.* 2008), 3) presence of PCR inhibitors (Reed *et al.* 1997) or 4) degradation of DNA (Idaghdour *et al.* 2003). These drawbacks can lead to genotyping errors and biased population size estimations. To reduce these biases several approaches can be applied, e.g., sampling of fresh faeces during winter to reduce DNA degradation (Maudet *et al.* 2004), optimising storage and extraction methods to reduce PCR inhibitors (Murphy *et al.* 2007) and optimising PCR to increase the amplification rate.

Here, we focused on increasing the target DNA concentrations for faecal samples to develop a reliable method for individual identification of wild boars. This method can be used, e.g., for population size estimations, mating system or phylogeography analyses. We evaluated eight combinations of storage and extraction methods by determining amplification success, genotyping success of three microsatellite markers and the target DNA amount using a quantitative PCR (qPCR) assay, which is useful for preselecting good quality faecal DNA extracts (Morin *et al.* 2001; Hausknecht *et al.* 2010).

Material and methods

We collected 141 wild boar faecal samples in a 4000 ha area of the Palatinate Forest in the south-western Germany (49.2°N, 7.8°E) in January 2006. Approximately 2 g of the upper part of each sample was taken using wooden toothpicks and placed in a 4 ml micro tube filled with 2 ml of 99.6% ethanol. The remaining samples were stored in plastic bags and frozen at -20°C until DNA extraction. To prevent contaminations, all DNA extractions were carried out in a designated room that was free of PCR products. A subset of 20 of the 141 faecal samples was randomly chosen for testing the following three storage procedures (1-3) and five extraction methods (a – e):

1) After storing for 24 hours in ethanol, faeces samples were removed from ethanol and divided into six portions à 300 µg each. Three portions were placed again in ethanol for 28 days. Afterwards they were air-dried at room temperature for 24 hours until they were completely dry.

2) The remaining three parts were dried in an exsiccator with silica gel for 28 days;

3) After storing faecal samples for six months at -20°C, 300 µg faeces were processed twice by scrapping the surface, without any preservative agent or drying;

a) QIAamp DNA Stool Mini extraction kit (Qiagen, Hilden, Germany) according to manufacturer's protocol;

b) Nucleo-Spin Tissue Kit (Macherey-Nagel, Düren, Germany) with the following modification: after the first centrifugation step working with supernatant instead of the pellet and incubation with proteinase K at 56 °C overnight;

c) Nucleo-Spin Tissue Kit according to the stool–protocol with the modification of an overnight incubation with proteinase K at 56 °C;

d) Nucleo-Spin Tissue Kit with the following modification: all DNA wash steps were repeated three times; and

e) Nucleo-Spin Tissue Kit with the following modification: after the first centrifugation step the supernatant instead of the pellet was used and all wash steps were repeated three times.

Subsequently, the three storage and five extraction methods were combined, resulting in eight combinations (1a, 1b, 1c, 2a, 2b, 2c, 3d, 3e) summarized in Table 1. To compare the eight storage/ extraction combinations, we genotyped the DNA extracts with three microsatellite markers (CGA, Sw742 and Sw2496; Rohrer *et al.* 1994; Lowden *et al.* 2002; Kolodziej *et al.* 2011). All PCRs were prepared using filter pipette tips on two clean benches (one for the master-mix and one for DNA addition) and were optimised for concentrations of $MgCl_2$ and bovine serum albumin (BSA). The PCR was conducted as described for faecal samples in Kolodziej *et al.* (2011). PCR was performed with a positive control (wild boar tissue) and a negative control (H_2O) and were visualised on an agarose gel to check for the expected target region. PCR products were analysed using a CEQ 8000 sequencer (Beckman Coulter, Krefeld, Germany) and scored with the corresponding software CEQ SYSTEM 9.0 to determine allele lengths. As quality control we independently scored the peaks for a second time in a random order without knowing the peak length from the first scoring. The presence of PCR-product on an agarose gel within the expected size range was counted as amplification success. Allele lengths which could be clearly assigned to a single-

locus genotype were counted as genotyping success. Amplification and genotyping success rates for each storage/extraction combination were calculated across three repeats per microsatellite locus and respective Chi^2 distributions were compared with the corresponding standardized normal distributions using the software IBM SPSS Statistics Base 19 (Chicago, USA).

Table 1: Overview of the eight combinations of storage and extraction methods of wild boar faecal samples tested

Combination	Storage	Extraction
1 a	Ethanol	Qiagen Kit
1 b	Ethanol	MN kit; supernatant; proteinase K overnight
1 c	Ethanol	MN kit; proteinase K overnight
2 a	Ethanol / Silica drying	Qiagen Kit
2 b	Ethanol / Silica drying	MN kit; supernatant; proteinase K overnight
2 c	Ethanol / Silica drying	MN kit; proteinase K overnight
3 d	Direct freezing	MN kit; triplicate wash step
3 e	Direct freezing	MN kit; supernatant; triplicate wash step

MN: Macherey & Nagel

We quantified DNA concentrations via qPCR with a single copy gene primer set TAGLN-Sus (Ebert *et al.* 2012) using a SYBR Green-based assay. Amplifications for qPCR were run twice per sample and combined on a Mastercycler ep realplex (Eppendorf, Wesseling-Berzdorf, Germany). The PCR was carried out in a 10 µl reaction volume containing (final concentration) 5 µl DyNaMoTM Flash SYBR® Green qPCR mastermix (Finnzymes, Vantaa, Finland), 0.2 mM of each primer, 0.1 µg/µl BSA and 1 µl of template. Thermal cycling conditions were as follows: an initial denaturation step at 95°C for 7 min, followed by 40 cycles of 10 s at 95°C,

30 s at 59°C annealing temperature and 20 s at 68°C, with a final melting curve analysis of 15 min (59°C to 95°C) to test if unspecific products were present. We performed the DNA quantification with a standard curve obtained from a wild boar embryo tissue sample of known DNA concentration, which was amplified three times per PCR. The standard curve consisted of six dilutions of the following DNA amounts: 250 ng, 100 ng, 10 ng, 1 ng, 0.1 ng and 0.01 ng. In all qPCR runs a positive control (wild boar tissue) with known DNA concentration and a negative control (H_2O) were additionally used. We quantified the DNA-yield for each run from the slope and Y-intercept of the trendline from the standard curve, which was obtained by plotting the log DNA amounts versus the Ct values, using the following equation: DNA yield = $10^{((Ct - Yint)/slope)}$ (Morin *et al.* 2001). Pearson correlations of amplification and genotyping success rates with DNA amount were calculated with the software IBM SPSS Statistics Base 19. The correlation was calculated using values from each sample and each storage/extraction combination (n = 160) consisting of 20 x 6 for ethanol storage and 20 x 2 for direct freezing.

Results and Discussion

In this study, we were looking for a practical method to maximise the success of faecal DNA extractions. Therefore, we tested various storage and extraction methods for their suitability to increase PCR and genotyping success rates for wild boar faecal DNA.

The storage of faeces is an important factor for inhibiting enzymes that degrade DNA (Beja-Pereira *et al.* 2009). Three types of faecal storage have been

recommended in previous studies: 1) removing water (DET´s buffer: DMSO, EDTA and Tris; Ethanol; Silica), 2) removing cations (Chelex®) and 3) using low temperatures. Murphy *et al.* (2002) found that DNA extraction of brown bear faeces in DET´s buffer and ethanol preservation performed well for the first week after storage, but the duration of storage had a significant negative impact on amplification success. Hence, we did not test the method using DET´s buffer. Frantzen *et al.* (1998) reported the best PCR success rate from faecal samples that were stored in ethanol (60%), dried (67%) and directly frozen (61 %). For that reason, we chose to test storage methods using ethanol, silica drying and direct freezing on wild boar faecal samples. For DNA extraction of forensic samples, many protocols have been reported in previous studies, which were reviewed by Beja-Pereira *et al.* (2009), e.g., phenol-chloroform, Chelex®, guanidinium thiocyanate-silica, the lysis buffer/column purification method and commercial kits. The best results were obtained by using commercial kits (Bhagavatula & Singh 2006; Beja-Pereira *et al.* 2009). Therefore, we tested extraction protocols using two commercial kits: the QIAamp DNA Stool Mini Kit and the Nucleo-Spin Tissue Kit, with several modifications. The manufacturers' protocols state that the pellets formed after centrifugation should be further processed because they contain cells of the study organism and that the supernatant should be discarded. However, when wild boar intestinal epithelia cells are partially destroyed, a certain amount of target DNA will be present in the supernatant; thus, we also used the supernatant.

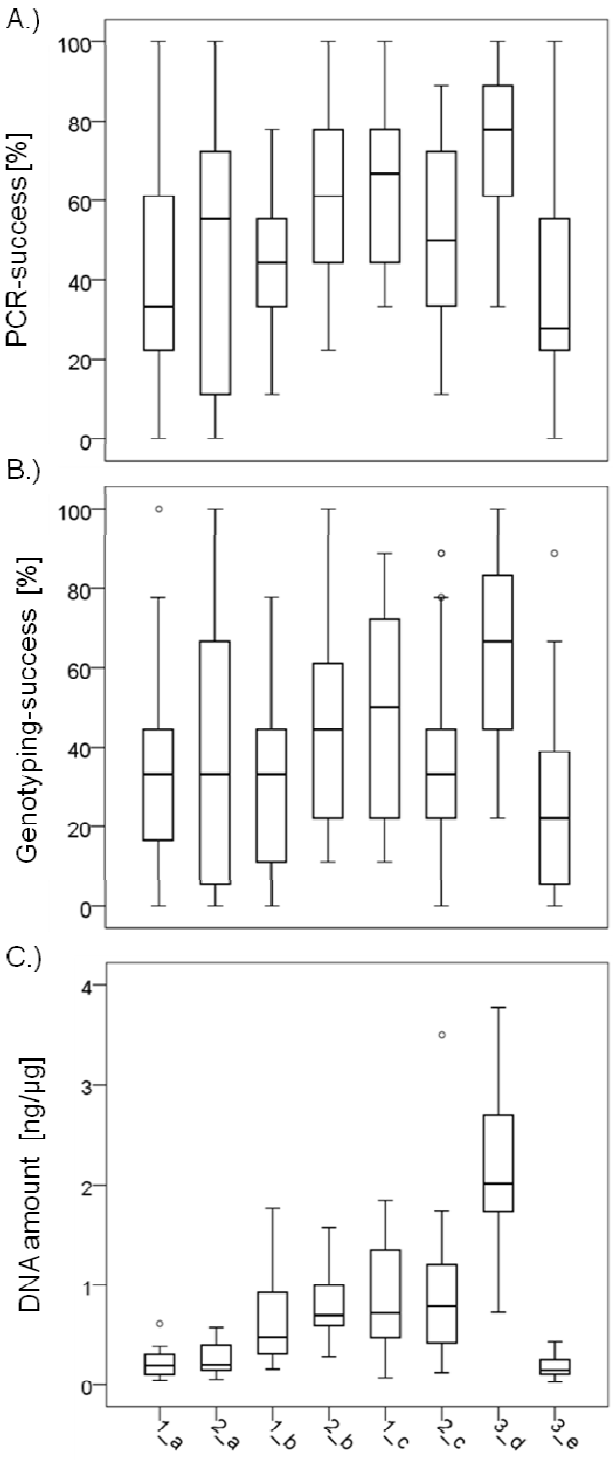

Figure 1: Comparison of eight combinations of storage / extraction protocols using 20 wild boar faecal samples. A) PCR success and B) genotyping success was determined using three microsatellite-markers (CGA, Sw742 and Sw2496) with three repeats per sample and locus. C) Wild boar DNA amount was quantified using quantitative PCR with primer TAGLN-Sus and two repeats per sample.
— Median; □ 25% - 75%; Error bars indicate standard deviation

