Causes of Spatial Genetic Structure in Mammals: A Case Study in the Atlantic Forest, Brazil

Anders Gonçalves da Silva

Submitted in partial fulfillment of the
Requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2007

ABSTRACT

Causes of Spatial Genetic Structure in Mammals: A Case Study in the Atlantic Forest, Brazil

Anders Gonçalves da Silva

One of the fundamental links between ecological and evolutionary processes at fine spatial scales is the association between dispersal and gene flow. Population genetic theory predicts that the degree of genetic differentiation among subpopulations is inversely related to the amount of gene flow. Landscape ecology demonstrates that the degree of landscape connectivity among subpopulations is a function of the species' dispersal capacity. Dispersal distance in mammals scales with both body size and trophic level, which suggests that these natural history characteristics could be good predictors of genetic structure in mammals. The actual dispersal distance is significantly dependent on the degree of heterogeneity of the surrounding landscape. Because dispersal and landscape connectivity, in most species, scale with body size and trophic level, it is plausible to think that genetic structure will also scale with these characteristics. To explore test this hypothesis, I studied the spatial distribution of genetic variation across species of mammals differing in both body size and trophic level within a single landscape.

Samples were collected in the fragmented landscape of the *Pontal do Paranapanema*, in the western tip of the State of São Paulo, Brasil. The area is contained within the Atlantic Forest of the Interior biome. Both dung and blood samples were

collected from collared peccary (*Tayassu tajacu*), white-lipped peccary (*Tayassu pecari*), lowland tapir (*Tapirus terrestris*), ocelot (*Leopardus pardalis*), and jaguar (*Panthera onca*). For jaguars, blood samples were also obtained from two other locations, allowing for exploration of genetic variation at larger scales in this species. Samples for pumas (*Puma concolor*) were also collect, but they were in insufficient number for analysis, and the species was removed from this study. Genetic variation was evaluated using microsatellites.

Two different approaches were taken to analyze the data. First, I inferred genetic structure using a Bayesian model, which simulation data showed to be robust to the conditions present in this study. I then overlaid the inferred structure on a map of the landscape to infer barriers to gene flow. Second, I investigated scaling of genetic structure within an isolation-by-distance framework in the collared peccary, white-lipped peccary, lowland tapir and ocelot. Here, I examined the correlation of genetic relationships among individuals of each species with different measures of geographic distance.

The results show significant genetic structure for all five remaining species. Habitat fragmentation affects all species, however, as expected, at different scales. The smaller collared and white-lipped peccaries show genetic partitions that are correlated with habitat fragmentation at smaller scales than the lowland tapir, or either of the two species of carnivores. Levels of genetic differentiation, when compared among species within the landscape, are higher for herbivores than carnivores. In addition, the scale of dispersal, as measured by the point of inflection in the isolation-by-distance curve, increases with body size being smaller for collared peccaries, intermediary for white-

lipped peccaries and highest for lowland tapirs. And, current results suggest that the dispersal scale may be very similar for lowland tapirs (large herbivore) and ocelots (small carnivore). However, the social structure displayed by both species of peccaries could be inflating the degree of genetic differentiation in these species, due to the smaller effective population size that social structure implies. But, the fact that tapirs still display a higher degree of structure than carnivores suggests that social structure may not be an important factor in determining differences in degree of genetic differentiation among the examined species. And, the short period (measured in number of generations) since fragmentation occurred in this landscape probably means that some of the observed patterns of genetic structure are in transition, and therefore there is a lack genetic drift – gene flow equilibrium. As such, some of the observed patterns might be reflective of past rather than present levels of connectivity and fragmentation.

These results are in general agreement with the initial predictions and support the hypothesis that genetic structure is scalable with body size and trophic. Collectively, these results indicate that natural history characteristics may be good predictors of genetic structure. In principle, this would allow for studies undertaken in an experimental setting using smaller animals (e.g. insects or mice) at a finer scale to be generalized to larger scales. However, independent studies should be carried out to further corroborate the findings published here. This would allow for more controlled experiments to examine further the effects of body size, trophic level, and other factors such as social structure, degree of landscape heterogeneity, population size, and non-equilibrium among evolutionary forces on the scaling of genetic structure. From a conservation point of view, this study shows that a landscape genetic approach can generate useful and

important information. However, the density of data needed for most analyses in landscape genetics still preclude the use of this approach in conservation settings were data are too scarce, and there is little prior ecological and historical information. I also make specific recommendations for the conservation of the *Pontal do Paranapanema* landscape.

RESUMO

Causas da Estruturação Espacial da Variação Genética em Mamíferos: Um Estudo de Caso na Mata Atlântica, Brasil

Anders Gonçalves da Silva

A associação entre dispersão e fluxo gênico é uma das associações fundamentais subjacentes à interface entre os processos ecológicos e evolutivos em pequenas escalas espaciais. A teoria da genética de populações prediz que o grau de diferenciação genética é inversamente proporcional à quantidade de fluxo gênico. A ecologia de paisagens, por outro lado, demonstra que o grau de conectividade da paisagem entre sub-populações é uma função da habilidade da espécie de dispersar. A distância de dispersão em mamíferos é especificada por uma função proporcional ao tamanho do corpo e nível trófico da espécie. Essa relação sugere que essas características podem ser bons indicadores dos níveis de estruturação genética em mamíferos. A real capacidade de dispersar, no entanto, é significativamente modificada pelo grau de heterogeneidade do contexto paisagístico. Porque a dispersão e a conectividade da paisagem são ambos proporcionais ao tamanho do corpo e ao nível trófico, é plausível imaginar que a o grau de estruturação genética também será proporcional a essas características. Com o intuito de explorar essas expectativas, este estudo examina a distribuição espacial da variabilidade genética em espécies de mamíferos de diferentes tamanhos e níveis tróficos em uma mesma paisagem.

Amostras foram coletadas na paisagem fragmentada do *Pontal do Paranapanema*, no extremo oeste to Estado de São Paulo, Brasil. A área está contida dentro do bioma da

Mata Atlântica do Interior. Amostras de fezes e sangue foram coletadas de catetos (*Tayassu tajacu*), queixadas (*Tayassu pecari*), anta-brasileira (*Tapirus terrestris*), jaguatirica (*Leopardus pardalis*), e onça-pintada (*Panthera onça*). Amostras de sangue de onças-pintadas também foram coletadas em duas outras localidades, *Anaurilândia* e *Ivinhema*, o que permitiu uma análise da estruturação da variabilidade genética na espécie em uma escala maior. Amostras de onça-parda (*Puma concolor*) também foram coletadas, mas o tamanho amostral foi insuficiente para realizar as análises propostas neste estudo e a espécie foi, consequentemente, removida. Variabilidade genética foi avaliada usando microssatélites.

Duas abordagens foram utilizadas durante as análises dos dados. Primeiro, a estrutura genética foi inferida através de um modelo Bayesiano, que dados de simulação indicam que é robusto às condições impostas por esse estudo. A estrutura estimada foi então sobreposta a um mapa da paisagem e barreiras ao fluxo gênico foram inferidas. Segundo, o escalonamento da estrutura genética foi investigado em catetos, queixadas, anta-brasileira, e jaguatiricas usando uma abordagem de isolamento-por-distância. Aqui, a correlação entre a associação genética entre indivíduos de cada espécies e diferentes medidas de distância geográfica foi investigada.

A estruturação genética foi significativa para todas as cinco espécies remanescentes. A fragmentação do habitat afeta todas as espécies, mas como esperado, em escalas diferentes. Em catetos e queixadas, as duas menores espécies de herbívoros, a estruturação da variabilidade genética ocorre em escalas menores que nas antas, ou ambas as espécies de carnívoros. Um padrão similar foi observado na análise de isolamento-por-distância. Neste caso, a escala de dispersão, medida pelo ponto de

inflexão da curva de isolamento-por-distância, aumenta com o aumento em tamanho de corpo, sendo menor para os catetos, intermediário para os queixadas, e máxima para as antas. Além disso, os resultados sugerem que a escala de dispersão é similar entre jaguatiricas (um carnívoro pequeno) e as antas (um herbívoro grande). No entanto, a estrutura social característica das duas espécies de pecarídeos pode estar causando um aumento no grau de diferenciação genética entre sub-populações, além do que seria esperado pelas diferenças em tamanho de corpo e nível trófico em relação as outras espécies. No entanto, o fato das antas também terem altos níveis de diferenciação genética entre sub-populações, maiores que as observadas em ambas as espécies carnívoros, sugere que a estrutura social não é um fator determinante de diferenças na estruturação genética entre essas espécies. No mais, o curto período de tempo (medido em número de gerações) desde a fragmentação do habitat provavelmente significa que alguns dos padrões de estruturação genética estão em transição, e não há equilíbrio entre deriva genética e fluxo gênico. Sendo assim, é possível que alguns dos padrões observados sejam um reflexo de graus de conectividade e fragmentação passados, e não presentes.

Em conclusão, os resultados corroboram, de maneira geral, as expectativas iniciais. E, sugerem que a estruturação genética pode ser escalonada com tamanho de corpo e nível trófico. Portanto, essas características são bons indicadores de níveis de estruturação genética. Em um contexto mais geral, isso permitiria estudar efeitos de paisagens em animais relativamente pequenos (por exemplo, insetos ou camundongos) e em escalas relativamente pequenas, e generalizar os resultados para escalas maiores. Isso permitiria experimentos controlados para examinar de forma mais profunda os efeitos de

tamanho de corpo, nível trófico, e outros fatores, como estrutura social, grau de heterogeneidade da paisagem, tamanho populacional, e situação de desequilíbrio entre forças evolutivas no escalonamento da estruturação genética. Do ponto de vista da conservação da paisagem, esse estudo demonstra que a abordagem da genética da paisagem pode ser uma ferramenta útil e importante. No entanto, a densidade de dados necessária para a maioria das análises restringe o uso da genética da paisagem em situação de poucos dados, e pouca informação ecológica e histórica. Ao final, recomendações específicas são feitas para a conservação da paisagem do *Pontal do Paranapanema*.

Table of Contents

List of Tables and Figures

List of Tables

List of Figures

Acknowledgments

The work presented in this dissertation would not have been possible without the help and support of my family and friends. In turn, I would like to thank Juliana José, Rodrigo Tristan, and Odalys Cabrera, who were my lab companions at the Universidade Estadual de Campinas (UNICAMP) during the early preliminary work of this dissertation, and Gonçalo Pereira and Vera Nisaka who allowed me to work in their labs during this early phase. I am also in debt to Dario Grattapaglia, who opened up the Plant Genetics Laboratory at the EMBRAPA-Recursos Genéticos e Biotecnologia allowing me to carry out the main part of the laboratory work for this thesis. I am also in debt to my Brasília lab partners Eva Mamani, Juliano Pádua, Marco Pessoa, Alexandre Missiaggia, Marília Pappas, Danielle Paiva, Nathália Bueno e Carolina Sansaloni who helped, encouraged and supported me while I fought against all odds to extract viable DNA from dung samples. And, when all else was not working, consoled me with a cold glass of beer and a few good laughs at Varandas. I also thank Regina, Beto, André and Jorge Vasquez for taking care of me, and providing me with a family away from home while I stayed in Brasília. I would also like to thank Luis Curi and Emília Ribeiro for their kindness and hospitality in taking care of me during other times I was in Brasília. And, finally, Eva Mamani and Juliano Pádua for sharing their small space with me whenever I needed.

In the field, I was helped by Patrícia Medici, Laury Cullen Jr., Alessandra Nava, Alexandre Uezu, George Velastin, Paulo Mangini, Joares May Jr., Cristina Tófoli, Robin Elliott, Arnaud Desbiez and Reem Hajjar. Samples were also collected with the help of Denis Sana, Kauê Abreu, Carlos Plateiro, and Cássio Roberto Leonel Peterka. Special

recognition is warranted to José Maria do Aragão, for his keen eye sight, entertaining stories, companionship and profound knowledge of the surrounding landscape. I am also thankful to all the IPÊ staff and researchers in the *Pontal*, who were always ready to help, and always greeted you with a friendly smile and an offer of coffee. And, a special thank you to Alexandre Uezu for helping me figure out ArcView, and supplying the maps needed for analysis.

In the early days of this thesis, while I was still in New York trying to figure out what my question may be and what the law professor was actually talking about, the friendship of Bea Perez-Sweeney, Evelyn Luciano, Juan Carlos Morales, Gonçalo Ferraz, Katherine McFadden, Chris Fiorello, Mariana Vale, Marcelo Weksler, Francisca Almeida, Aleksei Chmura, Andres Gomes, Elizabeth Nichols, Fernando Colchero, Dalia Conde, Rasit Bilgin, Tomas van Roosmalen, Robin Elliott, Camelia Minoiu, Claudia Minoiu, Todd Davis and Nancy Dammann were, and are still, dearly cherished. Later on, Claire Jouseau and Todd Osmundson joined the crowd on the 11th Floor, becoming good friends and companions in the trenches of graduate school. I would also like to recognize the help of Jessica Satkoski and Claire Jouseau while I was away from New York, sorting my mail and storing my books and papers. The assistance, smile and kindness of Evelyn Luciano, Katy Lopez, and Maria Meade will be missed. And, when I moved to the west coast, the transition was made easier with the help of Peter Wood, Gilia Angell, and Aaron Abrams. Finally, a special thank you to my friends Patrícia Medici, Juliana José, Arnaud Desbiez, and my parents, Jennifer and Cylon Gonçalves da Silva who were always there for me whenever I needed them; and to my future wife, Reem Hajjar, for her friendship and companionship, and taking care of me while I wrote this dissertation.

I would like to thank my advisor, Don J. Melnick, for his advice, guidance, and inspirational stories of conservation and science. My committee members Juan Carlos Morales for help in the lab, and early "acclimatization" at Columbia University; Claudio Pádua for his encouraging words and for opening up the infra-structure of IPÊ; Robert Rockwell, for his friendly advice and help; and Shahid Naeem for early discussions on the topic and merits of landscape genetics.

Funding for this project was provided by the Center for Environmental Research and Conservation (CERC), the Department of Ecology, Evolution and Environmental Biology (E3B), Graduate School of Arts and Science (GSAS), the Explorers Club, American Society of Mammalogists, Disney Wildlife Conservation Fund, and the Rufford Foundation.

This project was a component of the Landscape Detectives and the Brazilian Lowland Tapir Conservation Initiative, both programs of IPÊ - Instituto de Pesquisas Ecológicas (Institute for Ecological Research). The Initiative is grateful for institutional support from: Secretaria do Meio Ambiente / Instituto Florestal do Estado de São Paulo (Forestry Institute of São Paulo State); Parque Estadual do Morro do Diabo (Morro do Diabo State Park), Ministério do Meio Ambiente / Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis, Sorocaba Zoo, IUCN/SSC Tapir Specialist Group (TSG), Tapir Preservation Fund (TPF), American Association of Zoos and Aquariums (AZA) Tapir Taxon Advisory Group (TAG), European Association of Zoos and Aquaria (EAZA) Tapir Taxon Advisory Group (TAG), and WildTrack, Portugal. And, has received between 1996-2007 from the following institutions: Fundo Nacional do Meio Ambiente (FNMA); Fundação de Amparo à Pesquisa de São Paulo (FAPESP); Fundação

O Boticário para Proteção a Natureza; CFHF; Liz Clairborne Art Ortenberg Foundation; Smithsonian Institution, Wildlife Conservation and Management Training Program (WCMTP); Chicago Zoological Society; Scott Neotropical Fund; IUCN/SSC Tapir Specialist Group); Tapir Preservation Fund (TPF); Woodland Park Zoological Gardens; Idea Wild; Nellcor; The Ledder Family Charitable Trust; American Association of Zoo Keepers, the Puget Sound, the Los Angeles, the Houston Zoo and the Nashville Zoo Chapter; Columbus Zoological Park Association Inc.; Sophie Danforth Conservation Biology Fund; American Zoo and Aquarium Association (AZA) Tapir Taxon Advisory Group (TAG); IUCN Small Grants Programme, The Ford Foundation; USAID / Programa Natureza & Sociedade, Brazil; Parc Zoologique d'Amnéville, France; John Ball Zoo Society; Parc Zoologique Doué-la-Fontaine; Disney Wildlife Conservation Fund; North of England Zoological Society; The Oregon Zoo Foundation Conservation Fund; Cleveland Zoological Society; Givskud Zoo; Dutch Foundation Zoos Help; Discovery Channel Canada; Houston Zoo Inc.; Dallas Zoo & Dallas Aquarium at Fair Park; and Tapir Specialist Group Conservation Fund (TSGCF). Private donations were made by the following individuals (1996-2007): Alex & Suzan Sze, Alex Cardenas, Andre (Andi) Maerz, Andy Markley, Andy Schultz, Ayessa Rourke, Brandie Smith, Brenda Stringer, Bryony Angell, Carol & Mark Reid, Charles Foerster, Christine Kim & Rob Lyman, Corinna Bechko & Gabriel Hardman, Deanne Holsworth, Elaine & Paul Beckham, Gilia Angell & Aarom Abrams, Heidi Frohring, Kae Kawanishi, Katalin Pinter & Christian Fabris, Keith Sproule, Kevin Burkhill, Kristin Leus, Leonardo Salas, Michael Dee, Mickael Michault, Mitch Finnegan, Morty Ortega, Phil Schaeffer, Rudy Rudran, Sheryl Todd, and William Bob Harris. To all of you, thank you!

Dedication

I dedicate this dissertation to the field assistants, staff, and researchers at the Instituto de Pesquisas Ecológicas (IPÊ), for their 15 years of hard work and commitment to the conservation of biological diversity.

Chapter 1. Introduction: Genetic differentiation in natural populations

The patterns of genetic variation within and among populations, as well as in geographic space, predicted by population genetic theory are highly dependent on rates and patterns of reproduction and dispersal. Ecological data provide us with empirical estimates of these rates and patterns and thus the basis for applying population genetic theory to natural populations. The fundamental link between the ecology of a species and its population genetic structure can be found in the effect of reproduction and dispersal on effective population size (N_e). Mating system, social structure, mode of dispersal, and spatial distribution of populations are all ecological factors that affect the number of individuals producing offspring in each generation. In particular, from a spatial point of view, a species' ability to disperse within the limits imposed by the interaction between its natural history traits and the surrounding landscape, also known as landscape connectivity, is essential in determining, and therefore predicting, patterns of spatial distribution of genetic variation. In this study, I explore some of the putative ecological drivers of genetic differentiation in nature, within a comparative framework, aiming to identify some of the major factors contributing to patterns of distribution of genetic variation in space.

Population genetic theory

The basis of population genetic theory was laid out at the beginning of the 20th century, spurred by the re-discovery of Mendel's Laws of Inheritance. Three main figures heavily influenced the development of population genetic theory, R.A Fisher, Sewall Wright and J. B. S. Haldane. Their work, published in the early 1930's (Fisher, 1930; Haldane, 1932; Wright, 1931) revolutionized the study of evolutionary processes, helping create what is

referred to as the modern synthesis of evolution (Dobzhansky, 1937; Huxley, 1942; Mayr, 1942). In essence, it was proposed that population genetics processes could be understood by examining deviations from Hardy-Weinberg equilibrium, an elegant but simple model of population genetic expectations in the absence of any evolutionary forces. As such, the study of population genetic processes is the study of changes in allele frequencies through time in the presence of evolutionary forces.

Wright (1931) proposed analytical results that generated expectations about allele frequencies based on effective population size, mutation rate, strength of selection, and/or migration rates for a single population, describing some of the fundamental relationships in population genetics. A particularly important result of this early research was the observation that gene flow and genetic drift act in opposing directions. Too much of the former (one or more migrants per generation, as per Wright) will prevent random differentiation of populations. Too little, and genetic drift will take over to a degree inversely proportional to the effective population size. To study this relationship in nature, Wright developed a simple model of population structure, the "island model", which recognizes that populations are often subdivided into several different groups of individuals (i.e. subpopulations) that are, more or less, reproductively isolated from other such units. He applied this model to predict what the genetic relationships among the subpopulations would be given a certain level of gene flow. Wright's island model assumes all alleles are selectively neutral, no mutation, and equilibrium between gene flow and drift. In addition, it assumes an infinite number of subpopulations, each with the same number of diploid individuals (N), and that the spatial distribution of subpopulations has no effect on gene flow. Instead, each subpopulation exchanges a

certain number of individuals (proportional to *N*) each generation with a pool of migrants from all subpopulations. Therefore, any one population is equally likely to contribute migrants to any other population. Within this simplified version of a natural system, Wright derived what is one of the most commonly used equations in population genetics:

$$F_{ST} \approx \frac{1}{1+4N_e m} \quad \text{(Equation 1.1)}$$

F_{ST} has numerous interpretations, but the most common is the variance in allele frequencies among subpopulations standardized by the mean allele frequency (Weir & Cockerham, 1984), N_e is the effective size of the subpopulation, and m is the migration rate per generation; so, $N_e m$ is the effective migration rate. This relationship essentially means that the degree of population differentiation is inversely related to the amount of gene flow (as mentioned above). While Wright hinted at this relationship in 1931, at the time he only studied a small population as a subset of a larger one. A generalized island model was to appear later, and the relationship between gene flow and genetic differentiation depicted in equation 1.1 was first published 10 years later by Dobzhansky & Wright (1941).