Across 180 amplifications (consisting of 20 samples, three loci and three repetitions) for all storage/extraction combinations PCR was successful in 27.75% to 77.73% (Figure 1A) whereas the genotyping success rates ranged between 22.2% and 66.62% (Figure 1B). Testing the Chi2 distributions of both PCR and genotyping success provided significant values with χ^2_{PCR} = 72.71, χ^2_{GEN} = 57.97 compared to critical value 38.93 (α = 1%, df = 21; Bosch 2007) which showed that both success rates resulted due to treatment and not to coincidences. All DNA extractions from the pellet showed better results than the supernatant. The QIAamp DNA Stool Mini Kit resulted in intermediate to low success rates for PCR and genotyping and for determining the concentration of wild boar DNA. The method combination 3d (see Table 1) showed the best results for both wild boar DNA concentration and for PCR and genotyping success rates for microsatellite markers (Figure 1). Piggott & Taylor (2003) yielded a similar amplification rate of 70% after direct freezing of faecal samples. However, the poorest result was also associated with an extraction method combined with direct freezing (see method 3a in Figure1), implying that the storage method alone is not a decisive factor for determining the success of DNA extraction and genotyping.

The mean concentration of wild boar DNA yield was by far the highest using method 3d, (2.07 ng/µl), whereas method 3e exhibited the lowest yield, (0.15 ng/µl; Figure 1C). The remaining methods showed lower DNA concentrations, in the range of 0.18 ng/µl to 0.80 ng/µl). We approved that the success rate for PCR and genotyping of the microsatellite markers was positively correlated with the amount of DNA obtained, using qPCR of the single copy gene TAGLN-Sus (for PCR success r = 0.323; p < 0.0001 and for genotyping success r = 0.363;

$p < 0.0001$). Hausknecht *et al.* (2010) reported that qPCR could be applied for evaluating DNA sample quality and preselecting samples suitable for further genotyping analysis. They showed that a higher target DNA amount correlated with a better PCR success rate and a reduction in mismatched alleles; this correlation is supported in our study. Ebert *et al.* (2012) established a DNA concentration threshold for reliable extraction from wild boar faecal samples; all faecal samples with a DNA amount lower than 0.1 ng/µl should not be used for further analysis.

The main conclusion of our study is that PCR and genotyping success can be considerably increased by optimising storage and extraction conditions for non-invasive faecal samples. In the case of wild boar faecal samples, we strongly recommend direct freezing of the collected faeces and extracting the DNA with the Nucleo Spin Tissue Kit with triplicate wash steps. To our knowledge, this method combination has not been recommended or used in previous studies. Nevertheless, for other species we advise to conduct a well-designed comparison and evaluation of methods as described here before starting a non-invasive sampling in wildlife forensics.

Acknowledgments

We thank C. Ebert and T. Schikora for sample collection. We also thank C. Ebert, D. Huckschlag and U. Hohmann for helpful discussions about the manuscript. Furthermore, we wish to thank R. Heydenreich for proofreading the manuscript and statistical advices, and we are grateful to T. Bürgi for technical assistance. This project was supported by the Foundation "Rheinland-Pfalz für Innovation" and the Ministry for Environment, Forestry and Consumer Protection, Rhineland-Palatinate. K.K. was also supported through a PhD scholarship from the Lotto Foundation Rhineland-Palatinate.

References

Beja-Pereira A., Oliveira R., Alves P.C., Schwartz M.K. and Luikart G. (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources,* 9: 1279-1301

Bhagavatula J. and Singh L. (2006) Genotyping faecal samples of Bengal tiger Panthera tigris tigris for population estimation: A pilot study. *Bmc Genetics,* 7: 48

Bosch K. (2007) Statistik für Nichtstatistiker. *Oldenbourg Wissenschaftsverlag,* München

Ebert C., Knauer F., Spielberger B., Thiele B. und Hohmann U. (2012) Estimating wild boar (*Sus scrofa L.*) population size using faecal DNA and capture-recapture modeling. *Wildlife Biology,* in press.

Fickel J. and Hohmann U. (2005) A methodogical apprach for non-invasive sampling for population size estimates in wild boars (*Sus scrofa*). *European Journal of Wildlife Research*, 52: 28-33

Frantzen M.A., Silk J.B., Ferguson J.W., Wayne R.K. and Kohn M.H. (1998) Empirical evaluation of preservation methods for faecal DNA. Molecular *Ecology*, 7: 1423-1428

Hajkova P., Zemanova B., Bryja J., Hajek B., Roche K., Tkadlec E. and Zima J. (2006) Factors affecting success of PCR amplification of microsatellite loci from otter faeces. *Molecular Ecology Notes,* 6: 559-562

Hausknecht R., Bayerl H., Gula R. and Kuehn R. (2010) Application of Quantitative Real- Time Polymerase Chain Reaction for Noninvasive Genetic Monitoring. *Journal of Wildlife Management*, 74: 1904-1910

Idaghdour Y., Broderick D. and Korrida A. (2003) Faeces as a source of DNA for molecular studies in a threatened population of great bustards. *Conservation Genetics*, 4: 789-792

Lampa S., Gruber B., Henle K. and Hoehn M. (2008) An optimisation approach to increase DNA amplification success of otter faeces. *Conservation Genetics,* 9: 201-210

Kolodziej K., Theissinger K., Brün J., Schulz H.K. and Schulz R. (2011) Determination of the minimum number of microsatellite markers for individual genotyping in wild boar (*Sus scrofa*) using a test with close relatives. *European Journal of Wildlife Research*, 58: 621-628

Lowden S., Finlayson H.A., Macdonald A.A., Downing A.C., Goodman S.J., Leus K., Kaspe L., Wahyuni E. and Archibald A.L. (2002) Application of *Sus scrofa* microsatellite markers to wild suiformes. *Conservation Genetics*, 3: 347-350

Maudet C., Luikart G., Dubray D., von Hardenberg A. and Taberlet P. (2004) Low genotyping error rates in wild ungulate faeces sampled in winter. *Molecular Ecology Notes,* 4: 772-775

Morin P.A., Chambers K.E., Boesch C. and Vigilant L. (2001) Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology*, 10: 1835-1844

Murphy M.A., Kendall K.C., Robinson A. and Waits L.P. (2007) The impact of time and field conditions on brown bear (*Ursus arctos*) faecal DNA amplification. *Conservation Genetics*, 8: 1219-1224

Murphy M.A., Waits L.P., Kendall K.C., Wasser S.K., Higbee J.A. and Bogden R. (2002) An evaluation of long-term preservation methods for brown bear (*Ursus arctos*) faecal DNA samples. *Conservation Genetics*, 3: 435-440

Murphy M.A., Waits L.P. and Kendall K.C. (2003) The influence of diet on faecal DNA amplification and sex identification in brown bears (*Ursus arctos*). *Molecular Ecology*, 12: 2261-2265

Piggott M.P. and Taylor A.C. (2003) Extensive evaluation of faecal preservation and DNA extraction methods in Australian native and introduced species. *Australian Journal of Zoology*, 51: 341-355

Reed J.Z., Tollit D.J., Thompson P.M. and Amos W. (1997) Molecular scatology: the use of molecular genetic analysis to assign species, sex and individual identity to seal faeces. *Molecular Ecology*, 6: 225-234

Rohrer G.A., Alexander L.J., Keele J.W., Smith T.P. and Beattie C.W. (1994) A microsatellite linkage map of the porcine genome. *Genetics,* 136: 231-245

Sloane M.A., Sunnucks P., Alpers D., Beheregaray L.B. and Taylor A.C. (2000) Highly reliable genetic identification of individual northern hairy-nosed wombats from single remotely collected hairs: a feasible censusing method. *Molecular Ecology*, 9: 1233-1240

Sweitzer R.A., Van Vuren D., Gardner I.A., Boyce W.M. and Waithman J.D. (2000) Estimating sizes of wild pig populations in the North and Central Coast regions of California. *Journal of Wildlife Management*, 64: 531-543

Toigo C., Servanty S., Gaillard J.M., Brandt S. and Baubet E. (2008) Disentangling natural from hunting mortality in an intensively hunted wild boar population. *Journal of Wildlife Management*, 72:1532-1539

Wehausen J.D., Ramey R.R. 2nd and Epps C.W. (2004) Experiments in DNA extraction and PCR amplification from bighorn sheep feces: the importance of DNA extraction method. *Journal of Heredity*, 95: 503-509

Appendix II

Determination of the minimum number of microsatellite markers for individual genotyping in wild boar (*Sus scrofa*) using a test with close relatives

Karolina Kolodziej, Kathrin Theissinger, Jörg Brün, Holger K. Schulz, Ralf Schulz

Published in *European Journal of Wildlife Research* 58: 621-628

Abstract

In the context of developing a non-invasive, practicable method for population size estimation in wild boar, we present a stepwise procedure to reduce the number of required microsatellite markers for individual genotyping. Step1: An initial marker set of 12 microsatellite loci was tested for species specificity with non-target DNA and resulted in an exclusion of two markers. Step 2: A variability test regarding heterozygosity and deviations from Hardy Weinberg equilibrium led to the rejection of two further markers. Step 3: The remaining eight markers were tested for transferability across populations with three separate wild boar sample sets. Step 4: On the basis of probability of identity values a reduction from eight to five markers was possible. Step 5: A novel test using tissue samples from female wild boars and their embryos provided evidence that four variable microsatellite markers and one sex-marker are sufficient for individual identification of close

relatives. Step 6: Faeces samples were finally used to estimate PCR (PS) and genotyping success (GS). In conclusion, we recommend a specific four marker combination with both PS and GS > 50% for a reliable individual identification in non-invasive population size estimation of wild boar.