Isolation-by-distance

To account for the spatial distribution of individuals, Wright (1943) developed a model of isolation-by-distance. In Wright's view, in a continuously distributed population over an infinitely large area (the opposite extreme to the island model), an individual's parents are randomly drawn from a circle of fixed radius (R). This circle is dependent on

the variance in dispersal distance of the parents. Using path coefficients, Wright established that dispersal range is a fundamental variable in determining levels of local differentiation, with local levels of inbreeding being higher when there are severe restrictions to gene flow (i.e. with small local *N*, and an infinitely large continuous population, F_{ST} approaches 1). Malécot (1968) re-analyzed the problem (Malécot, 1948), looking at how the genetic correlation among individuals changes with distance. In his study, Malécot formalized the relationship between genetic differentiation and geographic distance, showing that individuals that are closer in space are also more genetically similar. Malécot's results are extremely important to this thesis, at least qualitatively. Malécot's principle illustrates shows that there is a strong relationship between geographic distance and genetic distance, implying that gene flow and dispersal are somehow related. However, as mentioned before, natural populations are rarely continuously distributed in space. Wright (1943) identified this shortcoming, and suggested that a more likely scenario in which immigrant individuals would come preferentially from populations that were spatially closer.

In 1953, Motoo Kimura (1953) filled the gap noticed by Wright by presenting the stepping-stone model of population structure. This model describes a large population subdivided into subpopulations aligned one after the other in one dimension, or on nodes of a grid in two dimensions, or on the nodes of a cube in three dimensions. Subpopulations are equally spaced, and migration is stepwise. In other words, individuals are only exchanged among adjacent subpopulations. In addition, allele frequencies only change in response to mutation and genetic drift pressures. Following the presentation of the stepping-stone model, Kimura published two papers in conjunction with George

Weiss in which they analyzed the effects of isolation-by-distance (Kimura & Weiss, 1964; Weiss & Kimura, 1965) within Kimura's model of population structure. The authors derived expectations as to the correlation in allele frequencies between populations that were at any distance (in this case, steps) apart, in one, two and three dimensions. Their results show that the correlation in allele frequencies among populations falls exponentially with distance. In addition, they show that the rate of differentiation increases with the number of dimensions. Essentially, they found the same results that Malécot and Wright did before, however they used a model of population structure that was much more realistic.

The studies of Wright, Malécot and Kimura on isolation-by-distance have been essential in establishing the role of space in shaping the distribution of alleles among populations and individuals. Furthermore, in spite of their different approaches and some inconsistencies in the assumptions of the models (in particular Malécot's model, see Felsenstein, 1975), they all obtained similar results (for a review see Nagylaki, 1989). In a broader context, much of the subsequent work in the field of theoretical population genetics has been based on the theories developed by Wright, Malécot and Kimura, but also those by Fisher and Haldane. Most of this work has been focused on (1) looking at different models of population structure (e.g. Sawyer & Felsenstein, 1983), (2) relaxing or violating some of the assumptions of the original models (e.g. the effects of finite population size in a continuous population, Maruyama, 1972), and (3) devising ways of measuring isolation-by-distance in natural populations (e.g. using coalescent methods, Slatkin, 1993; or, spatial autocorrelation methods, Sokal & Wartenberg, 1983).

Other patterns of spatial distribution of genetic variation

As important as it is, isolation-by-distance is not the only pattern of spatial distribution of genetic variation. Given the principles outlined by Wright (1931), other patterns can arise depending on the interplay between gene flow and genetic drift. In the island model, the degree of genetic differentiation among the different subpopulations is inversely proportional to N_em (Equation 1.1). If $N_em \rightarrow 0$, then $F_{ST} \rightarrow 1$ and the gene pools contained within each subpopulation start to evolve independently. In other words, genetic drift becomes stronger. Spatially, this translates into a break in the distribution of allele frequencies, often referred to as a genetic boundary to gene flow[1] (*cf.* Manel *et al.*, 2003). The barrier can be any number of ecological or physical characteristics of the landscape (e.g. Piertney *et al.*, 1998). On the other hand, if $N_em \rightarrow \infty$, then $F_{ST} \rightarrow 0$ and the subpopulations act as one large panmictic population, with alleles spread out randomly in space. And, finally, if we relax the assumption of the island model of constant size for each subpopulation, allow for varying levels of migration among subpopulations, and include the possibility of extinction and re-colonization a metapopulation genetic structure will emerge (Hanski, 1998). In this case, genetic variation tends to be structured according to the dynamics of the metapopulation. Varying levels of gene flow among different habitat patches will result in varying levels of differentiation among subpopulations. Ultimately, the amount of differentiation among subpopulations in this case will be $0 > F_{ST} < 1$. Exactly where a population and its subpopulations lie on this

[1] It is important to note that the term "genetic boundary to gene flow" is somewhat misleading, as the barrier is in fact not genetic. However, this is the term used in the literature, and as such it will also be used here.

continuum, and why they occupy that position is the focus of the research described in this thesis.

Finally, spatial gradients in allele frequency at a locus, or character, may be observed (Huxley, 1938). Clines, as they are usually called, can be caused by several processes acting on their own, or in combination with one another (Endler, 1973; Vasemägi, 2006). Among these are drift, restricted gene flow, and secondary contact hybridization. However, clines are more commonly associated with spatial differences in selective pressure. Habitat heterogeneity can lead to spatial differences in the distribution of alleles, as different variants at a locus have differential selective advantage in different parts of the habitat. The pattern that emerges will depend on the spatial distribution of selective forces, the degree of dominance at the locus, amount of gene flow and genetic drift, the comparative disadvantage of the heterozygote, and the spatial scale of habitat heterogeneity relative to the strength of the selective pressure (Slatkin, 1973; Slatkin & Maruyama, 1975). The steepness of the cline is indicative of the level of differentiation between different parts of a species' habitat; the steeper the cline, the more genetically differentiated the parts (Endler, 1973). And, under the assumption that the cline is adaptive (for a discussion on when this might not apply see Vasemägi, 2006), the rate of change in allele frequencies in space is a measure of the strength of natural selection (Haldane, 1948).

Conclusions

In conclusion, the main models used to develop current population genetic theory are based on assumptions about the structure of populations, the mode and rates of dispersal,

and other simplifying assumptions about the organism and its surrounding habitat. These simplifications are essential to make the models intelligible and informative. In addition, the simplifications are made to generate generalized, and testable, expectations about patterns of evolution, which help identify the most important factors underlying the observed pattern. As such, incorporating knowledge of ecological factors influencing evolutionary processes can refine these earlier models, particularly as our understanding of certain processes becomes clearer. Currently, two fronts are helping advance population genetic theory. First, a deeper understanding of how dispersal and other ecological factors influence the evolutionary process is helping test and refine theoretical models of genetic variation (for review see Lawson Handley & Perrin, 2007). Second, advances in statistics (coupled with computing) and DNA technologies are helping identify empirical patterns of genetic variation in nature at finer spatial scales (for review see Storfer *et al.*, 2007).

Ecological processes and population genetic patterns

Structuring of genetic variation arises when individuals do not breed at random with respect to genotype and relatedness. Ecologically, departures from panmixia can be related to mating systems and social structure (Storz, 1999), dispersal barriers (e.g. Ernest *et al.*, 2003) and spatial distribution of individuals or populations (e.g. Peakall *et al.*, 2003). These processes lead, directly or indirectly, to reproductive isolation and a decrease in N_e. Social structure and mating systems determine which group of individuals is more likely to contribute to the next generation's gene pool. Usually, this group is much smaller than the set of individuals of breeding age, and exactly how large that

subset is may vary among the sexes. Dispersal barriers and spatial distribution of individuals and populations also act to reduce the probability that an individual will mate. However, these impediments to mating arise from an interaction between the individual organism and its environment, while the social and reproductive constraints mentioned above arise from an interaction between individuals in a species.

Social structure

In mammals, social structure seems to be most commonly associated with polygynous mating systems coupled with male natal dispersal and female philopatry (Greenwood, 1980). A polygynous mating system implies that only a small, non-random portion of the gene pool will contribute to the following generation. In addition, sex-biased dispersal implies that part of the genetic variation stays fixed in space. If we assume an island model, in which each social cluster is a subpopulation, we can use Wright's equation (Equation 1.1) to derive expectations about the distribution of genetic variation among groups. In the case where only a small portion of males are breeding (i.e. small N_e), and in which almost all males disperse but only a small portion of them survive to breed in new groups (i.e. small m), the effective dispersal rate, N_em will be proportionally small, and F_{ST} will therefore be large. Indeed, this seems to be the case for some species (Storz, 1999; Sugg *et al.*, 1996), and is a direct response to the compounded effect of the social system on N_e and gene flow. In the case of howler monkeys (*Alouatta seniculus*, Pope, 1992), for example, migrating females are not permitted to establish themselves into troops that already have a dominant matriline. Furthermore, a male who wishes to enter into a troop has to remove the dominant male. Successful over-throwers do not come

along very often, less than one per female generation. These strong barriers to gene flow coupled with small N_e create a spatial distribution of genetic variation that tracked the troops' territories. Similar results are seen in Soay sheep (*Ovis aries*). Heft territoriality contributes to a spatial distribution of genetic variation that is highly correlated with the geographic location of each heft (Coltman *et al.*, 2003). In another case, wintering grounds coupled with female philopatry helped create genetic boundaries among different groups of white-tailed deer (*Odocoileus virginianus*) in spite of the potential for interbreeding during the summer months (Mathews & Porter, 1993).

Nevertheless, there are situations when social barriers are not strong enough, and other spatial patterns emerge. In the case of the howler monkeys, two sets of troops were observed in different, but adjacent, habitats. In one habitat, troops had been established for some time, while in the other, colonization was relatively recent. In the former, high F_{ST} values were observed among the troops, while in the latter genetic differentiation was not significantly different from zero. The difference in levels of genetic differentiation is believed to be related to the stage of maturity of the two populations within the metapopulation cycle and to the biology of the species. In the older habitat, troops were already established, territories were delineated, and matrilines had been formed. Migration among troops was rare, except for new alpha males that removed their predecessors. Therefore, there was little space for new troops, and it was hard to infiltrate into the established ones. In the habitat under colonization, new troops, which in this species are formed by unrelated individuals, were constantly being formed. Matrilines were established later by female competition; and migration into troops was still rather flexible. As such, genetic variation was more evenly spread in space than would be

expected if colonization had happened farther in the past, and when compared to the older population nearby. The same pattern of low genetic differentiation is observed among social groups when the reduction in N_e is offset by increased success of dispersing individuals to enter and breed in new groups (i.e. higher m). However, in these cases, high genetic differentiation should be observed among social groups if the chosen molecular maker is exclusively transmitted by the philopatric sex (e.g. mtDNA in cases where the female is philopatric, Melnick & Hoelzer, 1992). In so far as primates are concerned, this latter case seems to be more common (Melnick, 1987) than what was observed in howler monkeys.

From these examples, we can see that some particular features are common to establishing strong social barriers leading to genetic differentiation at the scale of social groups (McCracken & Bradbury, 1977). First, as reflected by the polygynous mating system, there is a high variance in reproductive success among males. Second, there is a strong geographic attachment of females. And, third, dispersal into groups can be difficult. But, it is possible that if some of these conditions are not met, the expected pattern in genetic differentiation will not be realized. For instance, the greater spear-nosed bat (*Phyllostomus hastatus*) is a highly social species that displays strong polygyny and female philopatry. However, genetic differentiation among harems is low because both male and female juveniles disperse randomly to other breeding groups (McCracken & Bradbury, 1977). Therefore, the presence of some degree of social structure does not necessarily imply genetic structure (at least at nuclear loci). On the other hand, lack of social structure does not imply a lack of spatial structuring in genetic variation. Storz (1999), for instance, cautions that the same patterns of genetic differentiation attributed to

social structure can arise without sociality, but simply through demographic parameters that lead to patterns of mating and dispersal similar to those of social mammals (for review, Lidicker & Patton, 1987). In these cases, spatial factors might also be at play.

Spatial distribution of individuals and populations

Geographical space will influence genetic structure through an interaction between spatial heterogeneity (*cf.* Li & Reynolds, 1995) and dispersal (Olden *et al.*, 2004). More specifically, spatial heterogeneity modifies how effectively distant (*cf.* Verbeylen *et al.*, 2003) individuals are from one another (Schooley & Wiens, 2003; With & Crist, 1995; With & King, 1999). As such, it changes the degree of functional connectivity among populations and individuals (McIntyre & Wiens, 1999; With & Crist, 1995). If we go back and think in terms of Wright's island model, functional connectivity is proportional to $N_e m$, or the effective number of dispersers. The higher the level of connectivity, the closer the different subpopulations would be to genetic homogeneity, and the closer the spatial distribution of genetic variation would be to random. On the other side of the spectrum, the lower the connectivity, the closer the subpopulations would be to behaving independently from one another, and the more geographically structured the genetic variation. But, before we can apply this model to natural populations and landscapes, we have to determine (or estimate) levels of functional connectivity within the studied landscape. To do this, we first have to take into account spatial scale.

Habitat heterogeneity, and therefore functional connectivity, is scale dependent. The appropriate scale is dictated by the species one wishes to study. For any particular species, there is a range in spatial scales, which include a range of landscape features to

which it is capable of responding. For any group of species, there can also be overlapping ranges (Calabrese & Fagan, 2004). As an extreme example, Wallace's line is beyond what most non-volant terrestrial species have been able to respond to over evolutionary time. In this case, functional connectivity across that line is very close to zero. The isolation caused by this landscape feature (i.e. persistent water barrier) has culminated in the evolution of two separate biotas in relatively close geographical proximity (Brown & Lomolino, 1998). On the other hand, a group of species might have very different responses to landscape features. For instance, for many species of bats, Wallace's line does not present a barrier for dispersal (Hall *et al.*, 2004).

Within this framework, the three basic expectations from Wright's island model as to the levels of genetic differentiation among subpopulations will arise. Subpopulations will either be essentially panmictic, or essentially genetically independent, or somewhere in between. The first pattern would arise if the spatial scale of the species' range were too small compared to the spatial range that the species responds to (e.g. bats species that can cross Wallace's line). The second is at the other end of the spectrum, arising when the spatial scale is too broad relative to the species capacity to respond (e.g. most terrestrial vertebrates on either side of Wallace's line). The third one would emerge when the spatial scale is within the range of the species ability to respond, and one would probably observe some departure from panmixia. While this is a simplistic view, it is useful to demonstrate the effect of scale. If, on the other hand, we assume isolation-by-distance in a stepping stone model of population structure, a more visual representation of the patterns can be obtained.

In this scenario, the first case implies that gene flow is much stronger than genetic drift over the species' range (i.e. case II in Hutchison & Templeton, 1999). In this case, we would expect low, and relatively constant genetic differentiation among subpopulations irrespective of distance. The second case implies that genetic drift is more powerful than gene flow over the species' range, and therefore there would be great variability in differentiation among subpopulations irrespective of distance (i.e. case III in Hutchison & Templeton, 1999). In the third case, two different patterns may arise. If the scale is broad enough that at the biggest distance between two subpopulations gene flow and genetic drift are close to equilibrium, then genetic differentiation should, on average, increase with distance. In addition, because drift's influence will increase with distance an increase in variability should also be observed (i.e. case I in Hutchison & Templeton, 1999). On the other hand, if the scale were broader, so that gene flow-genetic drift equilibrium is included somewhere within the species' range, a similar pattern as before would arise. But, genetic differentiation would only increase until the point of gene flow-drift equilibrium. After this point, genetic drift would be stronger than gene flow, and levels of differentiation would not be correlated with distance anymore (i.e. case IV in Hutchison & Templeton, 1999). For any given species, the pattern that is observed will depend, not only on the species' capacity to disperse over its distributional range, but also on the spatial scale of the study. It is easy to visualize how these scenarios could fit within the range of a single species. At the extremes we would see cases one and two, and in the middle we would see the scenarios in case three. Moreover, if the size of the landscape is kept constant, it is possible to hypothesize that different species would display different patterns, depending on their ability to respond to the selected landscape.

This brings us back to the start of this section. As mentioned earlier, it is the interaction between a species and the heterogeneity found in its surrounding landscape that determines the species' ability to respond to the landscape. Yet, the models reviewed here assume a uniform landscape. In some cases, this is entirely plausible, as can be assumed for some sections of the Amazon rainforest, or the northern forests of North America and North Asia, for example. Nevertheless, any landscape feature has the potential to hinder dispersal and lead to a structuring of genetic variation, unless proven otherwise. In pumas (*Puma concolor*) in California, for instance, patterns of spatial genetic structuring were found to correlate well with suitable habitat in the study area. In addition, some of the more dramatic barriers occurred over relatively short distances, where highly unsuitable habitat areas separate suitable areas (e.g. Figure 2 in Ernest *et al.*, 2003). Furthermore, significant isolation-by-distance was found along continuous areas of mountain lion habitat that were significantly larger than the species dispersal capacity ($r^2=0.52$ $p<0.0001$, Ernest *et al.*, 2003). In another study, conducted with red grouse (*Lagopus lagopus scoticus*) in northern Scotland, analysis of the genetic variation based on microsatellite markers revealed an important barrier to gene flow spatially correlated with a river system that was not initially marked as unsuitable habitat (Piertney *et al.*, 1998).

Overlaying spatial structuring of genetic variation onto geographical maps can help identify potentially important landscape features (Manel *et al.*, 2003). Quantifying landscape connectivity, or the effect of the landscape on the structure of genetic variation is less straightforward. One method combines GIS techniques with isolation-by-distance. Within a GIS database, a species' habitat preferences are used to define suitable

movement paths in a given landscape. Based on this map (sometimes referred to as a friction map, Broquet *et al.*, 2006), minimum effective distances are calculated among samples or populations. These distances are usually larger than simple Euclidian distances, because they attempt to maximize the length of the path that is contained within suitable areas for the species, while minimizing the total length of the path. Using an isolation-by-distance framework, one evaluates which of the two geographical distances (effective or Euclidian) better explains the variation seen in genetic differentiation (Keyghobadi *et al.*, 1999). If effective distances explain significantly more of the variance in genetic differentiation, and are significantly different from Euclidian distances, then one has quantitative evidence of a landscape effect. And, the magnitude of the effect and comparisons among species within the same landscape could be carried out by evaluating the residuals of a regression of effective distances on Euclidian distances (Broquet *et al.*, 2006).

Finally, landscape heterogeneity only modifies a species' potential to disperse. In other words, no matter how suitable the habitat may be, individuals are still limited by their natural history traits. A species' dispersal potential is determined by its life history traits. In mammals, dispersal distance is highly correlated with body size and significantly dependent on trophic level (Sutherland *et al.*, 2000). A more thorough discussion as to the reasons behind this will come later (see Chapter 4). For now, it is sufficient to outline that larger animals (i.e. greater body mass) disperse, on average, over larger distances than smaller ones; and, carnivores disperse farther, on average, than herbivores of the same size.

In conclusion, extant spatial structure of genetic variation will be affected by the dispersal potential of a species modified by features of the landscape across which the species is to disperse. The observed spatial structure will also depend on the scale of the analysis. The distinction between the realized and the observed genetic structure is important, and should be emphasized. Within a species range, the interaction between and individual and the surrounding landscape will determine if and where gene flow and genetic drift will be at equilibrium. In an ideal homogeneous landscape, the resulting genetic structure will depend on the species median dispersal distance. As the landscape becomes more heterogeneous, the resulting genetic structure will be increasingly more dependent on the landscape. The observed genetic structure, on the other hand, will depend on the scale chosen relative to the realized pattern of spatial structure, and will influence what conclusions one draws.