Introduction

In Europe, population sizes of wild boar (*Sus scrofa*) have been rapidly growing during the past three decades, causing increased agricultural damages and costs of compensation to farmers (Toigo *et al.* 2008). Population size estimation is inevitable for wildlife management, though often difficult to obtain (Valière *et al.* 2007). Traditional methods such as direct sightings, faecal drop counts, or hunting harvest result in relative estimates and population trends. Non-invasive genetic sampling methods (i.e. faeces, hairs, feathers) and molecular techniques for individual genotyping have increased within the last years, providing a more accurate, indirect way of population size estimation (Adams & Waits 2007). However, non-invasive sampling is often associated with technical problems due to low DNA quality and quantity, leading to genotyping errors (Broquet *et al.* 2007) and hence biased population size estimations (Hoffman & Amos 2005). It is therefore indispensable to assure maximized genotyping reliability.

The choice and number of microsatellite markers is of prime importance because it has consequences for all subsequent analyses (Taberlet & Luikart 1999; Broquet *et al.* 2007). The use of too many markers can increase genotyping errors, leading to false genotypes and overestimations of population sizes (Creel

et al. 2003). In contrast, using too few or insufficient variable markers can lead to underestimations of individuals and hence population sizes (Knapp *et al.* 2009). Ideally, a microsatellite locus should exhibit an expected heterozygosity (H_e) between 0.6 and 0.8 to provide best resolution (Taberlet & Luikart 1999). To define the minimum number of loci required for reliable multilocus genotyping, Waits *et al.* (2001) have developed the measure probability of identity (P_{ID}), i.e. the probability that two individuals drawn at random from a population will have the same genotype at multiple loci. The smaller the P_{ID} value the more informative and polymorphic the locus (Waits *et al.* 2001). The product P_{ID}sib is a conservative upper bound of the number of loci necessary to distinguish individuals and accounts for kinship in a population. It indicates the ability of a set of microsatellite loci to resolve between different individuals, including relatives and siblings (Woods *et al.* 1999; Mills *et al.* 2000). Waits & Leberg (2000) demonstrated that genotyping seven to ten loci can overestimate populations up to 200% because of increasing genotyping errors. This can be avoided by reducing the number of microsatellite loci to a minimum.

Here, we propose stepwise protocol for maximum reduction of the required microsatellite loci to resolve between wild boar individuals for non-invasive genotyping. We tested 12 commonly used wild boar microsatellite markers (Vernesi *et al.* 2003; Delgado *et al.* 2008; Poteaux *et al.* 2009) for species specificity, and calculated PCR and genotyping success as well as P_{ID}sib to define a minimum marker set suited for population size estimation using non-invasive sampling. We verified our results by transferring the determined minimum marker

set onto three disconnected wild boar populations and by using a novel test with closely related individuals.

Material and methods

Sampling localities

Samples were collected from three separate wild boar populations across Germany: Rhineland-Palatinate (RP; n = 420, south-western Germany, Palatinate Forest, 49.2°N, 7.8°E), Lower Saxony (LS; n = 100, north-eastern Germany, for details see Gethoffer *et al.* 2007), and Mecklenburg–Western Pomerania (MWP; n = 93, north-eastern Germany, for details see Keuling *et al.* 2008). For the populations RP and LS tissue samples were obtained by muscle biopsies of hunting bags in the hunting seasons from 2005 to 2007. For the population MWP hair samples (n = 75) and muscle biopsies (n = 18) were collected from 2002 to 2006. Hair samples were preserved in paper envelopes and tissue samples in denatured 99.6% ethanol. All samples were stored at 4°C. Additionally, wild boar fecal droppings (n = 20) were collected in the RP population in December 2006, stored in plastic bags, and directly frozen at -20°C until DNA extraction.

Laboratory procedures

A subsample of the RP population (RP.1; n = 44) was processed in the laboratories of the University of Koblenz-Landau, Germany. DNA extractions were conducted according to a standard phenol-chloroform protocol (Sambrook & Russel 2001). PCR was carried out in 13 µL reaction volume containing (final

concentration) 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.2 µM of each primer, 0.023 µg/µL bovine serum albumin (BSA), 0.02 u/µL Taq-DNA polymerase (Axon Labortechnik), and 1 µL template. Cycling conditions were as follows: an initial denaturation step for 3 min at 94°C, followed by 40 cycles of 30 s at 94°C, 30 s at locus specific annealing temperature and 30 s at 72°C, and a final elongation for 10 min at 72°C. PCR products were run on a CEQ 8000 sequencer (Beckman Coulter) and analysed on the corresponding software CEQ SYSTEM 9.0 to determine allele lengths.

The remaining RP samples (RP.2; n = 376), as well as the populations LS and MWP, were processed in the laboratories of the University of Bonn, Germany. Whole genomic DNA of hair and tissue samples was extracted using the NucleoSpin Tissue–Kit (Macherey-Nagel) for the MWP population, and the CHELEX method (Walsh *et al.* 1991) for the LS and RP.2 populations, following the respective manufacturer protocols. PCR was carried out in 10 µL reaction volume containing (final concentration) 0.2 mM dNTPs, 0.3 µM of each primer, 2 µL enhancer-solution P/Y (Peqlab), 0.017 U/µL Taq polymerase (Invitrogen) and 2 µL template DNA. Cycling conditions were as follows: initial denaturing step for 3 min at 94°C, followed by 30-35 steps for 30 s at 94°C, for 30 s at locus specific annealing temperature, for 30 s at 72°C, and a final elongation step for 10 min at 72°C. PCR products were run on an ABI Prism 377 automatic sequencer (Applied Biosystems) and analysed with the program GeneScan 2.1 (Applied Biosystems).
Faeces samples were processed in the laboratories of the University of Koblenz-Landau, Germany. About 250 mg of the dropping surface was used without drying for DNA-isolation. DNA from 20 wild boar faeces samples were extracted using the

Nucleo-Spin Tissue Kit (Macherey Nagel) according to the manufacturer's protocol, with a triplicate wash step. PCR was conducted in 15 µL containing (final concentration) 2.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.2 µM of each primer, 0.1 µg/µL BSA, 0.025 u/µL Go Taq Flexi DNA polymerase (Promega) and 1 µL template DNA. Cycling conditions were as follows: initial denaturation for 2 min at 95°C, followed by 40 cycles of 30 s at 95°C, 30 s at locus specific annealing temperature, and 30 s at 72°C, and a terminal elongation for 5 min at 72°C. PCR products were analysed as described for the RP.1 population.

To test for wild boar specificity, the chosen markers were tested with non-target DNA from roe deer and mouse tissue samples taken from ear biopsies. Wild boars feed on the carrion of these species and thus traces of alien DNA could be potentially present in wild boar faeces (Briedermann 1990). Furthermore, to exclude human contaminations we conducted the specificity test with human DNA obtained by buccal swabs. Extraction and PCR were conducted as described for the RP.1 population.

For the minimum marker test with close relatives, tissue samples taken from three pregnant wild boars (obtained from a driving hunt in 2008 in the Palatinate Forest) and their embryos (n = 23) were analysed. Embryo samples were taken from the internal organs. Female samples were taken from ear biopsies. All tissue samples were washed with 3 mL 1x phosphate-buffered saline before DNA-isolation. The sex marker (PigSRY) was combined with a microsatellite locus (TNFB) with similar annealing temperature and different size range as control marker to avoid bias due to PCR failures in males. The gender was determined on an agarose gel.

Statistical analyses

We selected 12 microsatellite markers designed for *Sus scrofa domestica* (see Table 5.1) by means of the respective number of alleles according to the literature (Rohrer *et al.* 1994; Alexander *et al.* 1996; Laval *et al.* 2000; Lowden *et al.* 2002), and the y-linked species-specific sex marker PigSRY (Kawarasaki *et al.* 1995). For faeces samples (n = 20), PCR success was calculated across five repeats per microsatellite locus and three repeats for the sex marker. The genotyping success was calculated by counting clearly assigned alleles within a single-locus genotype. Across the five repeats the inferred alleles had to be identical at least two times for heterozygotes and at least three times for homozygotes (Frantz *et al.* 2003; Arrendal *et al.* 2007).

Since the RP.1 and the RP.2 samples were processed in different labs and on different systems we could not combine the allelic data directly, but rather used the RP.1 data set as a reference population for initial loci selection and for subsequent population comparisons. For RP.1 the 12 loci were tested for linkage disequilibrium (LD) in GENEPOP 4.0.10 (Raymond & Rousset 1995) and for Hardy Weinberg equilibrium (HWE) and null allele frequencies (N_A) in CERVUS 3.0 (Kalinowski *et al.* 2007). Expected (H_e) and observed (H_o) heterozygosities were calculated with the program GIMLET (Valière 2002). Loci showing H_e and H_o within the preferred range of 0.6 and 0.8, indicating sufficient variability (Taberlet & Luikart 1999), were used for subsequent analyses. To test for flexibility and reproducibility of these eight markers the genotypic data of three separate wild boar populations (RP, LS, and MWP) were compared regarding H_e, H_o, HWE, as

well as by their inbreeding coefficient F_{IS} calculated in GENEPOP 4.0.10 (Raymond and Rousset 1995).

The product P_{ID}sib for all sets of loci was calculated across all populations with the program GIMLET (Valière 2002) to determine the minimum number of markers required for individual identification, using default parameters. GIMLET successively adds one locus according to its H_e, recalculates the P_{ID}sib each time, and hence provides an estimate for the required minimum marker set. To resolve between siblings in wildlife forensics, the product of P_{ID}sib should not be higher than 0.01 (Waits *et al*. 2001). This range was reached with a minimum of five loci. To test if a further reduction from five to four markers is still reliable, genotypic data from three pregnant wild boars and their embryos (n = 23) were analysed. Samples were initially genotyped across eight loci. Subsequently, 70 combinations of four-locus genotypes were tested for resolution among 23 individuals, using the modified EXCEL sheet GENECAP (Wilberg & Dreher 2004). To test if the relatedness of these individuals is conform to an *in vivo* situation and if the populations exhibit a comparable relatedness we used the unbiased r_{xy} statistics to calculate the relatedness coefficient by Queller & Goodnight (1989). We calculated a matrix of pairwise relatedness across all family groups (FG; n = 23), within (FGW) and among (FGA) the three family groups, and within each population (RP.1, RP.2, LS, MWP) with the program GenAlex 6 (Peakall & Smouse 2005). Tests for differentiation between populations were conducted with the software package R 2.9.1 (R-Development-Core-Team 2009). Linear mixed effect models (LME), package NLME (Pinheiro *et al.* 2009) were fitted to the measure differences in the Queller & Goodnight relatedness mean values (QGM) between

populations and FG while QGM was used as dependent variable and FG as explanatory variable. Animal identity was entered as a random factor. We first fitted a full model including all groups. Post-hoc, pairwise comparisons between the respective groups were also conducted. Reported p-values refer to the increase in deviance in model fit when the respective variable was removed (likelihood-ratio-tests lrt). To test differences between RP.1 and the family groups as well as within and among family group members a two-sample randomization tests between QGM means (10.000 iterations, Pop-Tools 3.2.2 Hood 2010) have been conducted.

Results and discussion

The selected 12 loci were polymorphic and exhibited 3 to 11 alleles among 44 RP.1 individuals. H_o ranged from 0.34 to 0.84; H_e ranged from 0.45 to 0.83. Only one locus (Sw936) exhibited deviations from HWE (see Table 1). No signs of null alleles were detected. Seven loci (S0005, CGA, Sw2496, S0068, Sw742, Sw461, Sw2021) were significantly linked with one or more loci across all populations. Nevertheless, due to different chromosome locations (see Table 1) we treated all markers as separate loci (Iacolina *et al.* 2009). In the following we present our results on the basis of a stepwise procedure to reduce the number of microsatellite loci for reliable individual genotyping in wild boar.

Table 1: Summary information of locus specific data based on the RP.1 population. Presented are the results of the specificity test with mouse (M), roe deer (RD), and human (H) DNA for the initial 12 microsatellite markers and the sex marker PigSRY. Moreover, expected (H_e) and observed (H_o) heterozygosities, Hardy Weinberg Equilibrium (HWE), chromosome number (Chr) of microsatellite location, allele size ranges and allele numbers (A) per locus are given. PCR (PS) and genotyping (GS) success with faeces samples are presented.

Locus	M	RD	H	H_e	H_o	HWE	Chr	size range [bp]	A	PS [%]	GS [%]
Sw742	-	+*	-	0.83	0.75	Ns	16	193-231	9	70	60
CGA	-	-	-	0.82	0.84	Ns	1	250-310	11	47	25
S0068	-	+*	-	0.82	0.77	Ns	13	211-281	9	37	20
S0005	-	-	-	0.80	0.82	Ns	5	205-261	11	59	45
Sw461	+*	+*	-	0.80	0.82	Ns	2	118-150	11	42	25
Sw2496	-	-	-	0.73	0.61	Ns	14	184-228	8	83	60
Sw2021	-	-	-	0.71	0.61	Ns	3	102-132	6	71	70
TNFB	+*	+*	-	0.70	0.80	Ns	7	170-212	6	69	65
Sw841	+*	+*	+	0.65	0.50	Ns	4	156-184	6	/	/
Sw936	-	+*	-	0.59	0.43	*	15	90-118	4	/	/
Sw957	+	+*	+*	0.55	0.43	Ns	12	115-157	3	/	/
Sw949	-	-	-	0.45	0.34	Ns	24	178-204	5	/	/
PigSRY	-	-	-	/	/	/	/	236	/	90	/

- no allele found, +* allele out of size range as described in literature, and differing peak morphology, + allele size found as described in literature; / no results; * $p < 0.05$, ns not significant; H_e, H_o, and HWE were calculated based on the RP subsample (N = 44)

Step 1: Specificity test. The results of the specificity test are summarized in Table 1. Five loci (CGA, S0005, Sw2496, Sw2021, and Sw949) and the sex marker PigSRY showed neither signals for mouse, roe deer, nor human DNA, while five other loci (Sw742, S0068, Sw936, Sw461 and TNFB) amplified a fragment in roe deer-DNA and/or mouse-DNA. However, these signals could be neglected because they were not within the allele size range expected for wild boar and did not show locus specific microsatellite patterns. The two remaining loci though (Sw841 and Sw957) amplified fragments for human and mouse-DNA within the

size-range of wild boars and were thus excluded from further procedures. This specificity test demonstrates that even species specific designed primer pairs can produce signals for non-target DNA, possibly leading to inaccurate allele counts. Therefore, we recommend to generally conduct a specificity test before starting a wildlife forensic study.

Step 2: Variability test. For the remaining ten loci the RP.1 samples exhibited H_e and H_o values ranging from 0.43 to 0.83 and from 0.34 to 0.84, respectively (Table 1). For eight markers (Sw742, CGA, S0068, S0005, Sw461, Sw2496, Sw2021, and TNFB) H_e and H_o ranked within the preferred range of 0.6 and 0.8 and were therefore selected for further testing.

Table 2: Comparison of expected (H_e) and observed (H_o) heterozygosities and Hardy Weinberg Equilibrium (HWE) for eight microsatellite loci, as well as inbreeding coefficient (F_{IS}) among different wild boar populations.

Pop	RP.1			RP.2			LS			MWP		
F_{is}	0.0350			0.0131			0.2213			0.0971		
Locus	H_e	H_o	HWE	H_e	H_o	HWE	H_e	H_o	HWE	H_e	H_o	HWE
Sw742	0.83	0.75	ns	0.85	0.86	ns	0.79	0.64	ns	0.78	0.78	Ns
CGA	0.82	0.84	ns	0.84	0.78	ns	0.85	0.85	ns	0.85	0.90	Ns
S0068	0.82	0.77	ns	0.83	0.83	ns	0.84	0.67	***	0.66	0.65	Ns
S0005	0.80	0.82	ns	0.87	0.74	***	0.88	0.80	ns	0.82	0.86	Ns
Sw461	0.80	0.82	ns	0.86	0.85	ns	0.75	0.61	ns	0.60	0.67	Ns
Sw2496	0.73	0.61	ns	0.76	0.74	ns	0.76	0.55	***	0.80	0.76	Ns
Sw2021	0.71	0.61	ns	0.73	0.68	ns	0.76	0.35	***	0.71	0.10	***
TNFB	0.70	0.80	ns	0.72	0.70	ns	0.68	0.48	***	0.77	0.72	Ns
mean	0.78	0.75		0.81	0.77		0.79	0.62		0.75	0.68	

ns not significant, * $p < 0.01$, *** $p < 0.0001$

Step 3: Transferability test. To test if these eight markers could be transferred onto other wild boar populations with similar results, we compared the genetic variation for geographically disconnected populations (RP. 2, LS, MWP). The results are presented in Table 2. H_e values ranged from 0.60 to 0.86 and did not differ significantly across populations. Moreover, between the two subsamples from the same population (RP.1, RP.2), which were analysed in different laboratories, the mean difference of H_e and H_o was below 0.03, indicating high transferability of the markers even across different laboratory systems. However, significant deviations from HWE due to homozygote excess were detected at one locus (S0005) in RP.2, at one locus (Sw2021) in MWP, and at four loci (S0068, Sw2496, Sw2021 and TNFB) in the LS population (Table 2). The deviation from HWE at one locus can be neglected due to possible occurrence of natural selection acting on a nearby gene (Kalinowski *et al.* 2007). A deviation from HWE by more than two loci could be an indicator for a substructure of a population causing a Wahlund effect (Cornuet & Luikart 1995). Nevertheless, all loci except for Sw2021 only in population MWP and LS, exhibited $N_A < 0.2$ (data not shown), indicating that null alleles should not bias our results (Dakin & Avise 2004). Since in another study (Vernesi *et al.* 2003) Sw2021 was successfully applied in two populations, with heterozygosities between 0.90 and 0.91 and without HWE deviations, we decided to use this marker. However, we recommend to initially test the locus Sw2021 for potential null alleles due to the increased null allele frequencies in two of four populations. The QGM means for RP.1 was -0.0236 ± 0.1946, for RP.2 -0.0025 ± 0.1810, for LS -0.0101 ± 0.1948, and for MWP -0.0109 ± -0.0109 (Figure 1A).

The population showed no significant differences, with the exception of RP.1 and RP.2 ($p < 0.01$; Figure 1A). This discrepancy is probably due to different sample sizes (RP.1: n = 44; RP.2: n = 376). While QGM of locally collected samples may exceed the population mean QGM of more regionally dispersed samples should decline due to the lack of collecting close relatives by chance. These results support the transferability of the chosen marker set across separate populations and laboratories.

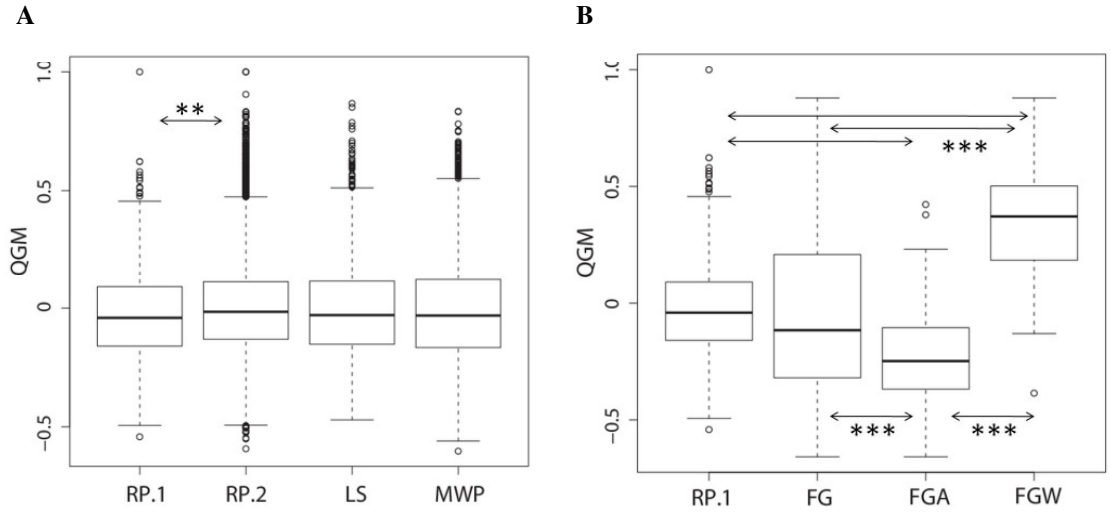

Figure 1: Queller and Goodnight relatedness mean values (QGM). A: QGM of three populations (MWP, LS, RP.1 and RP.2). B: QGM across family groups (FG); among family groups (FGA); within family groups (FGW) and within the reference population (RP.1); * $p < 0.01$, *** $p < 0.0001$

Step 4: Product P_{ID}sib. The P_{ID} for each locus calculated from RP.1 individuals are presented in Table 1. To obtain the minimum number of required microsatellite loci the product P_{ID}sib should not be higher than 0.01 (Waits *et al.* 2001). This applies to the combination of the first five microsatellite markers (Sw742, CGA, S0068, S0005, Sw461, Table 1). Therefore the number of markers could be reduced from eight to five. To check if this test could be transferred onto other populations we

calculated P_{ID}sib for three other populations (Table 3). The minimum number of required markers is five for all analysed populations. Thus, the reduction from five to four markers according to P_{ID}sib could be transferred in other wild boar populations.

Table 3: Comparison of products of P_{ID} among siblings (P_{ID}sib) across different wild boar populations.