Objectives

The principal objective of this study was to examine differences in spatial genetic structure among six mammal species differing in both their body size and trophic level within the same landscape. To accomplish this I (1) used Bayesian methods to re-construct the current spatial distribution of genetic variation for each species, (2) compared inferred structure among species using F_{ST}, (3) overlaid the inferred spatial structure on a map of the study site to identify for each species, and compare among species, important landscape features, and (4) compared Euclidian and effective distances among samples of each species within an isolation-by-distance framework to examine differences in landscape connectivity among the species.

Hypotheses

The nature of this study requires a comparative framework. To examine the effects of body size and trophic level, I chose to compare six mammal species within the same landscape context (Table 1.1). Four of the species (jaguars, pumas, ocelots and lowland tapirs) display no social structure, and individuals are solitary. This aspect is important, as I want to minimize the effects of social structure on the observed patterns of genetic differentiation. However, the remaining two species (white-lipped and collared peccaries) display different forms of social behavior. The landscape chosen is the *Pontal do Paranapanema*, located on the western tip of the State of São Paulo, Brazil (Figure 1.1). The *Pontal* harbors some of the last remnants of the Atlantic Forest of the Interior, and was ideal for this study because of the wealth of ecological information already available about the focal species. The biology of the species, and a history of the landscape are described in Chapters 4 and 5.

Table 1.1 Species sampled in this study.

Species	Common name	Trophic level	Average weight (kg)	Relative size
Panthera onca	Jaguar	Carnivore	135	Large
Puma concolor	Puma	Carnivore	45	Medium
Leopardus pardalis	Ocelot	Carnivore	15	Small
Tapirus terrestris	Lowland tapir	Herbivore	200	Large
Tayassu pecari	White-lipped peccary	Herbivore	32	Medium
Tayassu tajacu	Collared peccary	Herbivore	14	Small

The first thing that is needed to start building hypotheses about the spatial distribution of genetic variation in these species is estimates of the probability of

dispersal over a certain distance. Sutherland (2000) summarized dispersal distances of mammals generating equations that estimate the median dispersal distance depending on the average weight of the species. Sets of equations were generated for herbivores (Equation 1.2) and carnivores (Equation 1.3) separately.

$$D_{median} = 1.45 * M^{0.54} \quad \text{(Equation 1.2)}$$

$$D_{median} = 3.45 * M^{0.89} \quad \text{(Equation 1.3)}$$

Where, D_{median} is the median dispersal distance, and M is the average body weight of the species.

Using the same data, the authors derived the probability of dispersing a certain number of median distances (Equation 1.4 and Figure 1.3).

$$P(distance > x) = e^{-\frac{d}{1.5}} \quad \text{(Equation 1.4)}$$

Where, d is the number of median distances dispersed.

Using equations 1.2 and 1.3, I was able to estimate the median dispersal distance for each species, and scale the probability of dispersal for each species (Figures 1.4 and 1.5). Here, the simple negative exponential used generates the same shape for the curves between carnivores and herbivores, however the dispersal distances (compare *x*-axis of Figures 1.4 and 1.5) for carnivores are much larger than for herbivores, changing the probability associated with dispersing a certain geographic distance.

Figure 1.1 Location of the *Pontal do Paranapanema*

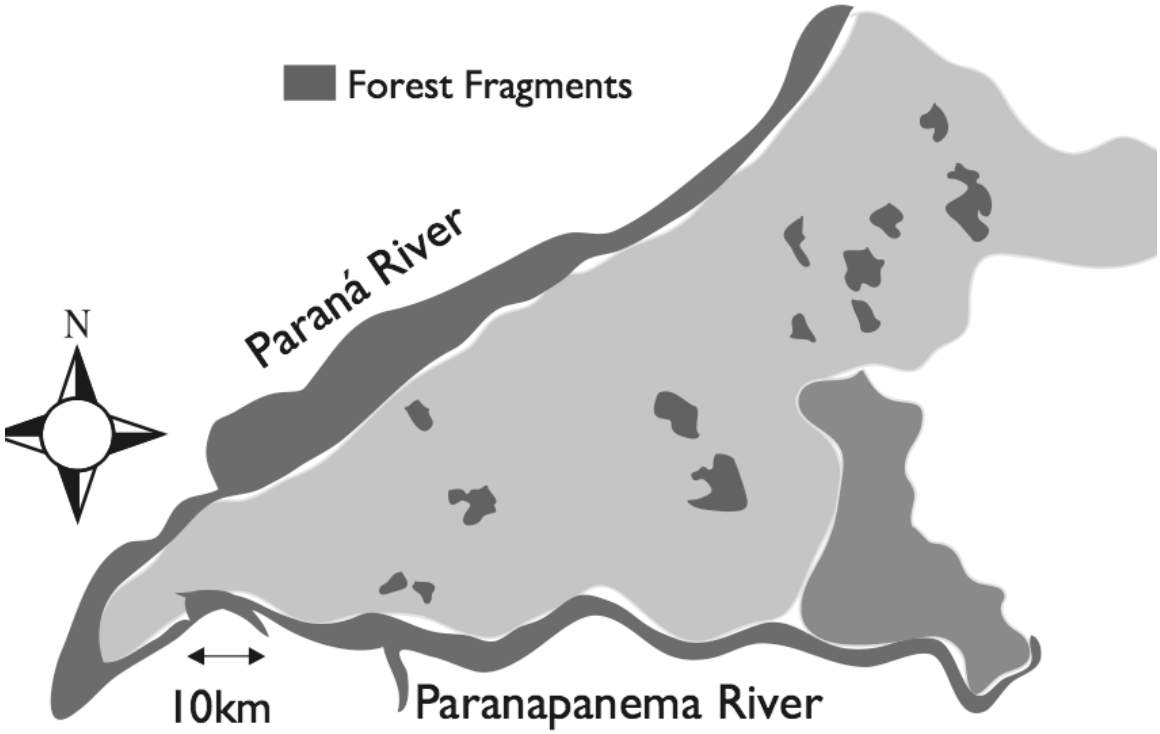

Figure 1.2 The forest fragments of the *Pontal do Paranapanema*

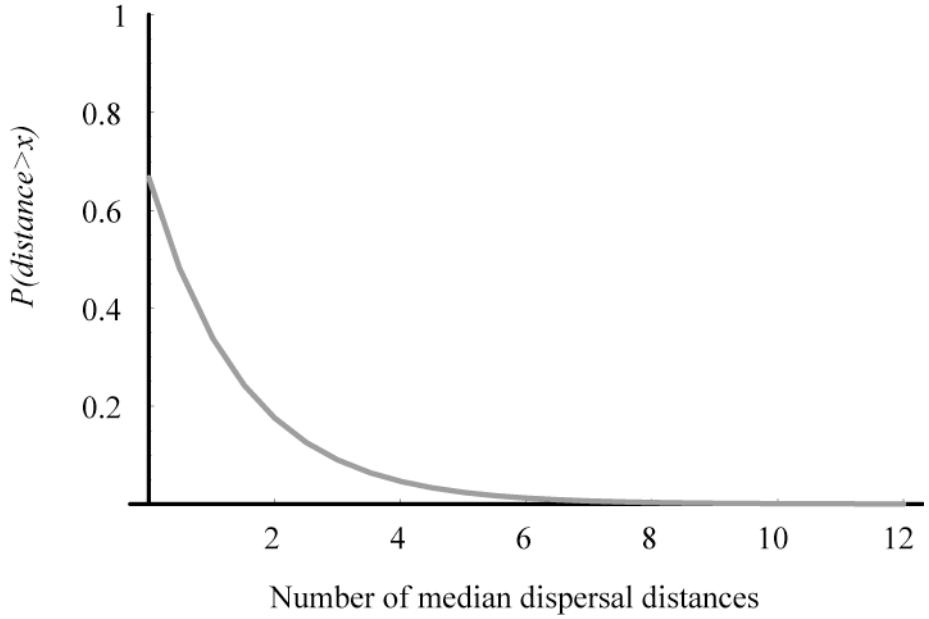

Figure 1.3 Probability of dispersing a certain distance scaled to median dispersal units

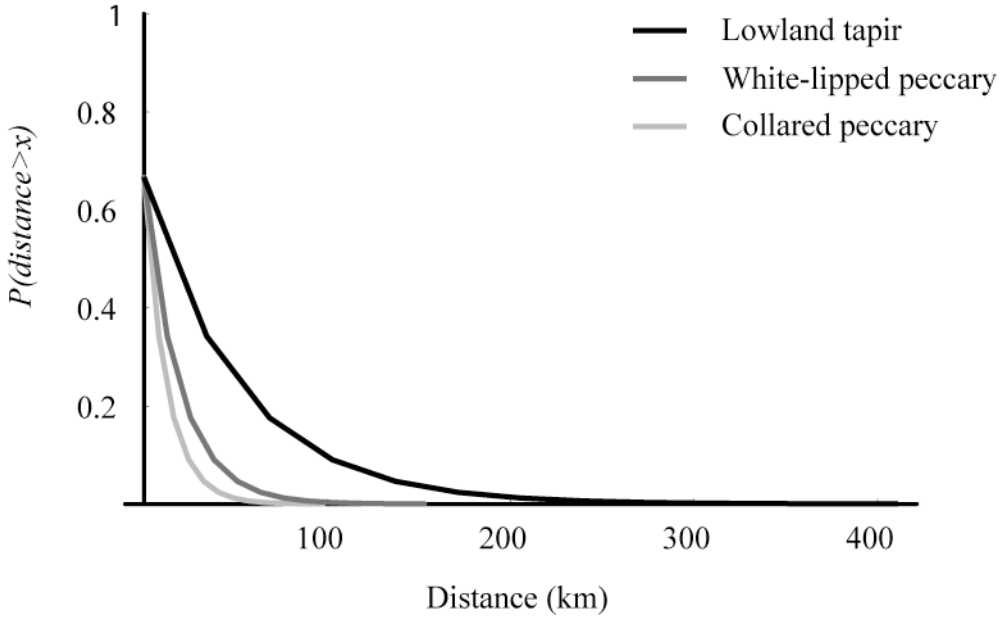

Figure 1.4 Herbivore probability of dispersal based on median dispersal distance estimated using equation 1.2

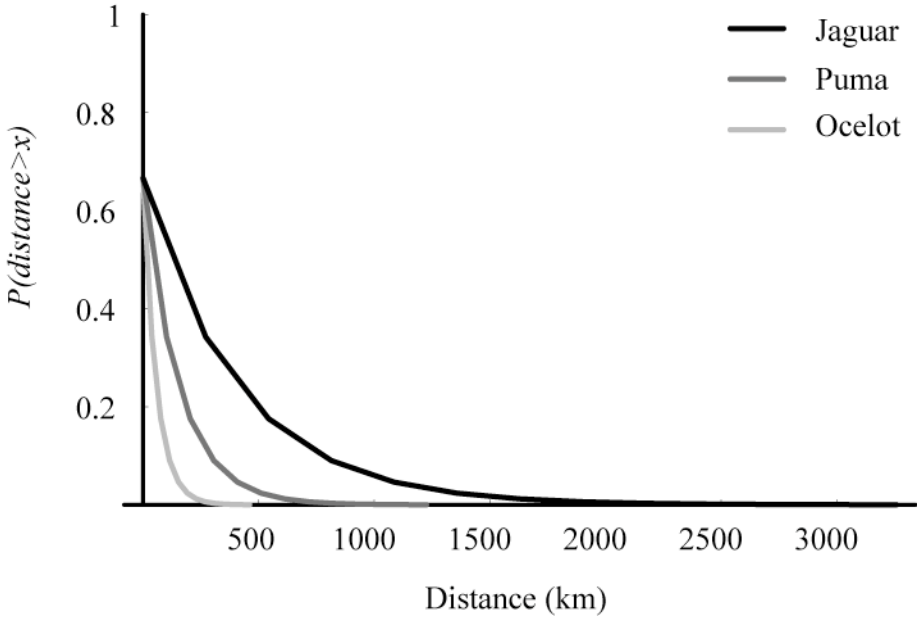

Figure 1.5 Carnivore probability of dispersal based on median dispersal distance estimated using equation 1.3

A priori *assessments of genetic differentiation*

Now that we have probabilities of dispersal for each species based on body size, it is possible to generate expectations as to the level of genetic differentiation that might arise in the landscape under study. If we assume that each forest fragment is a different subpopulation, then we can use Wright's island model (Equation 1.1). However, we would have to assume that emigration rates are constant, but immigration rates (*m*, in Equation 1.1) are modified by the probabilities of dispersing at the scale of the studied landscape. In other words, *m* will be equal to the emigration rate multiplied by the probability of dispersing a certain distance. In addition, for simplicity, we will assume that all species have the same population size of 100. Given these values, and that on average fragments are 20km apart, we can estimate levels of genetic differentiation for each species across different rates of emigration (Table 1.2).

Table 1.2 Estimates of genetic differentiation using Wright's island model across different levels of emigration.

	Emigration Rate			
Species	0.01	0.05	0.1	0.2
Jaguar	0.28	0.073	0.04	0.02
Puma	0.30	0.079	0.04	0.02
Ocelot	0.34	0.096	0.05	0.025
Tapir	0.35	0.100	0.052	0.026
White-lipped peccary	0.51	0.176	0.096	0.051
Collared peccary	0.65	0.278	0.16	0.087

The same can be done if we assume a stepping-stone model, which has the advantage of not depending on the size of the population to estimate gene frequency correlations in space. In Kimura's model, distance is measured in steps between subpopulations (called colonies by Kimura). For the studied landscape, there is an

average 20km between forest fragments. Therefore, if we scale the steps by the median dispersal distance estimated for each species, scale *m* in the same way proposed above for the island model, and assume a constant rate of long range migration (m_∞ in Kimura's notation) of 0.001, the following scenarios can be built (Figure 1.6).

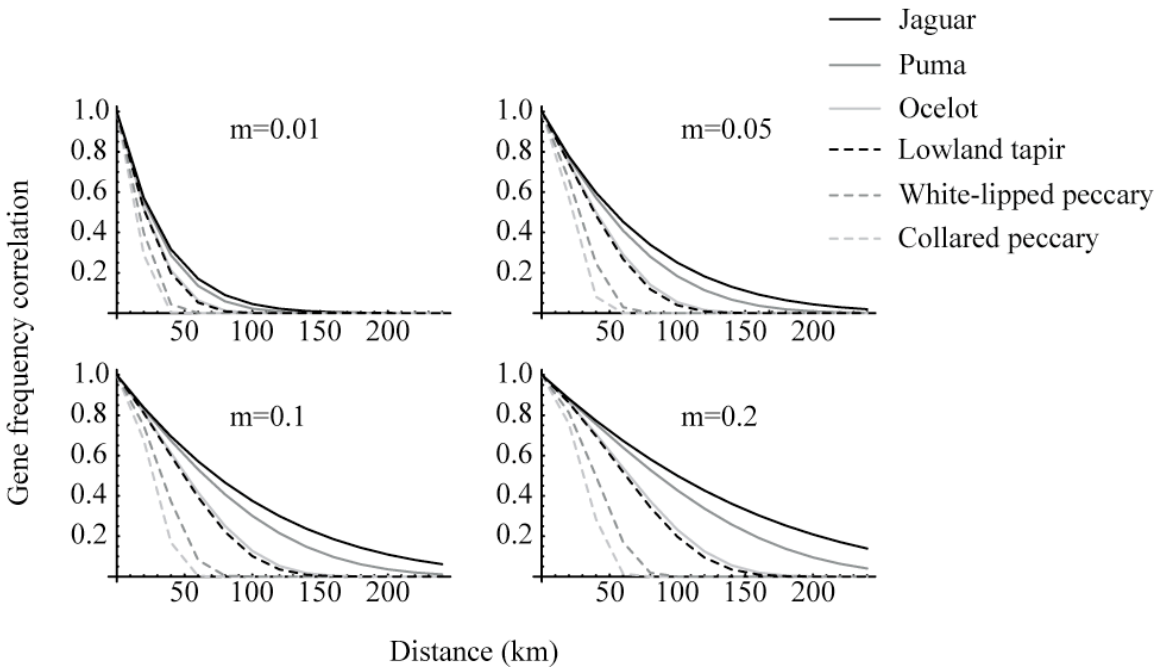

Figure 1.6 Scenarios of genetic variation assuming stepping-stone model in one dimension, across different rates of emigration.

From a qualitative point of view, these two different models generate similar patterns of genetic differentiation across the species being examined in this study. Three groups arise with similar levels of genetic differentiation across the different levels of emigration. According to the models, jaguars and pumas, ocelots and tapirs, and white-lipped and collared peccaries are expected, as pairs, to have similar levels of genetic differentiation within this landscape. However, these models do not take into account the effect of the heterogeneity of the landscape on connectivity. At the moment, there are no

models that incorporate these effects, allowing for a similar analytical approach as illustrated in Figure 1.6. Nevertheless, it is plausible to imagine that restricting connectivity, by deterring dispersal, would increase genetic differentiation. How much, and if the effect would be the same across all species analyzed here, is difficult to estimate. But, given that carnivores are expected to disperse farther, on average, than herbivores of equal size, the studied landscape is expected to have a bigger impact on the herbivores than on the carnivores, and therefore, landscape connectivity is expected to be lower for herbivores than for carnivores.

The models used above, and the conditions used, make several simplifying assumptions. As such, the estimated values are probably incorrect. Nevertheless, the relative magnitude of effects can be a basis for the following hypotheses:

H1.A: Genetic differentiation, as measured by F_{ST}, will decrease with increase in body size within trophic levels

H1.B: Genetic differentiation, as measured by F_{ST}, will be higher for herbivores than for carnivores

H2.A: Landscape connectivity will increase with body size within trophic levels

H2.B: Landscape connectivity will be smaller for herbivores than for carnivores

The structure of this thesis

In the following chapters, I shall explore these general expectations within the framework of landscape genetics. Data on genetic variation for each of the six species was collected from blood and dung samples. Chapter 2 is a critique of the current techniques used to obtain genetic information from dung samples. Chapter 3 is a modeling exercise using

hypothetical scenarios to test the power of different simulated sets of data in estimating genetic differentiation within the limits imposed by this study. Chapters 4 and 5 look at the roles of body size, trophic level, and the landscape itself in shaping genetic differentiation in these species within the *Pontal*. And, finally, chapter 6 discusses the results of this study in the larger picture of ecology and evolution, and of conservation of species and their landscapes.

Chapter 2. A review of methods for collection, preservation, extraction and PCR amplification of DNA from dung samples

To be submitted as:

Gonçalves da Silva A & Melnick DJ **What makes noninvasive dung-based samples work, and what needs to improve?**

Abstract

Our ability to infer patterns in animal behavior, ecology and evolution has been greatly improved by the use of noninvasively sampled DNA. This advance has allowed the collection of large numbers of samples from species that are either difficult to capture, or for which capture is not advised. Yet, while non-invasively sampled DNA allows for more efficient use of field time, it still suffers from a wide range of time-consuming technical limitations. Here, we review the literature on noninvasive dung DNA in search of patterns and protocols that "work" for each step of a field-based genetics project. While we find some common patterns in sample preservation methods, and in the preparation of PCR reactions, the most noteworthy finding is the idiosyncratic nature of the protocols used by each study. The lack of repetition of methods across studies stems, in part, from vague descriptions of protocols. Hence, in many aspects each researcher must re-invent protocols when first starting out in this line of work. Based on our review, we make several recommendations for reaching more efficient and standardized methods to process noninvasively acquired samples, in the hope that this will leave more time to focus on the biological questions of interest.

Keywords: dung, feces, collection, preservation, PCR, noninvasive sampling

Introduction

A little over a decade ago, Constable *et al.* (1995) published the first preliminary study using DNA extracted from dung to study population genetics of a natural population. Since then, studies using dung as the main source for DNA have proliferated in number and application, ranging from population genetic structure of dolphins (Parsons *et al.*, 1999) to species identification of carnivores (e.g. Farrell *et al.*, 2000). The advantages of using dung are obvious, and have been reviewed elsewhere (e.g. Fernando *et al.*, 2003a). However, the sampling of dung for DNA does not come without its limitations (e.g. Vigilant, 2002); the poor quality and small quantities of target DNA usually recovered from dung samples present many methodological problems. Even more problematic, poor samples are only detected once collection and extraction have already taken place. In other words, problems usually surface at the stage of amplification of target sequences with PCR, after considerable money and time have already been spent collecting and preparing samples.