RP.1		RP.2		LS		MWP	
Sw742	3.48E-01	S0005	3.24E-01	S0005	3.19E-01	CGA	3.32E-01
CGA	1.22E-01	Sw461	1.05E-01	CGA	1.06E-01	S0005	1.18E-01
S0068	4.36E-02	Sw742	3.56E-02	S0068	3.64E-02	Sw2496	4.34E-02
S0005	1.59E-02	CGA	1.21E-02	Sw742	1.36E-02	Sw742	1.65E-02
Sw461	5.85E-03	S0068	4.19E-03	Sw2496	5.30E-03	TNFB	6.37E-03
Sw2496	2.42E-03	Sw2496	1.64E-03	Sw2021	2.08E-03	Sw2021	2.74E-03
Sw2021	1.04E-03	Sw2021	6.77E-04	Sw461	8.20E-04	S0068	1.25E-03
TNFB	4.48E-04	TNFB	2.87E-04	Sw936	3.60E-04	Sw461	6.32E-04

Step 5: Test with close relatives. An optional step to test the validity of a reduction from five to four markers was a test with close relatives, i.e. three females and their embryos. Using the eight loci selected in step 2, 70 four-loci genotype combinations with gender information were compared for individual identification. The number of 32 four-loci combinations (Table 4) showed sufficient resolution between 23 closely related individuals, i.e. none of the 23 four-loci genotypes were identical. For eight combinations (14, 17, 19, 21, 24, 26, 29, 30, Table 4) the sex-marker was even redundant. These results demonstrate that varying combinations of four microsatellite loci and an additional sex marker might be sufficient to

resolve between closely related wild boar individuals, such as siblings and parents.

To test if the chosen sample of close relatives gives similar relatedness characteristics as in natural wild boar population, the relatedness coefficient was calculated for the three females with their embryos and a reference population (RP.1). The QGM mean for FG was -0.0455 ± 0.3377, for WFG 0.342 ±0.237, for AFG -0.235 ± 0.193 for RP.1 -0.0236 ± 0.1946 (see Figure 1B). All values were significantly different ($p < 0.001$) except for RP.1 and FG ($p > 0.183$) in the two-sample randomization tests. Compared to the study of Iacolina *et al.* (2009) our QGM values of our reference population and close relatives (FG) showed similar values to nongroups (wild boars moving separately in an area) or individuals with different age class associations. This corresponds to an open population similar to an *in vivo* situation. A relatedness coefficient of 0.5 is expected between parents and among full siblings but deviations from the theoretical value is quite common and inherent to the system (e.g. Queller & Goodnight 1989, Csilléry *et al*. 2006, Van Hoorn *et al.* 2008). Further on sample size, chance as well as multiple paternities in the litter may influence relatedness values. However all values are in an expected range and comparable to other studies and there is no significant difference between the distribution of relatedness values between the full-family groups and RP.1.

Table 4: Thirty-two four-loci combinations and additional sex marker (PigSRY) for reliable resolution of closely related wild boar individuals. Combinations are arranged by increasing product P_{ID}sib calculated per each combination by allele frequencies from the closely related individuals (n = 23)

	Sw742	CGA	S0068	S0005	Sw461	Sw2496	Sw2021	TNFB	PigSRY	P_{ID} sib
1	x	x		x	x				x	0.010
2		x	x	x	x				x	0.011
3		x	x		x			x	x	0.011
4		x		x	x			x	x	0.011
5	x	x	x		x				x	0.012
6	x		x	x	x				x	0.012
7	x	x	x	x					x	0.013
8		x		x	x		x		x	0.013
9		x	x		x		x		x	0.014
10	x	x			x		x		x	0.014
11	x	x	x				x		x	0.015
12	x	x		x			x		x	0.015
13	x			x	x		x		x	0.015
14		x		x	x	x				0.016
15		x	x		x	x			x	0.016
16	x		x	x			x		x	0.016
17	x	x			x	x				0.017
18	x	x			x			x	x	0.017
19	x	x		x		x				0.018
20	x			x	x	x			x	0.018
21	x		x		x	x				0.018
22	x	x		x				x	x	0.019
23	x			x	x			x	x	0.019
24		x			x	x	x			0.020
25		x			x		x	x	x	0.020
26	x	x				x	x			0.022
27		x		x			x	x	x	0.022
28	x	x					x	x	x	0.023
29	x		x			x	x			0.024
30		x			x	x		x		0.024
31	x			x			x	x	x	0.025
32*	x					x	x	x	x	0.026

* Four-loci combination recommended for faeces samples.

Step 6: Non-invasive samples. To test the reliability of the inferred marker set for non-invasive samples, the PCR (PS) and genotyping success (GS) was determined across faeces samples (n = 20) of the RP population (Table 1). PS ranged between 37% and 83% GS ranged between 25% and 70%. Four loci (Sw742, Sw2496, Sw2021, TNFB) showed a PS and GS above 50%, which is rather the upper bound in many wildlife forensic projects (see Broquet *et al.* 2007 for review). Accordingly, we recommend only the four-loci combination (No. 32, Table 4) with PS and GS above 50% for individual non-invasive genotyping of wild boars in the RP population. To transfer this combination onto other wild boar populations we recommend an amplification and genotyping test for faeces as described above.

In most other studies on wildlife forensics six to ten microsatellite markers are commonly used (e.g. Wilson *et al.* 2003; Hajkova *et al.* 2009; Marucco *et al.* 2009). However, a higher number of microsatellite markers increase the potential genotyping error rate. A single-locus error rate of 1% would add up to 10% using ten loci (Taberlet & Luikart 1999). Considering the maximum threshold of 5% genotyping errors for population size estimation (Lukacs & Burnham 2005), it could be one way to minimize potential error sources by reducing the number of microsatellite markers. The calculation of product P_{ID}sib gives an estimation of the number of loci necessary to distinguish between individuals with sufficient discriminating power. Our test using closely related individuals suggests further reduction of the required number of microsatellite markers for individual identification to a minimum of four loci in varying combinations.

The presented stepwise procedure of marker reduction and the test with close relatives could also be applied to other species, where it is possible to obtain pregnant females by regular hunting, e.g. in red deer. An additional advantage of the marker reduction could be the decrease of laboratory costs especially for single PCR methods, but also for multiplex PCR-methods due to the easy combination of four fluorescent dyes.

Acknowledgements

We thank C. Ebert, D. Huckschlag and U. Hohmann from the Research Institute of Forest Ecology and Forestry, Rhineland-Palatinate and G. Sodeikat and O. Keuling from the Institute of Wildlife Research in Hannover for the collection of samples. Furthermore, we thank T. Bürgi for technical advices and C. Wallnisch, J. Schürings and B. Müller for lab assistance as well as S. Baldauf for statistical support with the lme models. This project was supported by the Foundation „Rheinland-Pfalz für Innovation“, and the Ministry for Environment and Forestry, Rhineland-Palatinate. KK was supported through a two year PhD scholarship from the Lotto Foundation Rhineland-Palatinate.

References

Adams J. and Waits L.P. (2007) An efficient method for screening faecal DNA genotypes and detecting new individuals and hybrids in the red wolf (Canis *rufus*) experimental population area. *Conservation Genetics,* 8: 123-131

Alexander L.J., Rohrer G.A. and Beattie CW (1996) Cloning and characterization of 414 polymorphic porcine microsatellites. *Animal Genetics*, 27: 137-148

Arrendal J., Vila C. and Bjorklund M. (2007) Reliability of noninvasive genetic census of otters compared to field censuses. *Conservation Genetics*, 8: 1097-1107

Briedermann L. (1990) Schwarzwild. *Deutscher Landwirtschaftsverlag*, Berlin, Germany

Broquet T., Menard N. and Petit E. (2007) Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics*, 8: 249-260

Cornuet J.M. and Luikart G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, 144: 2001-2014

Creel S., Spong G., Sands J.L., Rotella J., Zeigle J., Joe L., Murphy K.M. and Smith D. (2003) Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Molecular Ecology*, 12: 2003-2009

Csilléry K., Johnson T., Beraldi D., Clutton-Brock T., Coltman D., Hansson B., Spong G. and Pemberton J.M. (2006) Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics*, 173: 2091-2101

Dakin E.E. and Avise J.C. (2004) Microsatellite null alleles in parentage analysis. *Heredity*, 93: 504–509

Delgado R., Fernandez-Llario P., Azevedo M., Beja-Pereira A. and Santos P. (2008) Paternity assessment in free-ranging wild boar (*Sus scrofa*) - Are littermates full-sibs? *Mammalian Biology*, 73: 169-176

Frantz A.C., Pope L.C., Carpenter P.J., Roper T.J., Wilson G.J., Delahay R.J. and Burke T. (2003) Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Molecular Ecology*, 12: 1649-1661

Hajkova P., Zemanova B., Roche K. and Hajek B. (2009) An evaluation of field and noninvasive genetic methods for estimating Eurasian otter population size. *Conservation Genetics*, 10: 1667-1681

Hoffman J.I. and Amos W. (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, 14: 599-612

Hood, G.M. (2010) PopTools version 3.2.3. Available on the internet. URL http://www.poptools.org

Iacolina L., Scandura M., Bongi P. and Apollonio M. (2009) Nonkin associations in wild boar scocial units. *Journal of Mammalogy*, 90: 666-674

Kalinowski S.T., Taper M.L. and Marshall T.C. (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099-1006

Kawarasaki T., Kohsaka T., Sone M., Yosshida M. and Bamba K. (1995) Detection of Y-bearing porcine spermatozoa by *in-situ* hybridization using digoxigenin-labeled, porcine male-specific DNA-probe produced by polymerase chain-reaction. *Molecular Reproduction and Development* 40: 455-459

Keuling O., Stier N. and Roth M. (2008) Annual and seasonal space use of different age classes of female wild boar *Sus scrofa* L. *European Journal of Wildlife Research*, 54: 403-412

Knapp S.M., Craig B.A. and Waits L.P. (2009) Incorporating Genotyping Error Into Non-Invasive DNA-Based Mark-Recapture Population Estimates. *Journal of Wildlife Management,* 73: 598-604

Laval G., Iannuccelli N., Legault C., Milan D., Groenen M.A., Giuffra E., Andersson L., Nissen P.H., Jorgensen C.B., Beeckmann P., Geldermann H., Foulley J.L., Chevalet C. and Ollivier L. (2000) Genetic diversity of eleven European pig breeds. *Genetics Selection Evolution*, 32: 187-203

Lowden S., Finlayson H.A., Macdonald A.A., Downing A.C., Goodman S.J., Leus K., Kaspe L., Wahyuni E. and Archibald A.L. (2002) Application of *Sus scrofa* microsatellite markers to wild suiformes. *Conservation Genetics*, 3: 347-350

Lukacs P.M. and Burnham K.P. (2005) Review of capture-recapture methods applicable to noninvasive genetic sampling. *Molecular Ecology*, 14: 3909-3919

Marucco F., Pletscher D.H., Boitani L., Schwartz M.K., Pilgrim K.L. and Lebreton J.D. (2009) Wolf survival and population trend using non-invasive capture-recapture techniques in the Western Alps. *Journal of Applied Ecology*, 46: 1003-1010

Mills L.S., Citta J.J., Lair K.P., Schwartz M.K. and Tallmon D.A. (2000) Estimating animal abundance using noninvasive DNA sampling: Promise and pitfalls. *Ecological Applications*, 10: 283-294

Peakall R. and Smouse P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295

Pinheiro J., Bates D., DebRoy S., Sarkar D. and R-Core-team (2009) nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-92

Poteaux C., Baubet E., Kaminski G., Brandt S., Dobson F.S. and Baudoin C. (2009) Socio-genetic structure and mating system of a wild boar population. *Journal of Zoology,* 278: 116-125

Queller D.C. and Goodnight K.F. (1989) Estimating Relatedness Using Genetic Markers. *Evolution* 43: 258-275

Raymond M. and Rousset F. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Heredity*, 86:248-249

R-Development-Core-Team (2009) R a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rohrer G.A., Alexander L.J., Keele J.W., Smith T.P. and Beattie C.W. (1994) A microsatellite linkage map of the porcine genome. *Genetics*, 136: 231-245

Sambrook J. and Russel D.W. (2001) Molecular Cloning: A Laboratory Manual. 3rd edition. *Cold Spring Harbor Laboratory Press*, New York, USA.

Sodeikat G. and Pohlmeyer K. (2007) Impact of drive hunts on daytime resting site areas of wild boar family groups (*Sus scrofa L.*). *Wildlife Biology in Practice,* 3: 28-38

Taberlet P. and Luikart G. (1999) Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*, 68: 41-55

Toigo C., Servanty S., Gaillard J.M., Brandt S. and Baubet E. (2008) Disentangling natural from hunting mortality in an intensively hunted wild boar population. *Journal of Wildlife Management*, 72: 1532-1539

Valière N. (2002) GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes*, 2: 377-379

Valière N., Bonenfant C., Toigo C., Luikart G., Gaillard J.M. and Klein F. (2007) Importance of a pilot study for non-invasive genetic sampling: genotyping errors and population size estimation in red deer. *Conservation Genetics*, 8: 69-78

Van Horn R.C., Altmann J. and Alberts S.C. (2008) Can't get there from here: inferring kinship from pairwise genetic relatedness. *Animal Behaviour*, 75: 1173-1180

Vernesi C., Crestanello B., Pecchioli E., Tartari D., Caramelli D., Hauffe H. and Bertorelle G. (2003) The genetic impact of demographic decline and reintroduction in the wild boar (*Sus scrofa*): A microsatellite analysis. *Molecular Ecology*, 12: 585-595

Waits J.L. and Leberg P.L. (2000) Biases associated with population estimation using molecular tagging. *Animal Conservation*, AUG: 191-199

Waits L.P., Luikart G. and Taberlet P. (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology*, 10: 249-256

Walsh P.S, Metzger D.A. and Higuchi R. (1991) Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques,* 10; 506-513

Wilberg M.J. and Dreher B.P. (2004) Genecap: a programm for analysis of multilocus genotype data for non-invasive sampling and capture-recapture population estimation. *Molecular Ecology Notes*, 4: 783-785

Wilson G.J., Frantz A.C., Pope L.C., Roper T.J., Burke T.A., Cheeseman C.L. and Delahay R.J. (2003) Estimation of badger abundance using faecal DNA typing. *Journal of Applied Ecology*, 40: 658-666

Woods J.G., Paetkau D. and Lewis D. (1999) Genetic tagging of free-ranging black and brown bears. *Wildlife Society Bulletin*, 27: 616-627

Appendix III

Comparison of established methods for quantifying genotyping error rates in wildlife forensics

Karolina Kolodziej, Holger K. Schulz, Kathrin Theissinger, Cornelia Ebert, Ulf Hohmann, Ralf Schulz

Publication in *Conservation Genetics Resources,* in press

Abstract

Several methods have been applied to calculate genotyping error rates (GER) for non-invasive population size estimations. However, there is a lack of comparability between these methods. Here we focused on the comparison of methods for determination of GER within one study using faeces samples of wild boars (*Sus scrofa*). Error rates were calculated by 1) comparison of reference tissue samples and rectum faeces samples 2) the number of deviations between replicates and the assumed consensus genotypes, 3) re-analysis of a subsample interpreted by allelic and genotype comparisons, and 4) a blind-test of anonymously subdivided faecal samples. The error rates differed widely between these four methods (0 to 57.5%) and underline the need of a consensus approach. The blind-test resulted in a GER of 4.3%. We recommend conducting such a blind-test for estimating realistic GER when starting a pilot study in wildlife forensics.

Introduction

Population size estimation using non-invasive sampling has emerged as an important field in conservation genetics and is of crucial importance for wildlife management. Faeces as DNA source is attractive due to the potentially easy sampling and the homogeneous capture probability (Wehausen *et al.* 2004). However, faeces genotyping implicates pitfalls like reduced amplification success due to low target DNA concentrations, interference from diet and bacterial DNA, presence of PCR inhibitors (Lampa *et al.* 2008) or degradation of DNA (Idaghdour *et al.* 2003). Further challenges are the occurrence of incorrect genotypes due to allelic dropout or false alleles (Wehausen *et al.* 2004).

These genotyping errors can either lead to underestimations of population sizes when genetic tagging is not unique, and/or to overestimations when genotyping errors lead to additional false genotypes (Creel *et al.* 2003). Several studies were carried out to minimize genotyping errors, by e.g. using specific primers (Broquet *et al.* 2007), optimizing storage and extraction methods (Flagstad *et al.* 1999; Murphy *et al.* 2007), sampling fresh faeces (Maudet *et al.* 2004), thorough selection of genetic markers (Broquet *et al.* 2007), and applying a multi-tubes approach (MA; Taberlet *et al.* 1996).

However, genotyping errors can never be completely avoided in a microsatellite dataset, and therefore it is crucial to quantify the genotyping error rate (GER) in order to estimate the reliability of inferred results (Hoffman & Amos 2005; Pompanon *et al.* 2005). Valière *et al.* (2007) reviewed the commonly used genotyping error calculations, but without actually comparing the respective methods (see also Broquet *et al.* 2004; Beja-Pereira *et al.* 2009). For this, a

comparison of different methods for quantifying genotyping errors using the same data set is needed.

Here, we developed a general approach for achieving reliable microsatellite datasets for population size estimations in wildlife forensics with a realistic GER of less than 5%, ensuring reliable population size estimations (Taberlet & Luikart 1999). It is important to consider the real GER instead of the theoretical GER because it reflects the accuracy of the individual assignment (Frantz *et al.* 2003). We genotyped 315 faecal samples from wild boars, calculated GER by applying different methods of MA and compared them. Finally, we present a blind-test for quantifying a realistic GER.

Material and methods

From 17 wild boars (*Sus scrofa*) obtained from a driving hunt in 2008 the Palatinate forest (SW Germany) tissue samples were taken from ear biopsies and faeces samples were taken from the rectum. Additionally, we collected 315 wild boar faecal samples in a 4000 ha area in the Palatinate forest (SW Germany) during a three week sampling period in December 2007. For the blind-test, eight additional faecal samples were collected within one day under the same conditions for method comparison. The tissue samples were stored in denatured ethanol 99.6% at 4°C and the rectum and faecal samples were stored in plastic bags and frozen at -20°C until DNA extraction. Tissue samples were washed with 3 mL 1x phosphate-buffered saline before DNA-isolation. The DNA was isolated using a Nucleo-Spin Tissue Kit (Macherey Nagel), following the manufacturer's protocol.

Approximately 250 mg of the surface of each faecal dropping directly used (without drying) for DNA-extraction, which was carried out according to the manufacturer's protocol from the Nucleo-Spin Tissue Kit (Macherey Nagel) with the modification of three wash steps. We genotyped the DNA extracts from tissue and faeces with a subset of four species-specific microsatellite loci and additional sex marker (TNFB, Sw2496, Sw2021, Sw742 and PigSRY) which have proven good genotyping success in wild boar faeces samples (Kolodziej *et al.* 2011). The PCR was conducted in 15 µl containing (final concentration) 2.5 mM $MgCl_2$ for faeces, 1.5 mM $MgCl_2$ for tissue, 0.2 mM of each dNTP, 0.2 µM of each primer, 0.1 µg/µl bovine serum albumin (BSA only for faeces), 0.025 u/µl Go Taq Flexi DNA polymerase (Promega) and 1µl template using a Primus 96 Cycler (Peqlab Biotechnologie GmbH) under following conditions: initial denaturation at 95°C for 2 min, followed by 40 cycles of 30 s at 95°C, 30 s at locus specific annealing temperature and 30 s at 72°C, and a terminal elongation step at 72°C for 5 min. The amplification products were visualized on agarose gel. Samples with visible bands were run on a CEQ 8000 Sequencer (Beckman Coulter) and analyzed on the corresponding software CEQ SYSTEM 9.0 to determine allele lengths.

In a comparative MA (MA_{comp}) PCR was repeated three to five times and individual alleles had to be identical across at least two repetitions for heterozygotes and three repetitions for homozygotes (Frantz *et al.* 2003). In case of five repeats with two heterozygotes and three homozygotes, the sample was assigned as heterozygote. In case of GER > 5% we altered the MA_{comp} to a conservative MA (MA_{cons}; Taberlet & Luikart 1999): homozygous genotypes had to be identical across at least eight repetitions. Samples with low genotyping success

and ambiguous MA outcomes were discarded from the analysis. Consensus genotypes were determined using the software GIMLET (Valière *et al.* 2002), with the default threshold value 2 for heterozygous, and 3 (MA_{comp}) or 8 (MA_{cons}) for homozygous genotypes. The consensus genotypes were manually scrutinized by re-checking the raw data of all one-mismatch pairs (Paetkau 2003).

Two sources of genotyping errors were defined: allelic dropout (ADO; one allele of a heterozygous locus was not amplified or not scored) and false allele (FA; an additional allele was misleadingly assigned to a heterozygous locus). Individual identification and genotype comparisons were calculated using the modified EXCEL sheet GENECAP (Wilberg & Dreher 2004). The following four methods were used to quantify GER: 1) Error rates were defined by using reference tissue samples and corresponding rectum faeces samples (n = 17). Tissue samples were analysed two times, whereas rectum faeces were analysed with MA_{comp}. The determined tissue sample consensus genotypes were compared with the determined consensus genotypes of the rectum faeces. 2) The number of deviations from the consensus genotypes were counted for locus and individual by comparing the repeated genotypes and the assigned consensus genotype (Broquet & Petit 2004) using GIMLET; 3) A re-analysis (Hoffman & Amos 2005), starting from DNA-extraction, was conducted for a subsample of 10% (N = 30) out of the whole sampling (N = 315). We compared the results by a) counting the allelic differences within consensus genotypes (Bonin *et al.* 2004) and b) comparing complete genotypes (Pompanon *et al.* 2005); 4) In a blind-test eight faecal samples from different wild boar individuals were anonymously divided into three to eight subsamples (N = 40) prior genotyping and subsequently assigned to

an individual by matching the consensus genotype obtained from MA_{cons}. The general procedure is demonstrated as a step by step scheme in Figure 1.