In general, two broad categories of problems can be distinguished, amplification failure and amplification error (Table 2.1). Solutions have been found for most of these problems. PCR inhibitors are becoming less of a concern with the mainstreaming of commercial extraction kits such as the QIAmp Stool DNA Kit (Qiagen, Inc.). Multiple extractions, physically separating extraction and PCR procedures in the lab, and multiple PCR reactions for each sample have helped to reduce problems of contamination. Higher fidelity *Taq*, improved PCR kits and advances in PCR methods, such as hot-start PCR, have increased PCR sensitivity and reliability, reducing allelic dropout, slippage, and failure rates. Primer design still remains a problem, but it can easily be controlled for

with extensive prior testing on reliable DNA sources (e.g. blood or tissue) in both agarose and acrylamide gels. Yet, while there has been some species-specific discussions of factors affecting PCR amplification using template DNA from dung samples (Frantz *et al.*, 2003; Hájková *et al.*, 2006; Nsubuga *et al.*, 2004), there have been few broad ranging analyses of this issue (e.g. Broquet *et al.*, 2007). Furthermore, reporting of methods and results in this body of literature are sometimes less than optimal. Many studies, for instance, simply indicate that dung samples were collected, without noting general sample condition at time of collection, the total amount collected or the amount of sample relative to the preserving agent when one is used, time before extraction, or amount used during extraction. Nevertheless, as pointed out by Fernando *et al.* (2003a), these are important considerations, and can be crucial in obtaining reliable genotypes and DNA sequences. Finally, we have yet to find a comprehensive synthesis of current methods used to recover reliable genetic data from dung samples, with a discussion of each step involved, from collection to reporting of results.

Table 2.1 List of PCR associated problems, and their potential causes.

	Amplification failure	Amplification error
Causes	1. PCR Inhibitors	1. Allelic Dropout
	2. Insufficient target DNA	2. Contamination by other samples or PCR amplicons
	3. DNA degradation post extraction	a. During collection
	4. Unreliable sample	b. During extraction
	a. Too old at collection	c. During PCR preparation
	b. Preserved inappropriately	3. Taq and marker limitations
	c. Species	a. Slippage
	5. Unreliable primers	4. Unreliable primers
	6. Human Error	5. Human errors

Here, we review genetic studies that rely heavily on dung as a main source of DNA in order to identify patterns in protocols that lead to reliable genetic information. We attempt this through a synthesis of current methods and reporting practices, and suggest reporting guidelines that can help in evaluating efficiency of methods and further standardize dung DNA technology.

Methods

We searched for studies in Google Scholar (http://scholar.google.com), an electronic database of peer-reviewed journals. Our search focused on studies of wild populations of mammals, for which sampling consisted mostly of dung, and where the QIamp DNA Stool Mini Kit (SMK, Qiagen, Inc.) was used to extract DNA. We found a large variety of methods for collecting and preserving samples, extracting DNA, PCR amplification, and reporting results. To narrow our search and keep studies as comparable as possible, we kept the DNA extraction method constant. We picked the QIamp DNA Stool Mini Kit because it is one of the most widely used extraction methods and apparently one of the most successful across a range of species, which allowed us to cover a wide range of species and habitats.

In our view there are four fundamental steps to any dung-DNA related project: (1) field collection and preservation of samples, (2) extraction of template DNA, (3) PCR amplification of target sequences, and finally, (4) analysis of reliability. The first three are straightforward; the last one consists of analyzing the obtained genotypes and DNA sequences to evaluate one's confidence in them. For each step, we noted patterns about

fundamental factors considered essential to obtaining reliable genetic data from dung samples (Table 2.2).

Table 2.2 Main factors affecting each step of dung-DNA genetics projects

Field collection and preservation	Extraction of DNA template	PCR amplification of target sequence	Reliability analysis
Species	Method	Marker type and length	Allelic dropout rates
Trophic level	Amount sample used per extraction	Amount extract/PCR reaction	False alleles rates
Local environmental conditions	Final elution volume	Type and amount of *Taq* DNA Polymerase	DNA sequence validation
Sample condition at collection		PCR adjuvants	
Preservation medium		PCR volume	
Ratio sample to preservation medium		PCR method (i.e. hot-startt PCR)	
Storage temperature		PCR success rate	
Storage time			

Results and Discussion

A total of 17 articles were surveyed (Table 2.3), covering 21 species of mammals (14 carnivores and 7 herbivores) and three different molecular marker systems, microsatellite (14), and mitochondrial (6) and nuclear DNA loci (3). Below, we examine each step in order.

Table 2.3 List of reviewed studies

Reference	Species	Location	Preservation Medium[a]	Molecular Markers[b]
Bellemain *et al.* (2005)	Brown bear	Sweden	E 95%	MS
Dálen *et al.*	Artic fox	Scandinavia	S	MS

(2006)				
Epps *et al.* (2005)	Bighorn sheep	Southern California	Dry	N, MD, MS
Frantz *et al.* (2003)	Eurasian badger	UK	D, E 75%, F	MS
Garnier *et al.* (2001)	Black rhinoceros	Zimbabwe	Dry	MS
Goossens *et al.* (2005)	Orang-utan	Malaysia	E 90%	MS
Hájková *et al.* (2006)	European otter	Czech Republic/ Slovakia	E 96%, ASL (Qiagen), SLB (Invitek), ISS (Invitek)	MS
Hedmark & Ellengren (2006)	Wolverine	Sweden	F	MS
Iyengar *et al.* (2005)	Dhole	Indonesia	D, S	MD, MS
Kurose *et al.* (2005)	Tsushima leopard cat Tsushima marten Siberian weasel Domestic cat	Japan	E 100%	MD
Lorenzini & Lovari (2006)	Roe deer	Turkey	E 80-85%	MS
Lukas *et al.* (2004)	Mountain gorilla Western gorilla	Central African Republic/ Republic of Congo	S, R	N, MS
Prugh *et al.* (2005)	Coyote	Alaska	F, D	MD, MS
Russello *et al.* (2004)	Amur tiger	Eastern Russia	S	MD
Scandura (2004)	Italian wolf	Northern Italy	E 90%	MS
Surridge *et al.* (2005)	Wild tamarins	Northern Peru	E	N
Wasser *et al.* (1997)	Black bear	Washington State	E, S	N, MD, MS

[a]D: DET (Seutin *et al.*, 1991); E: Ethanol; F: Frozen; S: Silica
[b]N: nuclear DNA sequence; MD: mtDNA sequence; MS: microsatellite

Collection and preservation of dung samples for genetic analysis

Most studies to date recommend that dung be collected as fresh as possible (e.g. Fernando *et al.*, 2003a), or in conditions that favor the preservation of DNA (e.g. dry, Epps *et al.*, 2005; or frozen, Hájková *et al.*, 2006). Not surprisingly, we found that a total of 41% of the studies reported to have collected samples from dung considered to be fresh. Additionally, 17% reported collecting samples found frozen, and another 17% specified that dung was within a certain age (e.g. up to 5 days old) but had no indication of the distribution of successful amplifications given the age of the sample. Finally, 25% did not report the conditions in which the collected dung was found (Table 2.4).

Table 2.4 Frequency of reported dung sampling conditions separated by climate type (according to the Köppen Climate Classification System, Strahler & Strahler, 1989)

	Fresh	Frozen	Other	Not reported
Alpine	0	1	0	0
Boreal Forest	2	0	0	2
Dry Mid-latitude	0	0	0	1
Mediterranean	1	0	0	0
Moist Continental	1	1	0	1
Tropical Moist Forest	3	0	1	0
Tundra	0	1	0	1
Wet-Dry Tropical Savanna	1	0	0	0

In addition, local environmental conditions are hypothesized to be a major factor contributing to amplification success in DNA originating from dung samples (Hájková *et al.*, 2006; Nsubuga *et al.*, 2004). Hájková *et al.* (2006), for instance, found that temperature at collection can have a significant impact on PCR amplification success (albeit, it is not clear whether the authors controlled for sample age). Because most authors do not report the temperature at collection, or rarely control for this variable during their studies, we assumed that if environmental conditions were an important

factor in determining amplification success, there would be a reporting bias towards climates considered favorable to DNA preservation. As such, we would expect an over representation of drier/cooler climates in the literature compared to wetter/warmer ones. Using the Köppen Climate Classification System (Strahler & Strahler, 1989), we classified each study according to average temperature and humidity. Sampling locations varied from southern California to Borneo, comprising eight different climate types (Table 2.4). Boreal and tropical moist climates were the two most common, with four studies each, suggesting that there is little bias in our sample of the literature in so far as temperature is concerned. Furthermore, it is possible to speculate that the freshness of the sample would confound any effect of local temperature.

Humidity can also be a factor. Presumably, drier conditions would favor the preservation of DNA in dung samples, leading to higher rates of success in amplification (Fernando *et al.*, 2003a). Of the eight climate types observed, 13 of the 17 surveyed studies are described as having 81cm or less of precipitation on average per year. This preference for relatively drier climates does not necessarily reflect a reporting bias towards drier climates, but could simply be related to the distribution of climates in the world, the availability of SMK to researchers, or some other factor.

A total of nine different preservation techniques were encountered, with ethanol (70-100%) and silica being the most widely used (Table 2.5). Four different storage temperatures were encountered, with room temperature and 4 ºC being the more common choices. The most commonly used combination was ethanol stored at -20 ºC, followed by silica at room temperature. However, there were eight instances (~50%) in which storage temperature was not reported. Finally, even though ethanol at -20 ºC was the most

common, no combination was clearly favored, with most combinations appearing only once.

Fernando *et al.* (2003a) also suggested that the ratio of sample to preserving agent may be important in obtaining successful amplifications. Of the surveyed studies, eight had complete information on approximate weight to volume of preserving agent, two had only partial information, five had no information, and for two the information was not relevant, as samples were preserved whole and dry. Of the reported values, five studies indicated using a ratio of 1:4 to 1:5 of sample to preserving agent, across all preserving agents.

Table 2.5 Frequency of storage medium given a certain storage temperature

	ASL[a]	DET[b]	Dry	Ethanol	Frozen	ISS[c]	RNAlater[d]	Silica	SLB[3]
RT	0	0	1	1	0	0	1	2	0
4	0	1	1	1	0	0	0	1	0
-20	1	0	0	3	0	1	0	0	1
-80	0	1	0	0	1	0	0	1	0
NR	0	1	0	4	2	0	0	1	0

RT – Room temperature; NR – Not reported
[a] Qiagen, Inc.
[b] Seutin *et al*. (1991)
[c] Invitek GmbH
[d] Ambion, Inc.

Storage time, even though considered by many as an important factor in determining amplification success (e.g. Palomares *et al.*, 2002, Rita Lorenzini, personal communication), was rarely reported, unless the study was specifically interested in the effect of this parameter on the outcome of PCR reactions. Of the studies that reported storage time, no clear pattern was found. Hájková *et al.* (2006) reported no effect on amplification success after a maximum of 234 days (although, there is no clear indication as to the distribution of success rates over sample storage times); Prugh *et al.* (2005) saw

some reduction in amplification success between samples stored in ethanol for 33 over 45 months (91% versus 79%, respectively), but samples preserved in DET (Seutin *et al.*, 1991) 19 to 31 months longer than ethanol preserved samples performed better than simply frozen samples (83.6% versus 65.9%, respectively); and Wasser *et al.* (1997), while testing several preservation protocols over several markers, also observed an effect of storage time. However, in this last case, there was no clear pattern, with certain protocol/marker combinations improving with age. These results suggest that there is great variability in PCR amplification success with the age of the sample, and that the effect of storage time will be determined by other factors.

Extraction of template DNA

As mentioned before, one of the criteria for choosing studies was that the SMK was used for extractions. However, reporting on the exact protocol used varied, with most studies indicating they used the protocol as suggested by the manufacturer, and some suggesting modifications to increase final yield (e.g. increased lysis and elution times). From a practical point of view, the two most important parameters are the starting amount of sample and the final elution volume. The first one because it determines the final ratio of target DNA to PCR inhibitors; and the second, because it determines the dilution of DNA, and the amount (volume) of extract available to PCR (i.e. the number of PCRs one can carry out on one extraction).

A balance between the target DNA and PCR inhibitor must be found for successful DNA amplification (Fernando *et al.*, 2003a). There is no clear rule to *a priori* decide how much sample one should use, and the amount must be found by trial and

error. For instance, Epps *et al.* (2005) found that too much pigment carried over to their final eluate when they used 100mg initial sample weight during their extractions. Consequently, poor results were seen during PCR amplifications. With a smaller initial sample weight (60mg), the authors were able to reduce pigment carry over and purify sufficient DNA to allow for successful amplifications. On the other hand, Frantz *et al.* (2003) found that 400-600mg of starting material was necessary for successful PCR reaction in the Eurasian badger. Nevertheless, the optimal amount of starting material is rarely reported. Only four studies unambiguously reported the amount of starting material, and a fifth cited a protocol that contained the appropriate information. Of these, three used an initial weight of 100mg, suggesting an initial weight for pilot studies.

According to the manufacturer of SMK using a final volume of 200µl ensures the recovery of ~100% of the DNA. Most studies did not indicate final elution volumes (N=11), four indicated using elution volumes smaller than recommended, and the remaining two followed the 200µl final elution volume suggested in the manufacturer's protocol. There seems to be an inverse relationship between starting sample weight and final elution volume, presumably because the species has a higher or smaller concentration of PCR inhibitors relative to DNA in the sample.

Finally, some authors used different extraction methods and kits along side the SMK. When compared to other methods, the efficacy of SMK (as measured by its ability to produce amplifiable products) varied from study to study. Wasser *et al.* (1997) found that SMK was the best method, alongside a method based on guanidine-thiocyanate combined with diatomaceous earth (GuSCN-DE, Gerloff *et al.*, 1995) for bear samples stored in ethanol and in silica. On the other hand, Frantz *et al.* (2003) found that for

Eurasian badger dung samples the GuSNC-DE method performed better for samples that were stored frozen or in ethanol, and SMK was better for samples preserved in DET. These differences illustrate the variability in results that can be obtained with different permutations of sample, and preservation and extraction methods. Hájková *et al.* (2006), while studying the effects of different parameters on PCR amplification for Eurasian otter fecal samples, decomposed the variance in PCR success into type of sample (three different types of excremental samples were collected: anal jellies, anal jellies/spraints, and spraints), preservation buffer, method of extraction, and an interaction component (sample type with preservation/extraction method). Sample type and the interaction component were the two most important variables accounting for ~98% of the variance in samples that were collected unfrozen, and for ~95% of the variance in samples collected frozen (Hájková *et al.*, 2006).

PCR components needed for successful amplification of target sequences

Obtaining successful PCR amplification from fecal DNA can be quite challenging. The combination of low quantity and quality of DNA with high relative concentration of PCR inhibitors greatly reduces the probability of successful PCR reactions. However, researchers have found different ways to address these issues and improve chances of success. The first thing to consider is the choice and length of the marker (for review see Broquet *et al.*, 2007). Companies specialized in wildlife genetic services, such as WGI, Inc. (Canada), usually restrict analysis of fecal samples only to mitochondrial markers. This is because, in general, there are at least 100-fold more mitochondrial than nuclear DNA copies per cell, making it theoretically easier to extract sufficient template DNA for

PCR of mitochondrial-based markers. In addition, smaller fragments generally have higher success rates than larger ones (Hájková *et al.*, 2006), with many authors designing primers for internal and overlapping fragments (nested-PCR) when attempting to sequence large fragments.

Wasser *et al.* (1997), when comparing amplification success of different size markers, had higher success rates in amplifying a 246bp than a 700bp fragment of the mitochondrial control region from bear samples collected in the wild (80% and 60%, respectively, from a total of 20 samples) and in captivity (90% and 65%, respectively, from a total of 50 samples). The nuclear loci, on the other hand, produced ambiguous results. Two different nuclear markers were tested, the sexing loci SRY and ZFY/X (of 224bp and 442bp, respectively) and the bear specific microsatellite locus G10H (252/262bp range). For samples from captive individuals, 67% of the samples were correctly sexed while 80% were successfully genotyped for G10H, following the expected pattern of lower success rate for larger fragments; but, for samples from wild animals, 85% were successfully sexed while only 67% were successfully genotyped. Thus, in wild samples the larger nuclear fragment had a higher success rate than the smaller one. This apparently counterintuitive result could be a consequence of the smaller average quantity of nuclear relative to mitochondrial DNA, which could conceivably lead to higher variability in results across different groups of samples. Therefore, while it is possible to use nuclear markers, there is a chance there will be high variability in results, particularly for longer fragments. All studies with microsatellites used loci with 250bp or less.

For the PCR itself, there seems to be a general pattern of using (1) higher concentrations of *Taq* DNA polymerase relative to that recommended by the manufacturers, (2) bovine serum albumin (BSA), and (3) hot-start PCR (Birch *et al.*, 1996) coupled with a high number of PCR cycles. These characteristics tend to make a PCR reaction more expensive than it normally would cost with higher quality DNA. However, they seem essential to offset the effects of PCR inhibitors and increase the specificity of the PCR reaction. Increasing the concentration of *Taq* is one way to counteract the effect of PCR inhibitors, with up to 10 times the recommended concentration being used in some cases (e.g. Pääbo *et al.*, 1988), since it assures that at least some enzymes will be available for the reaction. Most manufacturers (e.g. AppliedBiosystems, and Qiagen, Inc.) recommend 1.25U of *Taq* DNA polymerase for 50µl reaction, for a final concentration of 0.025U/µl. On average, the surveyed studies used 0.66U (± 0.32) of *Taq* for an average 21.7µl (± 14.97) reaction, translating into an average final concentration of about 0.030U/µl, or a 0.005U/µl increase from the recommended concentration.

PCR adjuvants are another way to offset PCR inhibitors, of which the most popular among the surveyed studies is bovine serum albumin (BSA). It has been suggested that phenolic compounds will bind preferentially to BSA, rather than to the DNA polymerase (although, it does not seem effective against bile salts), and therefore actively sequesters inhibitors from the solution (Kreader, 1996). Phenolic compounds are particularly abundant in degrading plant materials (e.g. tanic, humic and fulvic acids), and therefore abundant in herbivore dung (Kreader, 1996). The optimal concentration of BSA in a PCR reaction has been reported to be 0.4µg/µl (Birch *et al.*, 1996; Chou *et al.*,

1992). On average, published studies used 0.37µg/µl (range: 0.00032µg/µl – 1.6µg/µl), which is in range with what has been recommended. Finally, hot-start PCR works by precluding *Taq* activity until the PCR reaction is in the thermocycler. This increases the specificity of the PCR reaction and reduces secondary bands (Birch *et al.*, 1996). These factors are essential when dealing with highly degraded DNA in low quantities, allowing for a larger number of cycles (between 35 and 50) to be carried out without excessive background. A total of 10 out of 17 studies employed these methods, with another three studies using 35 or more cycles, even though they did not use hot-start PCR, and two others used hot-start PCR but did not indicate the number of used cycles.

Obtaining reliable PCR amplification

From a methodological point of view, PCR amplifications that are based on such poor quality DNA, and on such small quantities of DNA, face serious constraints on reliability and repeatability. This seems particularly problematic with genotyping (Taberlet & Luikart, 1996). In this particular case, one is generally faced with two problems: (1) allelic dropout (ADO), and (2) false alleles (FA) (Pompanon *et al.*, 2005; Taberlet *et al.*, 1996). Three basic solutions have been proposed to circumvent these problems: (1) repetition, (2) augmentation, and (3) quantification. Repetition, as it has been described in Taberlet *et al.* (1996), is based on using multiple tubes for each PCR (at least seven for homozygotes) to ensure having the correct genotype 99% of the time. However, due to the increase in cost of repeating each PCR so many times and the scarcity of most dung samples, the majority of studies either undertake a preliminary study with a small subset of samples to identify the optimal number of repetitions for their case (usually around 4),

and/or use a different approach for heterozygotes and homozygotes; undertaking fewer repetitions for the former, as true homozygotes are harder to detect (Taberlet *et al.*, 1996). Repetition, or multiple tube approach, is the most widely used alternative to date (8 of 17), probably because of its ease of implementation.