Results and discussion

The major challenge in non-invasive genotyping is to minimize genotyping errors and to quantify the real GER. A reliable method for non-invasive real GER calculation is the use of reference samples like blood or tissue and corresponding faecal samples to obtain the true consensus genotypes (Bayes *et al.* 2000). Faeces samples should be collected directly after defecating and tissue samples taken of the same individual as reported in Bayes *et al.* (2000). In practice this procedure is often difficult to acquire and can only be suggested for species where hunting is permitted and hunted individuals are available. In our study reference faeces samples were directly removed from the rectum of shot boars. Comparisons of reference tissue samples and corresponding rectum faeces samples (method 1; Table 1) revealed no genotyping errors, thus supporting the MA_{comp}.

Table 1: Estimation of genotyping errors rates (GER) [%] based on allelic drop outs (ADO) and false alleles (FA) for comparative multiple-tubes approach (MA_{comp}) and conservative multiple-tubes approach (MA_{cons}) using three different methods.

	MA_{comp}		MA_{cons}	
Method	ADO	FA	ADO	FA
1	0	0	/	/
2	26.8	7.3	31.5	4.5
3a	16.5	2.0	0	0.4
3b	49.0	8.0	0	8.0
4	/	/	0	4.3

Method 1) Comparison of reference tissue samples and rectum faeces samples; 2) GER calculated by deviations of repeats to consensus genotypes; 3a) re-analysis with GER calculated by allele comparisons; 3b) re-analysis with GER calculated by genotype comparisons; 4) blind-test of anonymously subdivided faecal samples

Compared to method 1 we expected a higher GER in method 2, since in method 1 the final consensus genotypes were compared to each other whereas in method 2 the single genotypes obtained from the repeats were compared to an expected final consensus genotype of each sample. For faeces samples we assumed that environmental influences, like UV-light, endogenous endonuclease activity or oxidative damage, could be a reason for low quality of the samples (Deagle *et al.* 2006). Therefore we determined the GER with method 2. The MA_{comp} used in method 2 resulted in 156 successful assignments out of 315 faeces samples, implying that 50% of the samples were discarded due to low DNA quality. This amount was in range with other non-invasive studies (Lucchini *et al.* 2002; Hedmark *et al.* 2004; Arrendal *et al.* 2007). The GER resulted in 26.8% ADO and 7.3% FA (Table 1). These values are in the range of those reported in comparable studies on otters (*Lutra lutra*; Ferrando *et al.* 2008; Hajkova *et al.* 2009), but lower

than those for black rhinoceroses (*Diceros bicornis*; Garnier *et al.* 2001) and higher compared to a study on coyotes (*Canis latrans*; Prugh *et al.* 2005).

Our calculations for method 2 were initially based on deviations of the repeats to the consensus and not to the real genotype. Thus, we validated the results of method 2 with a re-analysis of 10% of the samples (N=30; method 3 using MA_{comp}). Applying method 3a 216 of 240 allelic comparisons could be conducted. We identified 176 matches and 40 mismatches, which resulted in a considerably lower GER compared to method 2 (Table 1) and were in the same order of magnitude compared to a study on wolves (*Canis lupus*; Lucchini *et al.* 2002). According to method 3b, which is explained in the review of Pompanon *et al.* (2005), 23 of 30 consensus genotypes could be identified, and 10 out of 23 samples could be assigned to the correct individual, producing the highest GER in relation to the other methods (Table 1). To prove that the small sample size of N = 30 did not cause a bias, we calculated the GER with method 2 for N = 30 (ADO = 22.7%; FA = 7.3%). These results were similar to the GER calculated with N = 315.

The discrepancies of GER estimations for our dataset (Table 1) were expected but astonishingly high. The high GER, mainly caused by ADO, showed that three identical repeats for a consensus genotype are not enough to reliably describe homozygous loci. ADO occurred considerably more often than previously assumed, and the MA_{cons} method seems therefore preferable. However, after applying MA_{cons}, method 3 provided an ADO rate of zero, whereas method 2 led to ADO rates higher than obtained from MA_{comp} (Table 1). These results were due to the increased number of repetitions: a mis-identified consensus homozygous

genotype (due to low number of repeats) obscures the actual ADO rate; in contrast this mis-identified homozygous genotype was correctly identified during a re-analysis and therefore leading to a more realistic ADO rate. This is an argument to resign from the calculation of GER by using the deviation of repeats from the assumed consensus genotype or by calculating the ratio of observed ADO using the number of heterozygous genotypes. These methods produce, however, only a theoretical GER.

Method 3b (genotype comparisons) is favourable compared to 3a (allele comparison) because a single error in a multilocus genotype already leads to a false classification of individuals and results in biased population size estimation (Harris *et al.* 2010). Therefore, the number of incorrect genotypes should be counted rather than simply counting incorrect alleles. To validate the results of method 3b, we conducted a blind-test (method 4) as control procedure, which is illustrated in Figure 1. From eight collected samples (step 1) 40 faeces samples were divided (step 2) and 25 could be successfully genotyped (step 3 and 4). We could identify six of eight different individuals (step 5). Resolution of the blind-test showed that 24 samples were correctly assigned to match original samples or individuals (step 6). Only one sample could not be assigned properly due to one FA. Thus, the blind-test resulted in the lowest GER of 4.3% with ADO rate of 0 equal to results of method 3 (Table 1). This test implies a reliable and easily reproducible basis for calculating a realistic GER before starting a pilot study on population size estimates in wildlife forensics, especially when reference tissue or blood samples are not available. We recommend to use at least 20 faeces samples due to required statistical robustness (Bortz & Doering 2002) or if

possible a minimum number of 10% of collected samples in one collection period, which was suggested by Pompanon *et al.* (2005) for blind retyping, e.g. for re-analysis. It is important to provide a higher number of samples from different individuals than subsamples, due to the low DNA quality of faecal samples and thus a lower PCR success rate. This test can be transferred to other systems using different non-invasive samples (hairs or feather) with the assumption that samples are dividable or the sampling provides more than one sample per individual.

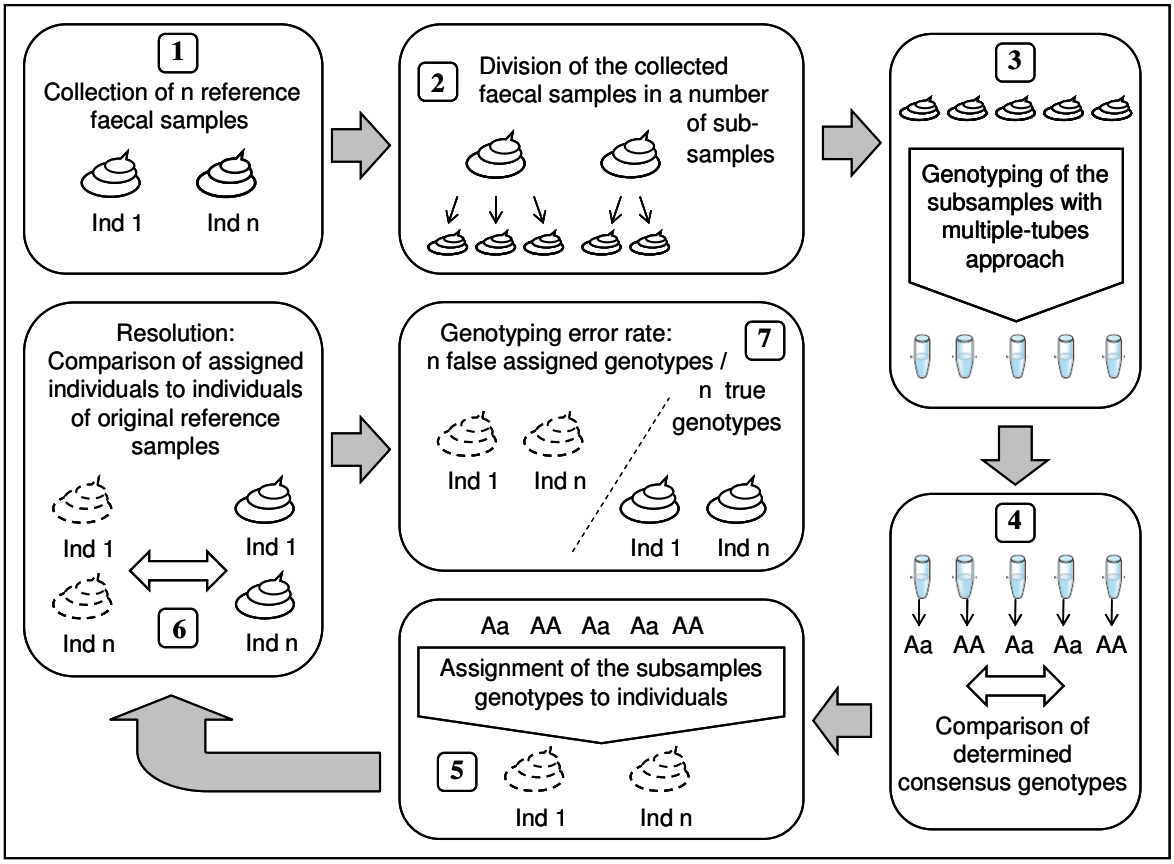

Figure 1: Flowchart of a blind-test to estimate realistic genotyping error rates. The number of collected reference faecal samples should not be lower than 20. We recommend a minimum number of 10% of collected samples in one collection period and to provide a higher number of samples from different individuals than subsamples. We suggest starting the test with a comparative multiple-tubes approach (fewer repetitions) in respect of reduced costs. In case of high error rates, the multiple-tubes approach could be modified by increased repetitions.

In conclusion, calculated GERs are often incomparable between studies due to the different approaches used. Therefore, different expected outcomes even for using the same approach but a different calculation are possible (see difference between genotype and allele comparison). Here, we want to highlight the importance to scrutinize the reliability of genotyping results provided by using only one method. Furthermore, we proved the necessity to calculate a realistic GER, opposed to a theoretical GER, which is based on the deviation of acquired genotyping repeats to an expected consensus genotype. This is important for studies dealing with e.g. population size estimations or mating system analyses. Finally, in Figure 2 we recommend a general stepwise approach for obtaining a realistic GER for reliable population size estimation in wildlife forensics.