Augmentation, as we call it, consists of running a two-step PCR, in which the first step is a short multiplex reaction and the second is a standard, single locus PCR using the multiplex reaction as a template (Piggott *et al.*, 2004). The logic is for the first step to increase the amount of template DNA for the second step of genotyping. In their study, Piggott *et al.* (2004) found this approach not only to be more reliable for the species they examined, but also to use a smaller amount of the DNA extract, allowing for a larger number of loci to be examined. However, this method is usually used in conjunction with the repetition approach rather than substituting for it, and therefore, does not necessarily reduce the number of PCRs one has to carry out to produce reliable genotypes. Furthermore, while the results were encouraging, the first step still requires a minimum amount of starting material. If there is ADO during the first step, it will carry over undetected to the second step. And, finally, it is not clear what effect the combination of primers might have on the multiplex reaction. Therefore, we would recommend extensive testing of the primer combinations on highly diluted DNA from a reliable and plentiful source (i.e. blood) before using this approach on scarce dung samples. Perhaps because of these issues, and its novelty, the method has not been tested further, with only one study employing it among the 17 surveyed.

As described above, problems associated with small quantities of DNA can be resolved by intensive sampling or by artificially increasing the template DNA, followed

by intensive sampling. Another way is to precisely quantify the amount of starting template, and use only the samples that are above a certain threshold that assures a good margin of reliability (Morin *et al.*, 2001). Quantification involves the use of real-time quantitative PCR to quantify the starting amount of template used in a PCR reaction. While studying chimpanzees, Morin *et al.* (2001) established that a PCR reaction should have at least 25pg of DNA of the target species to produce reliable genotypes. Because the amount of starting template can be known *a priori,* one does not need to perform several PCR reactions per sample per locus to obtain reliable genotypes. However, the authors warn that preliminary studies should be carried out with a subset of the samples to quantify the minimum amount of DNA required for the target species. In spite of the attraction of the precision afforded by this technique, it has not been widely used in fecal DNA studies (1 of 17). The apparent lack of enthusiasm is probably related to the cost of purchasing and maintaining a real-time PCR thermocycler.

Measuring error rates

Noninvasively sampled DNA is prone to errors. Therefore, detecting, correcting, and reporting error rates to allow for a critical appraisal of the quality of the data and confidence in the results are fundamental aspects of dung-DNA related projects (Paetkau, 2003; Pompanon *et al.*, 2005). In general, the literature on error rates in noninvasively collected samples focuses on genotyping errors at microsatellite loci (Taberlet *et al.*, 1997; Taberlet *et al.*, 1996; Taberlet & Luikart, 1996). However, sequence based molecular markers are also prone to errors (Yao *et al.*, 2004), and should not be ignored. First, we discuss errors related to genotyping, and then address errors in sequence data.

ADO and FA are detected through the comparison of a PCR from a replicate sample (or a replicate reaction) and a reference PCR from a reference sample (Broquet & Petit, 2004). Another option is to build a consensus genotype from several PCR reactions (Epps *et al.*, 2005). The reference sample, in natural population studies, can be a known individual (e.g. that was captured), or another noninvasively collected sample known to be of good quality (Broquet & Petit, 2004). ADO is defined by Pompanon *et al.* (2005) as the "stochastic non-amplification of an allele; that is the amplification of only one of the two alleles present at a heterozygous locus". Through this definition, ADO is detected as a mismatch between the reference heterozygote genotype and the replicate (Pompanon *et al.*, 2005), allowing for an unbiased quantification of ADO rates for each locus or over all loci (Broquet & Petit, 2004). However, because ADO can only be detected at heterozygous genotypes, care should be taken to include only heterozygous samples when calculating ADO rates, with the risk of otherwise underestimating ADO. Detecting FA, on the other hand, is not as straightforward. While FA is also discovered through a mismatch between replicate and reference genotypes, true FA can sometimes be confused with other types of error. FA is strictly defined by Pompanon *et al.* (2005) as "an allele-like artifact that is generated by PCR". This definition restricts errors to those generated by the PCR at the specified locus. Spurious amplifications due to contaminated samples or because of non-specific annealing of primers are not considered FA. It is not always possible to detect the difference among these types of errors, but liberal use of negative controls and exhaustive testing of primers beforehand can reduce their incidence (Broquet & Petit, 2004). Unbiased rates for each locus or over all loci can also be

calculated for FA (Broquet & Petit, 2004). But, calculations should include both heterozygous and homozygous genotypes, as FA can be detected for both.

In most reviews and studies centered on errors in genotyping noninvasively collected samples, strong comments are made about the importance of reporting error rates (e.g. Pompanon *et al.*, 2005), and clear reporting guidelines are presented (e.g. Broquet & Petit, 2004; Taberlet *et al.*, 1996). The importance of reporting ADO and FA rates cannot be overstated. These measures are analogous to measures of error in any statistic, for they allow the reader to evaluate the degree of confidence it has in the data presented. Yet many studies are still published without reporting these measures. In 13 reviewed studies that used microsatellite markers, only four reported ADO rates, another three reported FA, and only two of these reported both values. This means that in less than 15% of the cases we are able to judge appropriately the degree of confidence in published genotypic data.

The frequency and types of errors in DNA sequences are not as widely discussed in the noninvasively sampled DNA literature. Nevertheless, such markers are not immune to errors (e.g. Yao *et al.*, 2004), and one should take this into consideration when analyzing this type of data, particularly from low quality DNA. In general, five different types of error seem to affect sequence data: base shifts, reference bias, phantom mutations, base mis-scoring, and artificial recombination (Bandelt *et al.*, 2001). Discovery of these errors are based on phylogenetic analysis and careful comparisons with independently derived sequences, as found in GenBank, and with consideration to the coding sequence (Bandelt *et al.*, 2001). As such, detecting problematic sequences with the present methods seems more labor intensive than detecting problematic

genotypes. And, probably because of this, most studies that have attempted to quantify errors in sequence data are of a forensic nature or are interested in human evolution, which normally undergo a higher degree of scrutiny. Within this body of literature, there seems to be a growing concern about the quality of published sequence data, with several known examples of published mistakes coming to light in the past few of years (for review, Yao *et al.*, 2004). In the noninvasive DNA literature, such errors have not been reported, even though sequence data are being used. On the other hand, most studies seem to follow good practices to minimize errors, namely, using multiple tubes, sequencing both strands, and comparing sequences with sequences deposited in GenBank. Of the four studies that used sequence data, three reported using methods that allow checking for errors, and one was not clear what protocol was used, if any, to minimize error. However, none of these studies reported error rates, making it hard to gauge the pervasiveness of error and the efficiency of the protocols used.

Finally, errors seem to be as much about the samples and protocols used as they are about the technical capacity of people involved in the project (Paetkau, 2003; Pompanon *et al.*, 2005). Therefore, the variability among labs makes it difficult to compare error rates among studies. In any case, ADO rates were on average ~ 0.08 (± 0.07), and FA rates were on average ~ 0.009 (± 0.008) across studies. Suggesting that ADO is an order of magnitude more common than FA.

Conclusions

At the onset, this review had as one of its objectives to produce a synthesis of best practices in noninvasive genetic studies, and identify crucial parameters that affect the

outcomes of PCR reactions. However, there is still great variation in the description of methods in the published literature, with some descriptions leaving more questions than answers. One of the difficulties we found was variable reporting of success. Most studies reported the proportion of samples that produced amplifiable DNA. However, because measures of data quality are based on PCR success rate (i.e. number of PCRs that produce an amplicon over the total number of PCRs), it would be advisable to report this quantity too. In addition, measures of success rate based on individuals or samples depend on criteria used by the authors, which are not always clear, and vary in how rigorous they are (Petra Hájková, personal communication). Also, because the cost of most studies is measured in number of PCR reactions, this measure makes it easier to evaluate the cost of undertaking a project using such samples. Furthermore, studies using two or more protocols to collect and/or preserve and/or extract DNA, often reported only one, overall, measure of error, and of success rate. Not reporting separate error and success rates for each protocol precludes judging which protocol is more efficient. Additionally, some studies applied a factorial design in sampling (varying preservation techniques and extraction methods, for instance), but did not apply the same design when testing for the effect of the different parameters on PCR amplification success and on error rates. Instead, they tested each parameter separately and, in some cases, conducted tests with an unknown sub-sample of the total samples. This decoupled approach to the analysis hinders a deeper understanding of interactions among the different variables. The interaction among variables seems to be an important factor in PCR success, as was seen above for storage time, and in the two studies that actually tested for interactive effects (Frantz *et al.*, 2003; Hájková *et al.*, 2006). As such, it is currently difficult to draw

concrete conclusions about which method is more efficient. It is our suggestion that a carefully controlled experiment should be made in the field, following a factorial design and varying sample age, sampling methods, preservation techniques, and time to extraction, across several species, trophic levels, habitat types, and molecular markers. Furthermore, we recommend that in microsatellite studies both ADO and FA rates always be reported, and that PCR success rate always be published, regardless of the marker used. Nevertheless, there are certain patterns that are distinguishable across the studies, particularly with respect to PCR. These are:

- When collecting samples, ethanol and silica are good preserving agents. In particular, for a pilot project, collecting small amounts of sample across several tubes and preserving them in ethanol or silica seem to be the most widely used alternative;
- When choosing molecular markers, smaller sized fragments, <250bp, are more likely to amplify than larger ones;
- When preparing PCR reactions, increasing the final concentration of *Taq* polymerase relative to the recommended by the manufacturer seems important, as is the presence of BSA; and,
- When running PCR reactions, starting with a hot-start PCR coupled with 35 cycles or more is advisable.

Lastly, the number of studies using noninvasively collected samples continues to grow, a promising sign that the methods involved are becoming more common, and easier to use. The ability to use these samples, and extract meaningful information from

them, has certainly fueled a major advance in the study of the ecology and evolution of natural populations. But, this growth seems to lack a firm technical base on which to stand. Each study uses a completely new protocol, and most required extensive *a priori* testing of methods before the full study could be conducted. Yet, they have failed to report appropriately what they found. This lack of reporting standards leads to a greater expenditure of money with redundant experiments across studies. This is particularly worrisome in cases where the study focuses on endangered species, for which money is already in short supply. Therefore, in the interest of reaching better and more robust methods for collection, preservation and analysis of noninvasively sampled dung for DNA data, we recommend more thorough reporting of methods used, accompanied by the appropriate measures of data quality. In particular, we recommend that the following information be reported:

- Condition of samples at time of collection (e.g. fresh, dry, etc.);
- Amount sampled, amount of preservation buffer (if one is used);
- Storage conditions and time;
- Amount used for extraction, and dilution volume;
- PCR conditions and components;
- Number of PCR's, number of successful PCR's;
- ADO and FA rates for genotyping (using measures such as described in Broquet & Petit, 2004); and,
- BLASTing of sequences and search for base mis-scoring should be carried out in the case of sequencing, though standards are still in discussion.

We do recognize that the real interest, in the end, is the identification of the taxonomic unit or the degree of population structuring or whatever the biological question one is addressing, and therefore the technical aspects tend to be seen as secondary in importance. But, the biological conclusions of any study depend on the quality of the data reported, and better and more standardized methods would not only create incentives for more people to use them, but also improve our ability to draw meaningful conclusions about the biological system under study.

Chapter 3. The effects of dispersal, isolation, and population size and evenness on the capacity of a Bayesian model to detect deviations from panmixia

To be submitted as:

Gonçalves da Silva A & Melnick DJ **Optimum conditions for detecting deviations from panmixia with a Bayesian model.**

Abstract

In recent years, there has been an explosion of new methods dedicated to analyze genetic data. These methods, based mainly on maximum likelihood and Bayesian approaches, promise to substitute traditional population genetic analyses based on F_{ST}. With these new methods, we are able to use genetic data more efficiently, and consequently we are able to obtain finer details about the evolutionary processes shaping the distribution of genetic variation. However, these methods are not without fault. Their novelty means, among other things, that they have not been as thoroughly tested as F_{ST}, and therefore we are not entirely certain of their limitations. In particular, Bayesian approaches have become popular among researchers who seek to identify the number of unique genetic units in a population. In this case, minimum assumptions are made about the realized structure, and the data itself is used to infer which partitioning of individuals has the most support given the data. The conditions where the model is more powerful in detecting the correct number of clusters are still a matter of debate. Here, we simulate scenarios of recent population fragmentation, over different dispersal capacities, time frames, and population sizes and distribution, and empirically assess the power of BAPS, a software solution for a Bayesian model designed to infer population structure. We show that BAPS is able to infer the correct number of clusters over a wide range of conditions. However, the model's power is significantly affected by population size, differences in population size among subpopulations, and time since fragmentation. Dispersal was not an important factor, however this may be because of assumptions made about the nature of the fragmentation event. We discuss future research necessary to resolve several open issues.

Keywords: BAPS, Bayesian models, F_{ST}, fragmentation, population structure, power

Introduction

Most studies of population genetics have relied on measures of genetic differentiation developed by Fisher (1930), Wright (1931), and Haldane (Haldane, 1932). The frequentist model, as it is sometimes called (Pearse & Crandall, 2004), relies on F_{ST} to measure genetic differentiation among pre-defined groups of individuals, and subsequently infer ecological and evolutionary processes involved in shaping the observed pattern. While this approach has been and will continue to be widely used, as it is usually well understood and allows for easy comparison on levels of differentiation among studies (albeit not on cause and history), it makes use of only a small portion of the information contained in an individual's genotype (Manel *et al.*, 2005; Pearse & Crandall, 2004). This is particularly true in more modern datasets, which are typically made up of many hundreds of DNA bases per individual or several hypervariable microsatellite loci, but get averaged to obtain haplotype or allele frequencies to estimate F_{ST} (Mank & Avise, 2005). Focusing on such a small part of the genetic information available, and imposing an *a priori* organization to individuals may result if there is cryptic population structure, recent immigrants and/or hybrids (Mank & Avise, 2005).

To address some of these shortcomings of the frequentist model, numerous maximum likelihood (ML) and Bayesian analysis approaches have been proposed (for reviews, Manel *et al.*, 2005; Manel *et al.*, 2003), which focus on the individual as the unit of analysis (Mank & Avise, 2005). These approaches differ from the frequentist model in the fundamental way in which the data are treated. In the frequentist model, the data are seen as a sample from a known probability distribution, and are used to test a particular hypothesis (e.g. deviation from Hardy-Weinberg Equilibrium). In the ML and Bayesian

approaches, it is assumed that we do not know, *a priori*, the correct distribution underlying the sampled data, and the data are instead used to estimate the most likely parameters describing the distribution (ML) or the distribution itself (Bayesian). To do so, these methods make use of observed individual genotypes and alleles (and, optionally, sample locations, geographical coordinates, and the like), allowing for more efficient use of available information. As a result, biologists and conservation scientists are presented with potentially more sensitive tools to uncover, among other things, population history, past and present demography, gene flow patterns among populations, and number of distinct population clusters (Pearse & Crandall, 2004). Yet, these methods are not without fault, many are computationally intensive, do not provide user-friendly interfaces, require extensive datasets, and have varying assumptions as to the degrees of differentiation, genetic equilibrium and recombination (Pearse & Crandall, 2004). In addition, because of their novelty, the power and sensitivity to violation of assumptions of these models has either not been assessed, or has only been evaluated for a limited set of conditions (Manel *et al.*, 2005).

In the particular case of identifying population structure, Bayesian models are becoming the most widely used, with several different models having software solutions freely available to download from the Internet. These models differ in their use of the available information to inform prior probabilities (i.e. priors), and on how individuals are treated. Currently there are four main models. One of the most popular models, STRUCTURE (Falush *et al.*, 2003; Pritchard *et al.*, 2000), at its most basic form assumes a uniform (or uniformed) priors, and allows for admixture of individuals, but it estimates the number of clusters using an *ad hoc* procedure. PARTITION (Dawson & Belkhir, 2001),

also assumes a uniform prior, but will treat the number of clusters as a parameter to be estimated directly from the model, but unlike STRUCTURE, it does not allow for admixture. GENELAND (Guillot *et al.*, 2005a; Guillot *et al.*, 2005b) is similar to STRUCTURE, but uses the geographic coordinates of individuals to modify the model's priors. BAPS originally looked for the best partition of individuals taking into account the sampling design to modify the model's priors, and did not allow for admixture (Corander *et al.*, 2003). Newer implementations of the model allow for individuals to be the unit of the mixture analysis (Corander, 2004), and allow for admixture (Corander & Marttinen, 2006) and for use of geographical coordinates to inform the priors (Corander *et al.*, in press). The basic assumption underlying all the above models is that the true genetic units, which are to be uncovered, are in Hardy-Weinberg and linkage equilibriums (Excoffier *et al.*, 2006). The power of these models to correctly identify the number of clusters is directly related to the degree of allele frequency difference among the genetic units. Using simulated datasets with number of loci and alleles similar to what is found in microsatellite datasets, STRUCTURE and BAPS were equally able to distinguish among clusters at relatively low levels of differentiation ($F_{ST} \geq 0.03$, Latch *et al.*, 2006), while PARTITION required much greater levels ($F_{ST} \geq 0.09$, Latch *et al.*, 2006).

In this case, the power each model was evaluated under ideal conditions; at gene flow-genetic drift equilibrium, with enough time in isolation for complete lineage sorting, and equal population sizes for all subpopulations. These conditions are rarely present in nature, especially in a metapopulation setting and/or in cases of relatively recent fragmentation.

To assess the consequences of non-idealized population conditions, we examined the power of a Bayesian model, BAPS, to infer the correct number of clusters after an event of fragmentation that subdivides a large panmictic population into several small subpopulations interconnected by very low gene flow (average of 1 migrant/generation for the whole population). We also investigate the effects of sampling at different times since fragmentation (measured in number of generations) and different total population sizes and distribution of individuals among subpopulations (i.e. equal *N*/subpopulation, or one subpopulation has a larger *N* than all others). Additionally, we consider three different dispersal abilities (as specified by median dispersal distance).

We hypothesize that the power of BAPS at inferring the correct number of clusters when only a small number of generations have passed will be better at intermediate population sizes with even distribution of individuals among subpopulations, and with decreasing median dispersal distance. This hypothesis stems directly from population genetic theory. Even distribution of individuals among subpopulations leads to optimum rates in change of allele frequencies towards a new gene flow-genetic drift equilibrium, and equal rates of lineage sorting among subpopulations (Hartl & Clark, 1997). Smaller median dispersal distances translate into individuals not dispersing further than adjacent subpopulations, increasing the rate of genetic differentiation among populations that are farther apart. Additionally, intermediate population sizes will diverge at faster rates than larger populations, and will contain more genetic information than smaller populations. With increasing number of generations, these differing effects should become less important, as longer periods of isolation lead to increased genetic differentiation and

complete lineage sorting. Therefore, at this later stage, BAPS should be equally able to infer the correct number of clusters over all scenarios.

We focus on BAPS because of its versatility, which allows the user to explore a range of biological questions within one software solution (from identifying the number of population clusters to the estimating number of immigrants), and because of two key functional characteristics that make it a more attractive option than the other models. First, BAPS differs from the other models in its inference framework. The first three models use MCMC (Markov Chain Monte Carlo) algorithms to heuristically sample the posterior distribution of population allele frequencies and/or the number of genetic units (clusters). This imposes a constraint on time and computational power needed to carry out the analyses. Additionally, this framework requires the user to monitor the MCMC chains for convergence (which can be subjective) and to repeat runs several times to assure that the best results have been found. BAPS, on the other hand, will, for reasonable sample sizes (10 sampling locations or less), analytically derive the whole posterior distribution (in a similar approach to Fisher's Exact Test). This not only saves time, but also simplifies the work and knowledge needed to apply the model. Second, out of all the above models, BAPS has the most user-friendly software, with a multi-platform solution that is constantly updated, and which will accept input files formatted for GENEPOP (Raymond & Rousset, 1995), a commonly used format.

Methods

Building scenarios of population structure

To examine the effects of dispersal, number of generations since fragmentation and subpopulation size and evenness, we simulated populations using EASYPOP v2.1 (Balloux, 2001). Three different dispersal levels, three different periods after fragmentation, and six different population sizes (three with equal *N* for all subpopulations, and three with one subpopulation with a larger *N* than all others) were modeled in a factorial design for a total of 54 scenarios (Table 3.1). For each scenario, we simulated diploid randomly mating organisms, with two sexes at equal ratio. In all cases, nine subpopulations were simulated on a square grid using the spatial model option of EASYPOP (Figure 3.1). A total of five unlinked loci with five alleles were simulated. We assumed a mutation rate of 10^{-4} per generation following a k-allele model (KAM, Crow & Kimura, 1970). Populations were simulated over 200 generations before fragmentation, which is sufficiently long for populations to be well mixed but sufficiently short for mutation to be negligible. During the pre-fragmentation period, subpopulations behaved as one large panmictic population. After fragmentation, dispersal was limited to an average one individual per generation. Each scenario was replicated 10 times.

Table 3.1 Parameter values used in this study

Parameter		Values		
Dispersal[a]		Short (1/2)	Medium (1)	Long (2)
Generations since fragmentation		5	15	30
Population Size[b]	Equal	20 (180)	50 (450)	100 (900)
	Unequal[c]	10/100 (180)	16/97 (450)	40/130 (900)

[a] Numbers in parenthesis indicate the ratio of median dispersal distance to the shortest distance among subpopulations

[b] Numbers in parenthesis indicate total population size

[c] First number indicates *N* for smaller subpopulations and the second number indicates the population size of the largest subpopulation

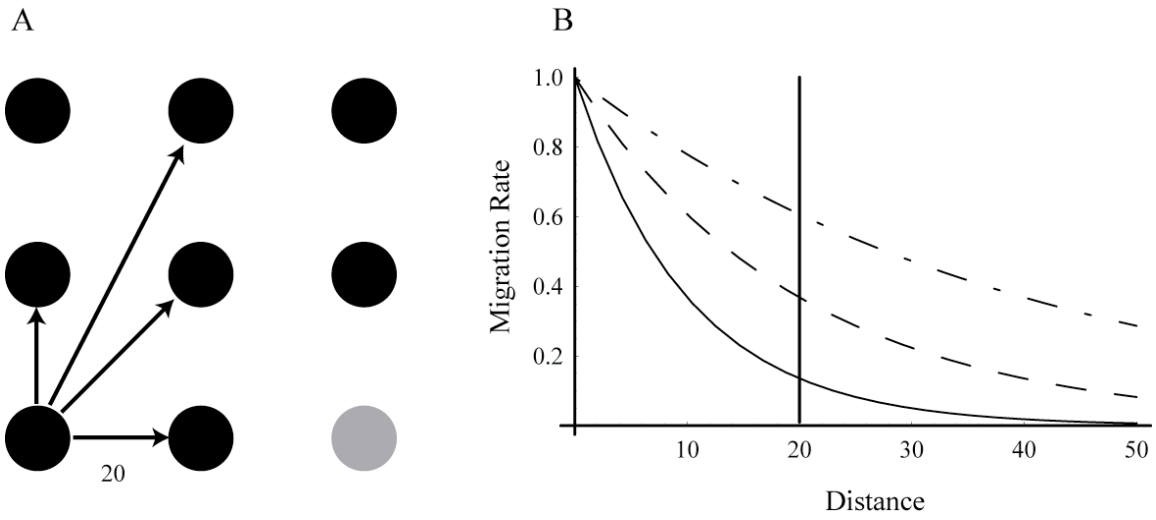

Figure 3.1 A – Spatial distribution of subpopulations. Individuals could potentially disperse to any subpopulation according to the probability function described in panel B. For the simulations carried out in this study, the shortest distance among subpopulations was arbitrarily set at 20. When modeling unequal population sizes, the population with the larger *N* was always assigned to the bottom right hand corner (gray circle). B – Within the spatial model option of EASYPOP (Balloux, 2001), the probability of dispersal from population *i* to *j* ($i \neq j$) is an exponential function of the distance between the population and the species' median dispersal distance. We modeled three different dispersal scenarios: short median dispersal (continuous line); medium median dispersal (broken line); and long median dispersal (broken/dotted line). The vertical bar denotes the largest possible probability of dispersal among subpopulations for the modeled scenarios.

Inference of population structure

At end of each simulation run, an input file was produced with the genotypes of each individual from the last generation grouped by subpopulation suitable for analysis using BAPS 4.14 (Corander *et al.*, 2006). For each of these input files, we inferred the optimal posterior partitioning of individuals by clustering groups of individuals (Corander *et al.*, 2003) and by spatial clustering of groups of individuals (Corander *et al.*, in press) as implemented in the population mixture analysis section of BAPS 4.14 (Corander *et al.*, 2006). We analyzed each input file 30 times, specifying a range of clusters from 1 to a maximum of 20 for each run. The best partition was the one that had the smallest posterior marginal log-likelihood (log(ml)). For each analysis, we noted the number of clusters in the optimal partition, log(ml), and the posterior probability of the partition.

Statistical analysis

To test the effect of each parameter, we performed an ANOVA with the number of inferred clusters as the dependent variable, and dispersal, generations since fragmentation and population size and distribution as independent variables. We also tested for interactions between all pairs of variables, and among all three variables. Two separate ANOVA's were performed, one for results from clustering of groups of individuals and another for results from spatial clustering of individuals. ANOVA was carried out using the built-in function "ANOVA" in *MATHEMATICA* (Wolfram Research, 2004). In addition, we correlated the number of inferred clusters and F_{ST} to test the hypothesis that the ability of the model to detect the correct number of clusters is related to the degree of genetic differentiation among subpopulations. In this hypothesis, it is expected that the

higher the genetic differentiation among subpopulations the more accurate the model will be in estimating the correct number of subpopulations.

Results and Discussion

Scenarios of population structure

To examine and compare the levels of population differentiation in the different scenarios, and ascertain that they are behaving as would be expected by population genetic theory, we plotted average F_{ST} onto number of generations since fragmentation (Figure 3.2). As would be expected, F_{ST} values grow with increasing time since fragmentation. Probably because the effect of fragmentation was equal over all median dispersal distances (i.e. all were reduced to 1 migrant/generation), there was no difference in overall rates of F_{ST} increase among the different dispersal scenarios. When comparing different population scenarios, there is a trend for F_{ST} to increase at faster rates in smaller populations than in larger, and this rate seems to accelerate if, in addition to being small, the subpopulations are of unequal size. Because genetic drift is expected to be stronger in smaller populations, and therefore rates of genetic differentiation are expected to be faster, the observed trends in F_{ST} suggest that the simulated scenarios are behaving in accordance with population genetic theory.

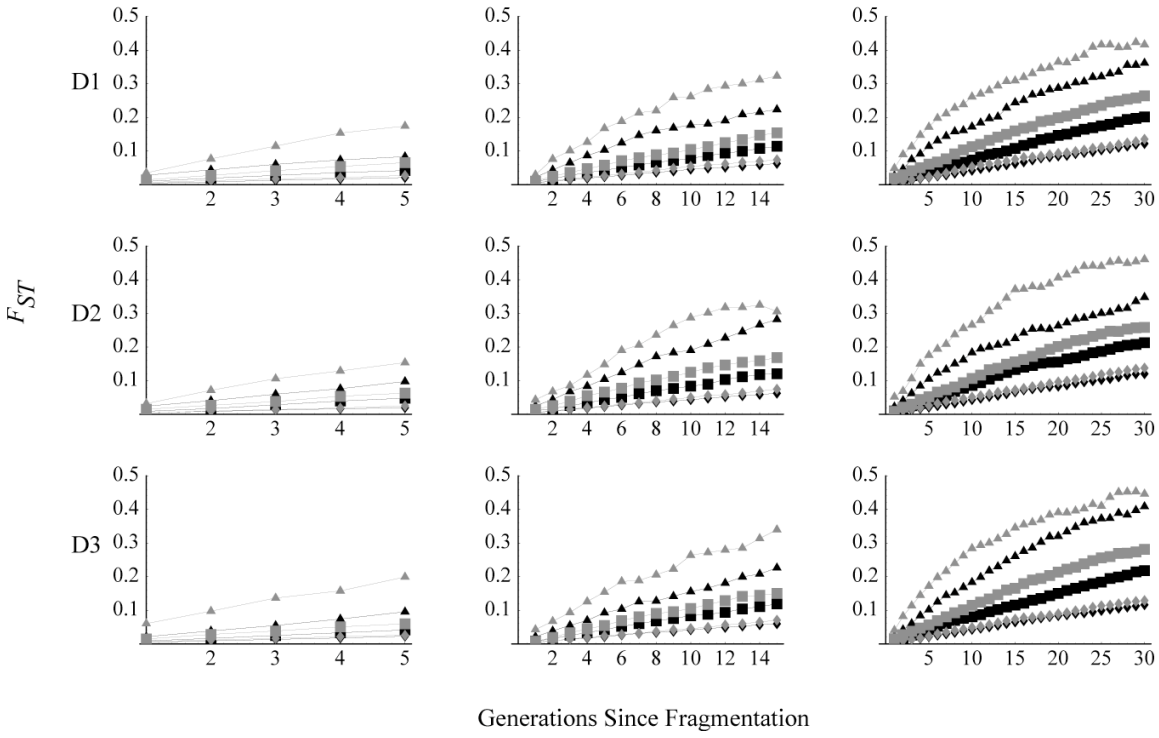

Figure 3.2 Average F_{ST} for 10 replicates for each of the 54 simulated scenarios after the onset of fragmentation. D1, D2, and D3 refer to the short, medium and long median dispersal scenarios, respectively. Black lines and symbols refer to scenarios where all subpopulations are of equal size. Gray lines and symbols refer to scenarios where one subpopulation has a larger population than the other eight. Triangles refer to small population size scenarios, squares to medium population size scenarios, and diamonds to large population size scenarios.

Factors affecting inference of correct number of clusters

The correct identification of the number of clusters (i.e. nine) was primarily affected by population size and number of generations since fragmentation for both clustering of groups (Table 3.2) and for spatial clustering of groups (Table 3.3). However, the interaction component population*generations since fragmentation was also of importance in both types of analyses. This significant interaction makes it hard to examine each component with clarity. Nevertheless, when examining the plot of average number of clusters identified for each scenario (Figure 3.3), it is clear that five

generations after fragmentation is generally too short a period for BAPS to correctly identify the number of clusters in all examined scenarios. However, BAPS was closer at estimating the correct number of clusters for scenarios of intermediate population sizes with even number of individuals across subpopulations, partially corroborating our hypothesis. After 15 generations, the ability of BAPS to identify the correct number of clusters increased significantly. While the increase was noticeable for all examined scenarios at 15 and 30 generations after fragmentation, BAPS still underestimated the number of clusters in smaller populations.

Finally, for the clustering of groups analysis there is a slightly significant effect of the interaction component dispersal*generations since fragmentation (Table 3.2). This is due to differences in the average number of inferred clusters between short and medium median dispersal scenarios at five and 15 generations since fragmentation (Figure 3.4). However, while the same pattern was observed for spatial clustering of groups (Figure 3.4), it was not significant for this analysis (Table 3.3). As such, it is hard to attribute a biological significance to this result without a larger number of replicates to rule out sampling error.

Table 3.2 ANOVA table for clustering of groups

Source	df	*SS*	*MS*	*F*	*p*-value
Dispersal (d)	2	3.14	1.57	2.18	0.11
Number Generations (t)	2	1240.56	620.28	860.81	0
Population (p)	5	82.66	16.53	22.94	0
d t	4	7.18	1.79	2.49	0.04
d p	10	11.82	1.18	1.64	0.09
t p	10	90.26	9.03	12.52	0
d t p	20	14.99	0.75	1.04	0.41
Error	486	350.2	0.72		
Total	539	1800.81			

Table 3.3 ANOVA table for spatial clustering of groups

Source	Df	*SS*	*MS*	*F*	*p*-value
Dispersal (d)	2	2.56	1.28	1.53	0.22
Number Generations (t)	2	1483.79	741.89	887.64	0
Population (p)	5	123.21	24.64	29.48	0
d t	4	5.29	1.32	1.58	0.18
d p	10	13.44	1.34	1.60	0.10
t p	10	83.67	8.37	10.01	0
d t p	20	15.70	0.78	0.94	0.54
Error	486	406.2	0.83		
Total	539	2133.88			

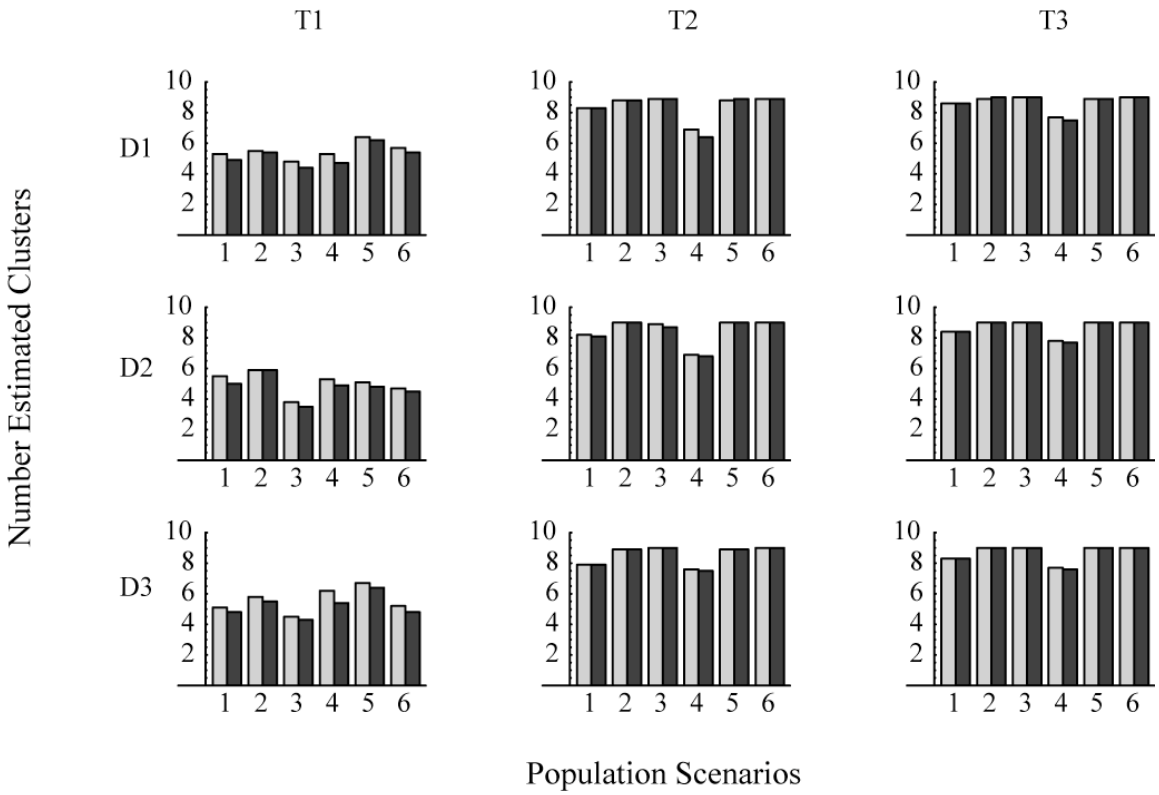

Figure 3.3 Average number of estimated clusters for examined scenarios for both clustering of groups (gray bars) and spatial clustering of groups (black bars). D1, D2, and D3 refer to the short, medium and long median dispersal scenarios, respectively. T1, T2, and T3 refer to 5, 15 and 30 generations since fragmentation, respectively. Scenarios 1 to 3 include subpopulations of equal size, and 4 to 6 include one larger. Scenarios 1 and 4 are the small population, 2 and 5 are the medium and 3 and 6 the large, as in Table 3.1.

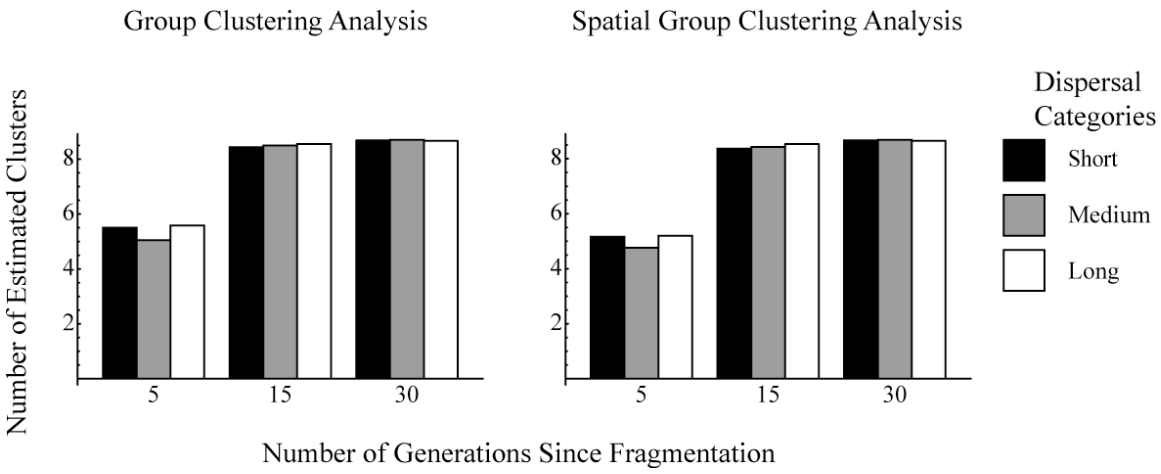

Figure 3.4 Average number of estimated clusters over all population scenarios for each combination of dispersal and generations since fragmentation discriminated by type of analysis.

Relationship between estimating correct number of clusters and F_{ST}

As discussed above, it has been shown that the ability of Bayesian model to estimate the correct number of clusters is intimately associated with the degree of differentiation among subpopulations (e.g. Evanno *et al.*, 2005; Latch *et al.*, 2006). This is because the model itself requires a certain degree of differentiation at allelic frequencies among subpopulations to be able to accurately estimate the number of clusters (Corander *et al.*, 2003). Indeed, our results show that the precision of the model increases with overall genetic differentiation among subpopulations (Figure 3.5). In the examined scenarios, the smallest F_{ST} value that yielded a correct estimate of the number of clusters was 0.048, which is very close to the 0.05 found by Latch *et al*. (2006) as the minimum amount of differentiation to have a 97% confidence that the correct number of clusters has been estimated using BAPS. If we examine the relationship in the different scenarios separately (Figures 3.6, 3.7 and 3.8), it becomes apparent that while increases in F_{ST} lead to increases in power in the estimation of the number of clusters, the relative levels of

differentiation at which BAPS is able to identify the correct number of clusters varies considerably across different scenarios. In general, increases in time since isolation and population size lead to increased power at smaller F_{ST} values than for shorter periods since isolation and smaller populations sizes.

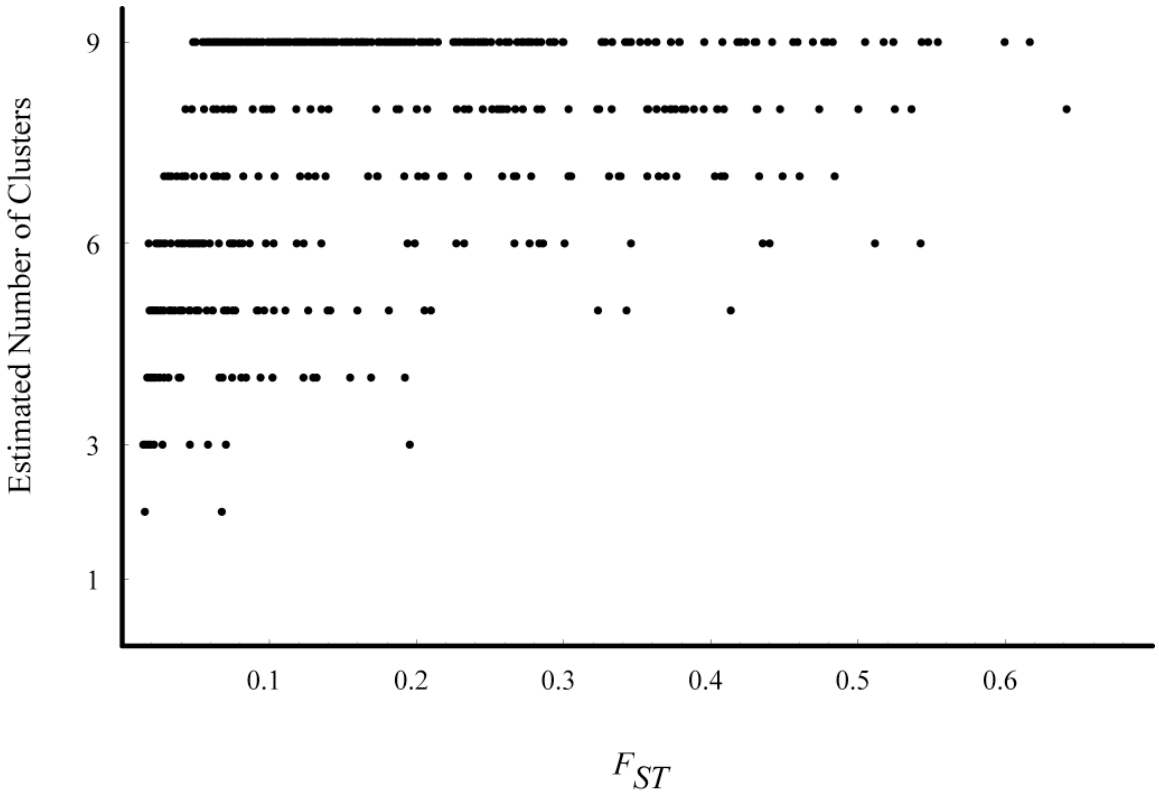

Figure 3.5 Variation in estimated number of clusters in relation to overall F_{ST} for each replicate from each examined scenario, using group clustering analysis. Results of correlation indicate a significantly positive slope (a=0.023, p=0, R^2=0.11).