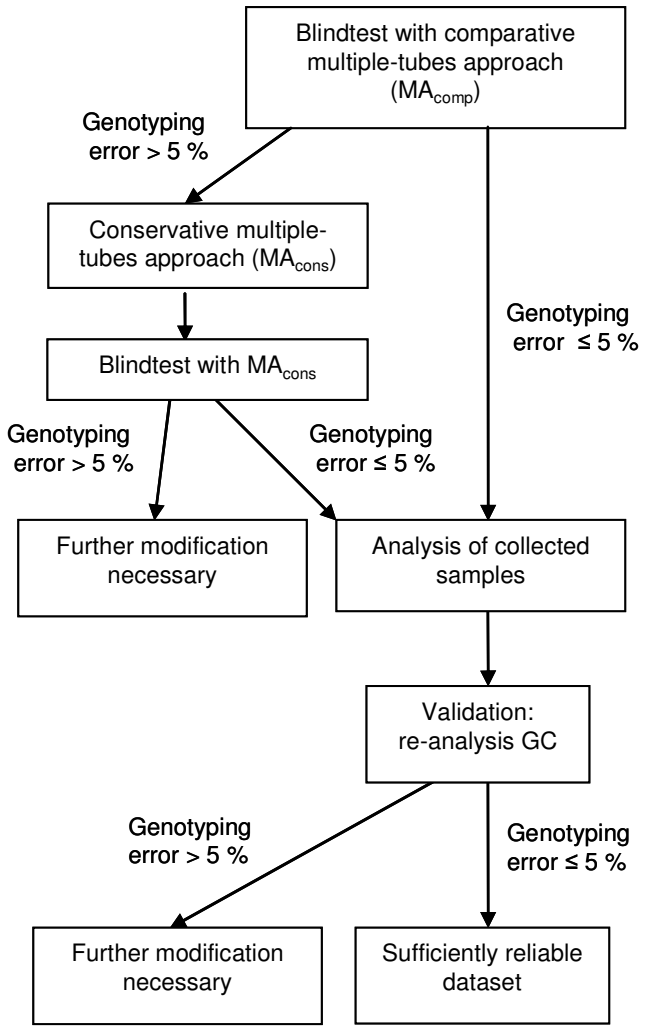

Figure 2: Flowchart presenting a general approach to obtain realistic genotyping error rates for reliable population size estimation in wildlife forensics. A genotyping error rate of 5% is assumed to be adequate for population size estimation. The blind-test can be used in a pilot study to test multiple-tubes approaches and can be validated with re-analysis GC (genotype comparison) after analysis of collected faecal samples for a study.

Acknowledgments

We wish to thank D. Huckschlag for good advices for the manuscript. We also thank R. Heydenreich for proofreading of this manuscript. Furthermore, we are grateful to T. Bürgi for technical assistance. This project was supported by the Foundation „Rheinland-Pfalz für Innovation“ and the Ministry for Environment, Forestry and Consumer Protection, Rhineland-Palatinate. K.K. was supported through a PhD scholarship from the Lotto Foundation Rhineland-Palatinate.

References

Arrendal J., Vila C. and Bjorklund M. (2007) Reliability of noninvasive genetic census of otters compared to field censuses. *Conservation Genetics*, 8: 1097-1107

Bayes M., Smith K., Alberts S., Altmann J. and Bruford M. (2000) Testing the reliability of microsatellites typing from feacal DNA. *Conservation Genetics,* 1: 173-176

Beja-Pereira A., Oliveira R., Alves P.C., Schwartz M.K. and Luikart G. (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Ressources*, 9:1279-1301

Bonin A., Bellemain E., Bronken Eidesen P., Pompanon F., Brochmann C. and Taberlet P. (2004) How to track and assess genotyping errors in population genetics. *Molecular Ecology*, 13: 3261-3273

Broquet T., Menard N. and Petit E. (2007) Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics,* 8: 249-260

Bortz J. and Doering N. (2002) Forschungsmethoden und Evaluation. *Springer Verlag*, Berlin, Heidelberg. Zweie Auflage.

Broquet T. and Petit E. (2004) Quantifying genotyping errors in noninvasive population genetics. *Molecular Ecology*, 13: 3601-3608

Creel S., Spong G., Sands J.L., Rotella J., Zeigle J., Joe L., Murphy K.M. and Smith D. (2003) Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Molecular Ecology*, 12: 2003-2009

Deagle E.B., Eveson J.P. and Jarman S.M. (2006) Quantification in DNA recovered from highly degraded samples- a case study on DNA in faeces. *Frontiers of Zoology,* 3:11

Ferrando A., Lecis R., Domingo-Roura X. and Ponsa M. (2008) Genetic diversity and individual identification of reintroduced otters (*Lutra lutra*) in north-eastern Spain by DNA genotyping of spraints. *Conservation Genetics,* 9: 129-139

Flagstad O., Roed K., Stacy J.E. and Jakobsen K.S. (1999) Reliable noninvasive genotyping based on excremental PCR of nuclear DNA purified with a magnetic bead protocol. *Molecular Ecology*, 8: 879-883

Frantz A.C., Pope L.C., Carpenter P.J., Roper T.J., Wilson G.J., Delahay R.J. and Burke T. (2003) Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Molecular Ecology*, 12: 1649-1661

Garnier J.N., Bruford M.W. and Goossens B. (2001) Mating system and reproductive skew in the black rhinoceros. *Molecular Ecology*, 10: 2031-2041

Hajkova P., Zemanova B., Roche K. and Hajek B. (2009) An evaluation of field and noninvasive genetic methods for estimating Eurasian otter population size. *Conservation Genetics,* 10: 1667-1681

Harris R.B., Winnie J., Amish S.J., Beja-Pereira A., Godinho R., Costa V. and Luikart G. (2010) Argali Abundance in the Afghan Pamir Using Capture-Recapture Modeling From Fecal DNA. *Journal if Wildlife Management,* 74: 668-677

Hedmark E., Flagstad Ø., Segerstroem P., Persson J., Landa A. and Ellegren H. (2004) DNA-based individual and sex identification from wolverine (*Gulo gulo)* faeces and urine. *Conservation Genetics*, 5: 405–410

Hoffman J.I. and Amos W. (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology,* 14: 599-612

Idaghdour Y., Broderick D. and Korrida A. (2003) Faeces as a source of DNA for molecular studies in a threatened population of great bustards. *Conservation Genetics*, 4: 789-792

Kolodziej K., Theissinger K., Brün J., Schulz H.K. and Schulz R. (2011) Determination of the minimum number of microsatellite markers for individual genotyping in wild boar (*Sus scrofa*) using a test with close relatives. *European Journal of Wildlife Research,* 58: 621-628

Lampa S., Gruber B., Henle K. and Hoehn M (2008) An optimisation approach to increase DNA amplification success of otter faeces. *Conservation Genetics*, 9: 201-210

Lucchini V., Fabbri E., Marucco F., Ricci S., Boitani L. and Randi E. (2002) Noninvasive molecular tracking of colonizing wolf (*Canis lupus*) packs in the western Italian Alps. *Molecular Ecology*, 11: 857-868

Maudet C., Luikart G., Dubray D., von Hardenberg A. and Taberlet P. (2004) Low genotyping error rates in wild ungulate faeces sampled in winter. *Molecular Ecology Notes*, 4: 772-775

Murphy M.A., Kendall K.C., Robinson A. and Waits L.P. (2007) The impact of time and field conditions on brown bear (*Ursus arctos*) faecal DNA amplification. *Conservation Genetics*, 8: 1219-1224

Navidi W., Arnheim N. and Waterman M. (1992) A multiples-tube approach for accurate genotyping of very small DNA samples by using PCR - Statistical considerations. *American Journal of Human Genetics*, 50: 347-359

Paetkau D. (2003) An empirical exploration of data quality in DNA-based population inventories. *Molecular Ecology,* 12: 1375-1387

Pompanon F., Bonin A., Bellemain E. and Taberlet P. (2005) Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics*, 6: 847-859

Prugh L., Ritland C., Arthur S. and Krebs C (2005) Monitoring coyote population dynamics by genotyping faeces. *Molecular Ecology*, 14: 1585-1596

Taberlet P., Griffin S., Goossens B., Questiau S., Manceau V., Escaravage N., Waits L.P. and Bouvet J. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, 24: 3189-3194

Taberlet P. and Luikart G. (1999) Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*, 68: 41-55

Valière N. (2002) GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes*, 2: 377-379

Valière N., Bonenfant C., Toigo C., Luikart G., Gaillard J.M. and Klein F. (2007) Importance of a pilot study for non-invasive genetic sampling: genotyping errors and population size estimation in red deer. *Conservation Genetics*, 8: 69-78

Wehausen J.D., Ramey R.R. 2nd and Epps C.W. (2004) Experiments in DNA extraction and PCR amplification from bighorn sheep feces: the importance of DNA extraction method. *Journal of Heredity*, 95: 503-509

Wilberg M.J. and Dreher B.P. (2004) Genecap: a programm for analysis of multilocus genotype data for non-invasive sampling and capture-recapture population estimation. *Molecular Ecology Notes* 4: 783-785

Wilson G.J., Frantz A.C., Pope L.C., Roper T.J., Burke T.A., Cheeseman C.L. and Delahay R.J. (2003) Estimation of badger abundance using faecal DNA typing. *Journal of Applied Ecology*, 40: 658-666

Erklärung

Ehrenwörtliche Erklärung zu meiner Dissertation mit dem Titel:
Entwicklung einer Methode zur Populationsschätzung von Wildschweinen (*Sus scrofa*) mittels Genotypisierung nicht-invasiv gewonnener Proben

Hiermit erkläre ich, dass ich die beigefügte Dissertation selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel genutzt habe. Alle wörtlich oder inhaltlich übernommenen Stellen habe ich als solche gekennzeichnet.

Ich versichere außerdem, dass ich die beigefügte Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und, dass diesem Promotionsverfahren keine endgültig gescheiterten Promotionsverfahren vorausgegangen sind.

_____	_____
Ort, Datum	Unterschrift

Curriculum vitae

Name:	Karolina Kolodziej
Born:	31.12.1980 in Opole / Poland
Adress:	Friedrich-Ebert-Straße 20 76829 Landau Germany

Education and career

since 10/2011	Trainee position in European Commission JRC, Institute for Health and Consumer Protection in Ispra, Italy
since 04/2007	PhD-Student at University of Koblenz-Landau, Institute for Environmental Sciences, Germany
01/2006 –11/2006	Diploma-thesis: Identification of an uncultured protist by fluorescence *in situ* hybridization and scanning electron microscopy
10/2000 –11/2006	Study of Biology at the University of Kaiserslautern, Germany with main focus on ecology and molecular biology
9/1997 – 6/2000	Finaly secondary school examination (Matura) at the Carl-Bosch-Gymnasium in Ludwigshafen, Germany