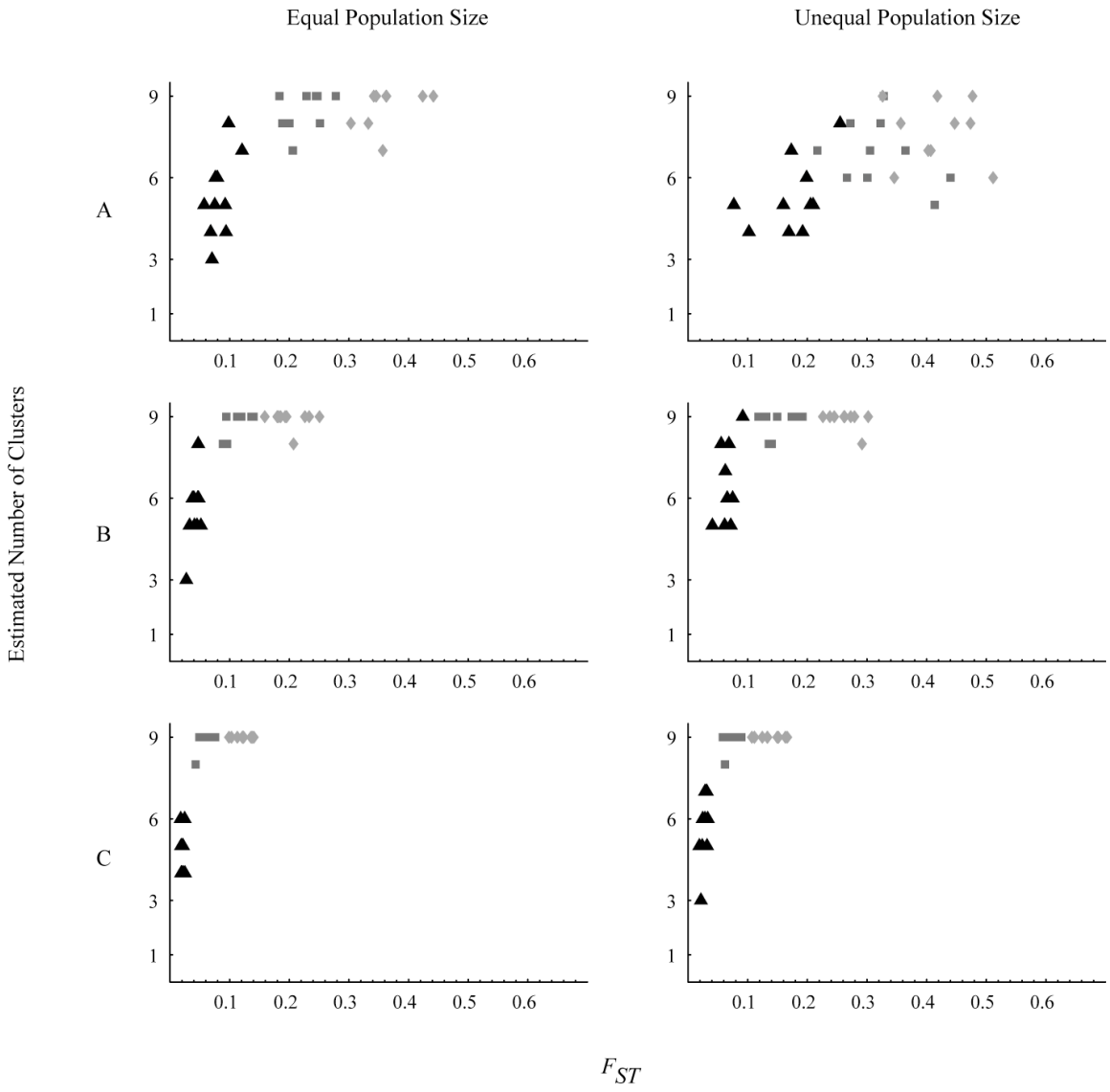

Figure 3.6 Variation in estimated number of clusters in relation to F_{ST} for all scenarios with short median dispersal. A – small population size; B – medium population size; C – large population size. Black triangles – 5 generations since fragmentation; dark gray squares – 15 generations since fragmentation; light gray diamonds – 30 generations since fragmentation.

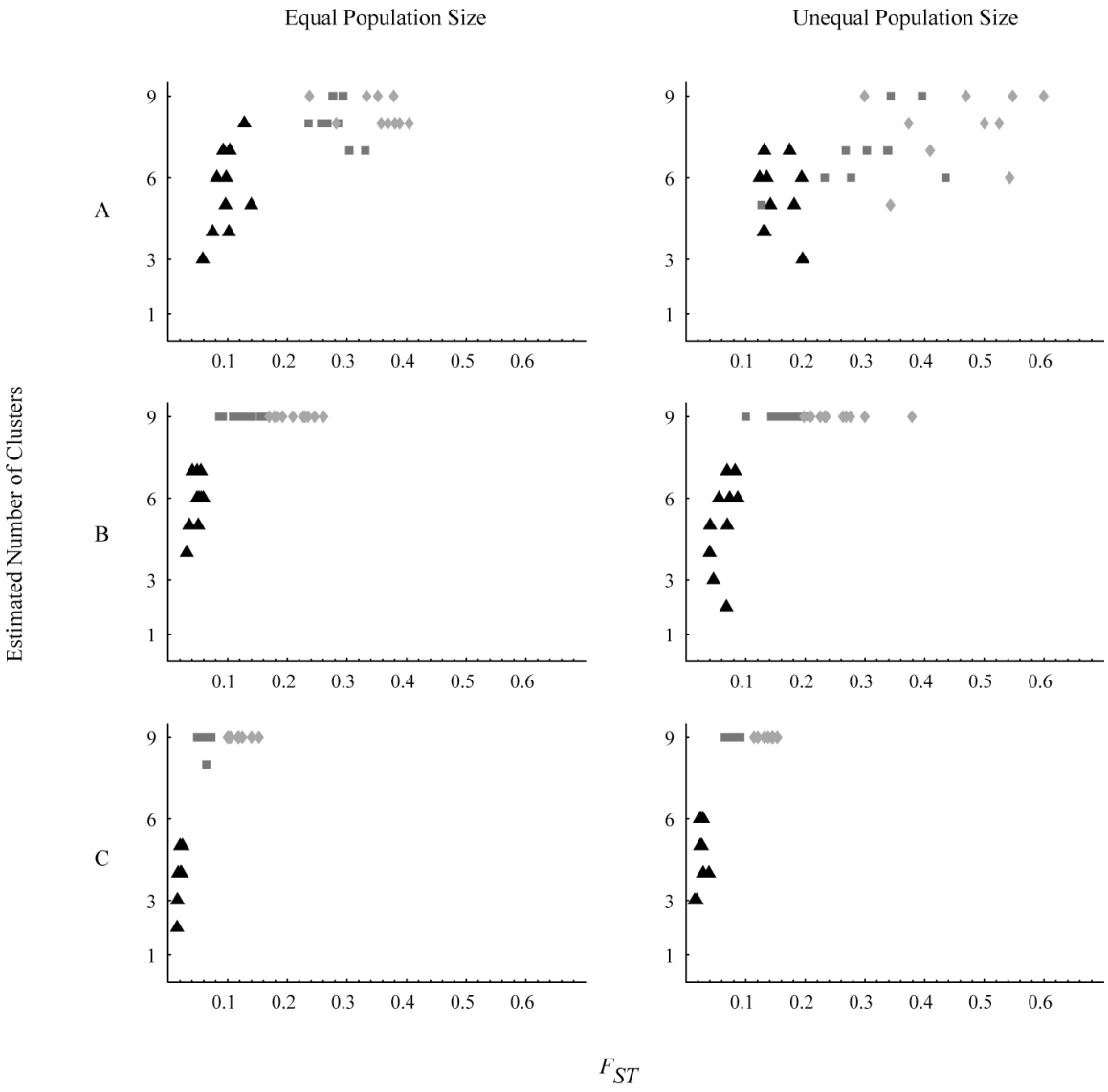

Figure 3.7 Variation in estimated number of clusters in relation to F_{ST} for all scenarios with medium median dispersal. A – small population size; B – medium population size; C – large population size. Black triangles – 5 generations since fragmentation; dark gray squares – 15 generations since fragmentation; light gray diamonds – 30 generations since fragmentation.

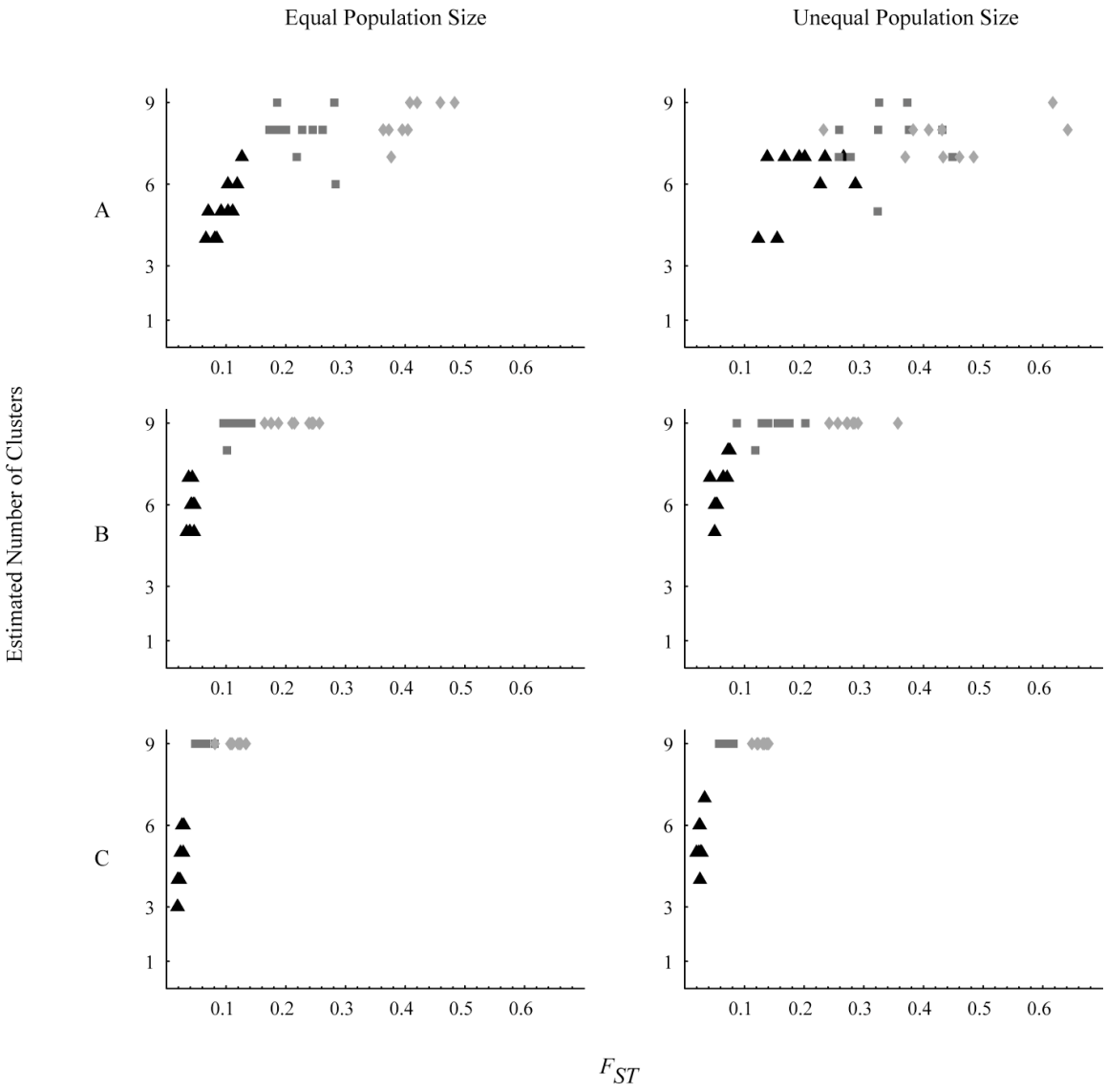

Figure 3.8 Variation in estimated number of clusters in relation to F_{ST} for all scenarios with long median dispersal. A – small population size; B – medium population size; C – large population size. Black triangles – 5 generations since fragmentation; dark gray squares – 15 generations since fragmentation; light gray diamonds – 30 generations since fragmentation.

Conclusions

Here we examined the ability of BAPS 4.14 to infer the correct number of clusters in a hypothetical situation of recent fragmentation. Within this context, we tested the ability of the model to infer the correct number of clusters over several scenarios varying in

dispersal capacity, number of generations since the fragmentation event, and in both population size and in the evenness in the distribution of individuals among the subpopulations. As is summarized in Figure 3.9, we found that the model was robust enough to identify the correct number of clusters under several different situations. Nevertheless, it consistently underestimated the number of clusters in the cases when only a small number of generations had passed since the fragmentation event, independently of the size of the population. But, less biased results were obtained for almost all scenarios when there was an increase in the number of generations since fragmentation. The improvement varied with population size, and how evenly distributed individuals were among subpopulations (Figure 3.9). The model proved to be less powerful in cases with small population sizes, and with extremely uneven distributions of individuals among subpopulations. The fact that median dispersal distance was not an important factor may be attributed to the simplifying assumption that isolation was the same for all species. Future tests should relax this assumption, as it is plausible to think that, within certain limits, species that have the capacity to disperse farther (relative to the size of the fragment) are probably also going to be less affected, in relative terms, by fragmentation than species that disperse shorter distances. In addition, one should perhaps consider a wider range of median dispersal distances. It is possible that the range studied here is too small to produce a significant effect within the conditions of number of individuals, loci and alleles examined.

BAPS never overestimated the number of clusters, but it always inferred the number of clusters to be larger than one. This suggests that in cases where isolation among subpopulations is relatively recent, but there is isolation nonetheless, some degree

of structuring will be found. At neutral loci, the degree of structuring among the true subpopulations will be related to the realized rates of differentiation due to genetic drift. In effect, each subpopulation is a realization of the random process of genetic drift. As such, it is possible that unrealistic clusters, in which subpopulations that are not immediately adjacent on the grid are grouped, appear in these scenarios because they have not yet diverged sufficiently. Indeed, when examining Voronoi tessellations quite often we found clusters that grouped disparate subpopulations. This pattern might be used to infer that isolation is an important factor generating allele frequency divergence among subpopulations, even if we cannot yet know for sure how many clusters really exist.

In comparing group clustering and spatial group clustering, there is a high correlation between the estimated number of clusters using the different methods. This is expected given the high correlation between the identity of the groups (as used in group clustering analysis) and the spatial coordinates (as used in the spatial group analysis). There were, however, differences in the composition of the clusters and, in general, spatial clustering resulted in higher posterior probability associated with the best partition than group clustering. Given that specifying the spatial location of the groups narrows the space of possible results to be searched, this is not surprising.

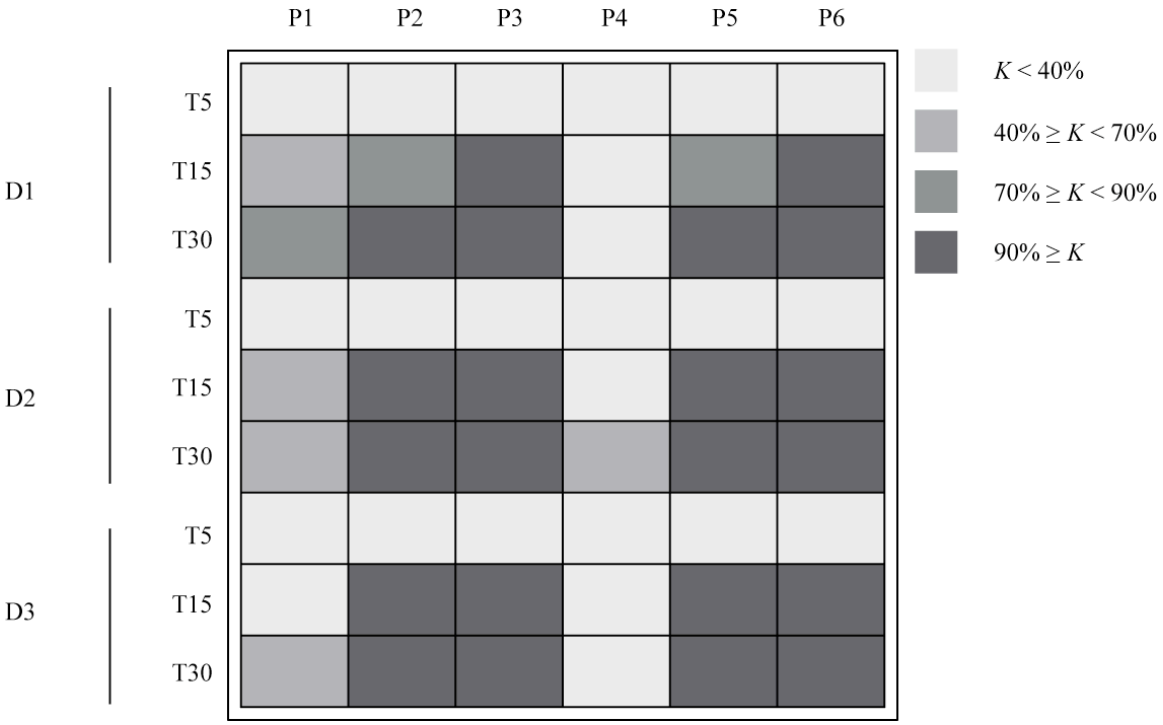

Figure 3.9 Power of BAPS to detect the correct number of clusters (K=9) for each of the tested scenarios. Gray coloring refers to the proportion of replicates (N=10) from which BAPS identified the correct number of clusters in each scenario. D1, D2 and D3 refer to the three dispersal scenarios, short, medium and long, respectively. T5, T15 and T30 refer to number of generations since fragmentation, 5, 15 and 30, respectively. P1-P6 refer to the different population scenarios. P1-P3 are scenarios with even number of individuals for all subpopulations, with total number of individuals increases from P1 to P3. P4-P6 are scenarios in which one population has more individuals than the other eight. Total number of individuals increase from P4-P6.

Finally, while BAPS is a powerful model, there still remains the question of how much the incorporation of the sampling design into the priors override the genetic signal. In the case of BAPS, simulations have shown that, in spite of sampling design, the model will group individuals when there is insufficient evidence to separate them (Corander *et al.*, 2003). With real data, STRUCTURE was unable to identify any structure in cases of low genetic differentiation or incomplete lineage sorting (Coulon *et al.*, 2006; Mank & Avise, 2005). However, the results improved markedly when information on the origin of samples (using STRUCTURE Mank & Avise, 2005) or on the geographic coordinates of

samples (using GENELAND Coulon *et al.*, 2006) were used to modify the priors. In the case on freshwater turtles described by Mank & Avise (2005), in spite of the modification of the prior, animals known to be of the same genetic ancestry were correctly grouped together even though they were located in separate *a priori* groups. In the case of Coulon *et al.* (2006), the modification of the prior grouped individuals of European roe deer from opposite sides of a highway, which is consistent with restricted movement within the species. Therefore, in these two cases, modifying the prior increased the power of the Bayesian model to detect biologically meaningful structure, and, at least in the first case, we have evidence that the modification did not overpower the genetic information. However, future research might examine the effects of informed priors on type II errors (i.e. rate of false clusters).

Bayesian models show great promise by allowing for more efficient use of collected genetic data. In the case of detecting the number of unique genetic units in a population, BAPS is very powerful. And, even in the cases where it was not able to detect the correct number of groups, it was still able to detect that there was some effect of isolation.

Chapter 4. Spatial distribution of genetic variation: Biological correlates of population genetic structure

To be submitted as:

Gonçalves da Silva A, Cullen Jr. L, Medici EP, Nava A, Valladares-Padua C & Melnick DJ **Scale dependent fragmentation effects on terrestrial mammals.**

Abstract

The relationship of body size and diet to the response to landscape structure is well documented in landscape ecology. These features are correlates of dispersal capacity, and thus, landscape connectivity. Here, we use a Bayesian approach to examine the spatial clustering of individuals from different mammal species in the same landscape. Six species were chosen (collared peccary, white-lipped peccary, lowland tapir, ocelot, puma and jaguar). Variation in body size and trophic level allowed us to examine the spatial genetic consequences of these characteristics. We predicted that larger species would respond to landscape features at larger spatial scales than smaller ones, and herbivores would be significantly more affected at smaller spatial scales than carnivores. Our results suggest this to be the case. The collared peccaries displayed fine spatial structure, with clusters falling along the lines of forest fragments. White-lipped peccaries displayed similar levels of clustering, though they also seem affected by finer landscape features (e.g., roads). The much larger tapirs featured less structuring than either peccary, but it may be related, in part, to past landscape conformation. Nevertheless, the results imply an effect of landscape features at a larger scale. Ocelots were less affected than herbivores, but still seem to be affected by fragmentation at the scale of the studied landscape. Jaguars show no structure at the sampled scale. Using samples from nearby areas, indicates a fragmentation effect, but at larger spatial scales than defined by the original study landscape. Body size and trophic level seem to be good predictors of population genetic structure.

Keywords: landscape genetics, landscape ecology, Bayesian models, dispersal, mammals

Introduction

Geographic space is rarely homogeneous. Instead, it is better described as an intricate and diverse matrix of different habitat types of varying complexity (Crist *et al.*, 1992; D'Eon *et al.*, 2002; Wiens *et al.*, 1995). This heterogeneity (*cf.* Li & Reynolds, 1995) is a fundamental aspect of landscapes that affects how effectively distant individuals are from one another, and therefore, the degree to which populations and individuals are functionally connected within a species (McIntyre & Wiens, 1999; With & Crist, 1995). The level of connectivity for any given species will depend on the scale of the landscape heterogeneity relative to the species natural capacity to disperse (Calabrese & Fagan, 2004). Furthermore, the degree to which different species within the same landscape are affected will be determined by differences in their dispersal capacity. Wallace's line, for example, has effectively isolated many groups of terrestrial non-volant species, causing the evolution of two separate biotas in relative close geographical proximity (Brown & Lomolino, 1998). Conversely, the same geographical feature seems to have very little effect on bats, and there is little distinction among the species composition on either side of Wallace's line (Hall *et al.*, 2004).

The degree to which a species is interconnected in a landscape is measured by the functional landscape connectivity (Merriam, 1984; Turner, 1989), which is, as mentioned above, correlated with the species' ability to disperse relative to the spatial scale of landscape heterogeneity. In general, dispersal capacity is seen as a function of the species' natural history traits. Different ecological and biological correlates have been used to predict a species' dispersal capacity in a given landscape, and therefore the level of connectivity. Two commonly used correlates are body size (Crist *et al.*, 1992) and diet

or trophic level (Peterson & Denno, 1998). In mammals, an individual's potential dispersal distance will be highly correlated with body size, explaining 88% of carnivore and 74% of herbivore variation in median dispersal distance (Sutherland *et al.*, 2000); and scaled by trophic level – carnivores' median distance is 1.2 to 4.5 larger than herbivores' of the same size (Sutherland *et al.*, 2000). Body size correlates with longevity and resource search time (Brown *et al.*, 2000), leading to bigger animals generally dispersing over larger distances than smaller animals (Sutherland *et al.*, 2000). Trophic level, on the other hand, correlates with spacing behavior and home range size (Harestad & Bunnel, 1979). Hence, carnivores disperse, on average, relatively farther than herbivores of similar size (Sutherland *et al.*, 2000). Nevertheless, how far individuals will actually disperse will depend on the landscape context in which they are found (Crist *et al.*, 1992; Mayer *et al.*, 2002). In different landscapes, individual dispersal distance may be significantly shorter than expected if, for instance, they suffer higher mortality rates because of lack of food (Crockett & Pope, 1993), higher predation rates (Alberts & Altmann, 1995), or even hostility by conspecifics (Shrader & Owen-Smith, 2002). Also, depending on the scale at which habitats are distributed in a landscape, individuals may not be able to detect heterogeneity (Andreassen *et al.*, 1996; Zollner & Lima, 1997), or may be unwilling to cross the matrix that surrounds suitable habitat (Keyghobadi *et al.*, 1999). Finally, as some critical thresholds in the scale of landscape heterogeneity are met, functional connectivity becomes completely disrupted, and dispersal among habitat patches within a landscape effectively stops (With & Crist, 1995).

Geographic space has long been recognized as an important component in shaping the evolutionary process. Sewall Wright (1943) was the first to formalize the role

of space within species in the concept of isolation-by-distance. Later, Ernst Mayr (1942) postulated that geographical isolation among populations is essential for speciation, and Kimura and Weiss elaborated on a variety of dispersal scenarios and their genetic consequences (Kimura, 1953; Kimura & Weiss, 1964; Weiss & Kimura, 1965). However, only recently have geneticists been able to examine population genetic processes at the landscape ecology scale (Manel *et al.*, 2003). New methods have allowed geneticists to examine finer-scale structuring, by resolving allelic frequency changes at very fine temporal and spatial scales (Manel *et al.*, 2003; Sunnucks, 2000). The basic premise underlying the study of population genetics at this scale is that while the same evolutionary forces apply, the smaller the temporal and spatial scales examined the more likely one is to identify population genetic correlates of fine scale ecological processes. These tools have been used to investigate the effects of behavior on the structuring of genetic variation (Piggott *et al.*, 2006; Rosenbaum *et al.*, 2002); and, to understand the effects of habitat fragmentation on genetic variation (Dallas *et al.*, 2002; Stow *et al.*, 2001; Sumner *et al.*, 2004) in individual species. In all of these studies there is a basic recognition of the role of geographical space in shaping genetic variation, either implicitly by choosing an appropriate sampling scale relative to spatial structure, or explicitly by testing its effect on the structuring genetic variation. However, there are few comparative studies that focus on understanding ecological and biological correlates of population genetic spatial structure in multiple species in a given landscape.

Here, we examine the spatial distribution of genetic variation in mammal species with differing dispersal capacities across the same landscape. By keeping the landscape constant and varying species ability to disperse in the landscape, we wish to gain insight

into how a species' biology interacts with the surrounding landscape to structure genetic variation. Our predictions are that (1) larger species will display less spatial structure than smaller species within the same trophic level at the same scale, and (2) carnivores will display less spatial structure than herbivores at the same scale. To test this hypothesis we selected six mammalian species – the jaguar (*Panthera onca*), the puma (*Puma concolor*) ocelot (*Leopardus pardalis*), lowland tapir (*Tapirus terrestris*), collared peccary (*Tayassu tajacu*), and the white-lipped peccary (*Tayassu pecari*), comprising three orders (Carnivora, Artiodactyla and Perissodactyla), and three size categories in herbivores and two in carnivores, in a single fragmented landscape - the *Pontal do Paranapanema*, in the western tip of the State of São Paulo, Brazil. The choice of these species across this particular landscape enabled us to examine genetic variation among a wide array of dispersal abilities relative to the same landscape features, allowing us to identify when a critical threshold is reached and landscape connectivity is disrupted. In particular, we are testing the effect of the main road bisecting the landscape, and the effect of forest fragmentation throughout the landscape.

Material and Methods

Study site

The main site for this study was the *Pontal do Paranapanema* Region of Brazil (Figure 4.1). More specifically, the site encompasses an area of roughly 270,000ha that was once the *Grande Reserva do Pontal do Paranpanema* (The Great Reserve of the *Pontal do Paranapanema*, Ditt, 2002, and references therein). This region is located at the western tip of the State of São Paulo (Brazil), and is delimited by the Paranapanema River in the

South, the *Paraná* River in the West, the *Anhumas* Brook in the North, and the Dividing Ridge of the *Paranapanema-Paraná* River Basins in the East (Ditt, 2002). The area is part of the Atlantic Forest Complex (Morellato, 2000), and its original vegetation cover is classified as Atlantic Forest of the Interior (Ditt, 2002). In addition, some jaguar samples were obtained from two other surrounding areas: *Anaurilândia*, immediately to the North-West on the other side of *Panará* River; and from the *Parque Estadual das Várzeas do Rio Ivinhema* (*Ivinhema*), which is located about 145km South-West of the the *Pontal*, following the *Paraná* River.

Historically, the region was first divided into three large fragments around 1950, the protected *Parque Estadual Morro do Diabo* (PEMD) that has changed little in size since, and two others that have since been fragmented into smaller fragments (Uezu, 2007), the *Ponte Branca* and *Tucano* complex and the *Santa Mônica* and *Santa Maria* complex (Figure 4.1).

Field collection and laboratory analysis

Due to the nature of the species being studied (i.e. elusive and occurring at low densities), most of the sampling for this study was based on noninvasively collected dung samples. In addition, researchers from the Institute for Ecological Research (IPÊ) kindly provided blood samples from all five species.

Dung samples were collected opportunistically on roads surrounding and on and off trails within local forest fragments. Sampling was carried out in the mornings, when temperatures were cooler. When collecting dung samples, approximately 2ml of dung was scraped off the surface of the bolus using a dry twig (one per sample) and put into a

5ml sampling tube containing ~4ml of RNAlater (Ambion, Inc.). The tube was subsequently capped, sealed with parafilm, well mixed, and kept at room temperature until reaching the lab, where the samples were kept at -20 ^{o}C until DNA extraction. Blood samples were preserved in equal volume of Easy Blood buffer (Tris HCl, EDTA and SDS), and kept at -20 ºC until extraction.

DNA extraction followed the CTAB based method described in Ferreira and Grattapaglia (1998) with modifications. First, preceding the extraction, 100mg of wet sample was separated and washed with 500µl of PBS, and after the first chlorophorm:isoamyl extract, DNA was further cleaned using the QIAquick Gel Extraction Kit (Qiagen, Inc.) by adding 3 volumes of QB buffer and proceeding as instructed by the manufacturer. All extractions were performed with at least two negative extraction controls, one at the beginning of the series and one at the end. To avoid issues of contamination and genotyping errors extraction and PCR reaction safety protocols were followed as recommended by Fernando *et al.* (2003a) and Taberlet *et al.* (1999). Isolation of DNA from blood samples was carried out using the DNeasy Tissue Kit (Qiagen, Inc.) following manufacturer's protocol. Feline dung samples were identified to species level by following the protocol delineated by Farrell *et al.* (2000) using reference sequences amplified from blood samples. Herbivore dung was identified based on their unique morphology.

To quantify neutral genetic variability for each species, we used nuclear DNA microsatellite markers. Mitochondrial sequences were not chosen because it is highly unlikely that they would have been informative at this scale for these species (e.g. Eizirik *et al.*, 2001). A battery of species-specific and cross-specific primers was used (Table

4.1). For blood samples, PCR was performed in 6µl reactions with 10-15ng of DNA, 0.3µM of each primer, 3µl of QIAGEN PCR Multiplex Master Mix (Qiagen, Inc.). Fecal samples were processed under similar conditions, except 3µl of DNA extract was used. PCRs were carried out over 35 cycles, and 45 cycles for blood and dung samples, respectively. PCR reactions from DNA extracted from dung were repeated 2-7 times depending on the quality of the DNA, and the repeatability of the genotype. Consensus genotypes were built using the method described in Frantz *et al.* (2003). If after 7 repetitions, we could not obtain genotypes for at least five loci the sample was removed from the study. PCR reactions for blood samples were performed once, with additional reactions in case of failure. In addition, to measure allelic dropout and false alleles rates, a small portion of the blood samples were chosen for additional PCR reactions. Allelic dropout and false allele rates were calculated over all dung samples producing usable genotypes. PCR for sequencing reactions followed Farrell *et al.* (2000). Cycle sequencing reactions were carried out directly from PCR amplification products using the ABI PRISM® BigDye™ Terminators v3.1 Cycle Sequencing Kit 3.1 (Applied Biosystems) preceded by fragment cleanup using the QIAquick Gel Extraction Kit (Qiagen, Inc.). Unincorporated BigDye™ dyes were removed from the cycle sequence reactions using the EDTA/Ethanol protocol suggested by the manufacturer (AppliedBiosystem). Each strand was sequenced three times, totalizing six sequencing reactions per sample.

Finally, all DNA extractions and amplification products were checked on standard agarose gels with Ethidium Bromide staining. DNA from blood samples was quantified by comparison to High DNA Mass Ladder (Invitrogen Corporation). Genotypes were obtained on ABI PRISM® 3100 Automated DNA Sequencing machine and processed

using GeneScan 3.7 (Applied Biosystems) and Genetyper v3.7 (Applied Biosystems). Sequences were obtained on an ABI PRISM® 3700 Automated DNA Sequencing machine and processed with Sequencher v4.5 (Gene Codes Corporation). All lab work was carried out at the Laboratory of Plant Genetics at the Brazilian Agricultural Corporation – Division of Genetic Resources and Biotechnology (Brasília, Brazil).

Table 4.1 Annealing temperature and size range for each primer pair. Collared and white-lipped peccary primers were originally developed for domestic pigs (*Sus scrofa*). Lowland tapir primer Tte12 was developed using lowland tapir libraries, while Tba20 and Tba23 were developed using Baird's tapir (*Tapirus bardii*) libraries, HMS7, HTG4 and HTG7 were originally developed for domestic horses (*Equus cabalus*). Primers used for feline species were originally developed for domestic cat (*Felis catus*).

	Annealing temperature (ºC)	Size range	Source
Collared peccary			
S0225	55	173-181	Archibald *et al.* (1995)
SW444	55	100-108	
S0226	55	174-184	
SW957	55	119-129	
IGFI	55	127-143	
SW857	55	128-154	
White-lipped peccary			
S0225	55	181-201	Archibald *et al.* (1995)
SW444	55	94-102	
S0226	55	180-194	
SW957	55	122-148	
IGFI	55	131-137	
SW857	55	130-134	
Lowland tapir			
Tte12	63	166-174	Norton & Ashley (2004)
Tba20	63	244-260	
Tba23	63	222-230	
HMS7	60	156-164	Bowling *et al.* (1997)
HTG4	60	128-136	
HTG7	60	113-129	
Ocelot			
FCA441	55	123-133	Menotti-Raymond *et al.* (1999)
FCA094	55	199-215	

FCA201	55	130-136	
FCA290	55	212-220	
FCA105	55	174-190	
FCA211	55	99-115	
FCA347	55	165-173	
Puma			
FCA441	55	141-153	Menotti-Raymond *et al.* (1999)
FCA094	55	184-188	
FCA201	55	130-136	
FCA290	55	218-228	
FCA105	55	174-180	
FCA211	55	107-119	
FCA347	55	161-173	
Jaguar			
FCA441	55	135-151	Menotti-Raymond *et al.* (1999)
FCA094	55	213-217	
FCA201	55	132	
FCA290	55	214-220	
FCA105	55	197-217	
FCA211	55	113-125	
FCA347	55	167-175	

Population genetics descriptive statistics for whole samples

For each species, we quantified average number of alleles per locus, observed and expected overall heterozygosity (corrected for sample size), Weir and Cockerham's (1984) *F*-statistics, and tested for departures from linkage equilibrium and Hardy-Weinberg Equilibrium (HWE) using GENEPOP (Raymond & Rousset, 1995). Significance levels were corrected for multiple tests using the standard Bonferroni technique (Rice, 1989).

Inferring spatial structure of genetic variation

Because we are interested in understanding how animals interact with their surrounding landscape it is essential that structure be inferred from the data. As such, we chose two

clustering methods, based on Bayesian models, to infer the spatial structure in genetic variation: BAPS (Corander *et al.*, 2003) and STRUCTURE (Falush *et al.*, 2003; Pritchard *et al.*, 2000). We chose these models because of their versatility and proven use in uncovering cryptic spatial structure (Evanno *et al.*, 2005; Latch *et al.*, 2006). A discussion on the differences between the methods has been published elsewhere (Latch *et al.*, 2006). However, because these models are still relatively new, using both has also provided us with an opportunity to test their utility under a wider set of conditions.

BAPS

BAPS treats both subpopulation (cluster) allele frequencies and the number of genetically distinct clusters (*K*) as random variables in a Bayesian model (Corander *et al.*, 2003). The number of clusters is treated as a variable to be estimated, and groups of individuals based on sampling design are randomly fused and split in an attempt to find the best number of clusters given the data (mixture procedure). In its most basic form, it treats all combinations to be equally likely *a priori* (i.e. a uniform prior), and will analytically derive the posterior distribution of allele frequencies and substructure for a relatively small number of groups. It is also possible to use information, such the spatial location of individuals or behavioral characteristics, to inform the prior and improve the estimates of the number of clusters, particularly in cases with sparse data (Corander *et al.*, in press). Here, we clustered both individuals and groups using the non-spatial and spatial mixture options available within the program. For group clustering, samples for each species were grouped according to different criteria (see results). Program parameters and sequence of analysis were set as suggested by Corander *et al*. (2006). We calculated the posterior

probability for the best and second best partitions using the formula described in Corander *et al.* (2006).

STRUCTURE

STRUCTURE uses a Bayesian model to jointly estimate population allele frequencies and the number of populations (*K*) given the sampled genotypes. In essence, the model introduces population structure, and attempts to find groupings of individuals that are in Hardy-Weinberg and linkage equilibriums (Pritchard *et al.*, 2000). Nevertheless, the estimation of *K* is *ad hoc*, in that in each run of the model, *K* must be fixed (Pritchard *et al.*, 2000). To identify the most likely *K*, we compared the mean posterior likelihood of the data across different values of *K* (Pritchard *et al.*, 2000), and compared the mean rate of change in the log posterior likelihood of the data across different *K* (*deltaK* method, Evanno *et al.*, 2005).

Population genetic statistics of inferred populations

Descriptive statistics for inferred populations were obtained with GENEPOP (Raymond & Rousset, 1995), and tests of the significance of population differentiation based on the clusters found were performed with AMOVA using ARLEQUIN (Schneider *et al.*, 2000). Significance levels for multiple tests were corrected using the standard Bonferroni technique (Rice, 1989).

Results

Samples

In total, 222 samples were collected or provided to this study (Table 4.2), of which 94 samples were dung and 128 were blood. Reliable peccary dung samples were not found. Feline and tapir dung were collected whenever they were found, unless it appeared to be too old or had been destroyed by rain or dung beetles. Most samples were probably no more than a week old. Of these, all blood samples produced usable DNA, however only 15 dung samples produced DNA in sufficient quantity and quality to allow for genotyping of at least five loci (Table 4.2). This resulted in a total of 143 samples spread out in various forest fragments of the region (Figure 4.1). Of the 38 feline samples, seven produced DNA sequences of sufficient quality for species identification. Sequences were compared with Genbank nucleotide (nt) database using the BLAST tool. Four of the sequences were identified as *Canis lupus*, and are probably the results of some contamination by a canid (possible prey), and the samples were removed from the study. Of the remaining three sequences, two were closely matched to ocelots (score= 194; $E=7e^{-47}$) and the third was matched to puma (score=301; $E=4e^{-99}$). Subsequent comparison using maximum parsimony to reference sequences produced from blood samples of the region corroborated the BLAST results (Figure 4.2). However, only one of the ocelot samples produced adequate genotypes. In addition, one dung sample, identified as an ocelot by its tracks, was added to the final genotypic database, even though it did not produce satisfactory identifying sequences. Finally, because only three samples produced useable genotypes, pumas were excluded from further analyses, leaving a total of five species.

Table 4.2 List of samples used in this study. In addition, there were 32 unidentified or contaminated feline samples (see text) that were excluded.

Species	Blood	Dung (Total collected)	Total
Collared Peccary	33	-	33
White-lipped Peccary	52	-	52
Lowland Tapir	17	15 (58)	32 (77)
Ocelot	10	2 (3)	12 (13)
Puma	3	0 (1)	3 (4)
Jaguar	11	-	11
Total	**126**	**17 (62)**	**143 (190)**

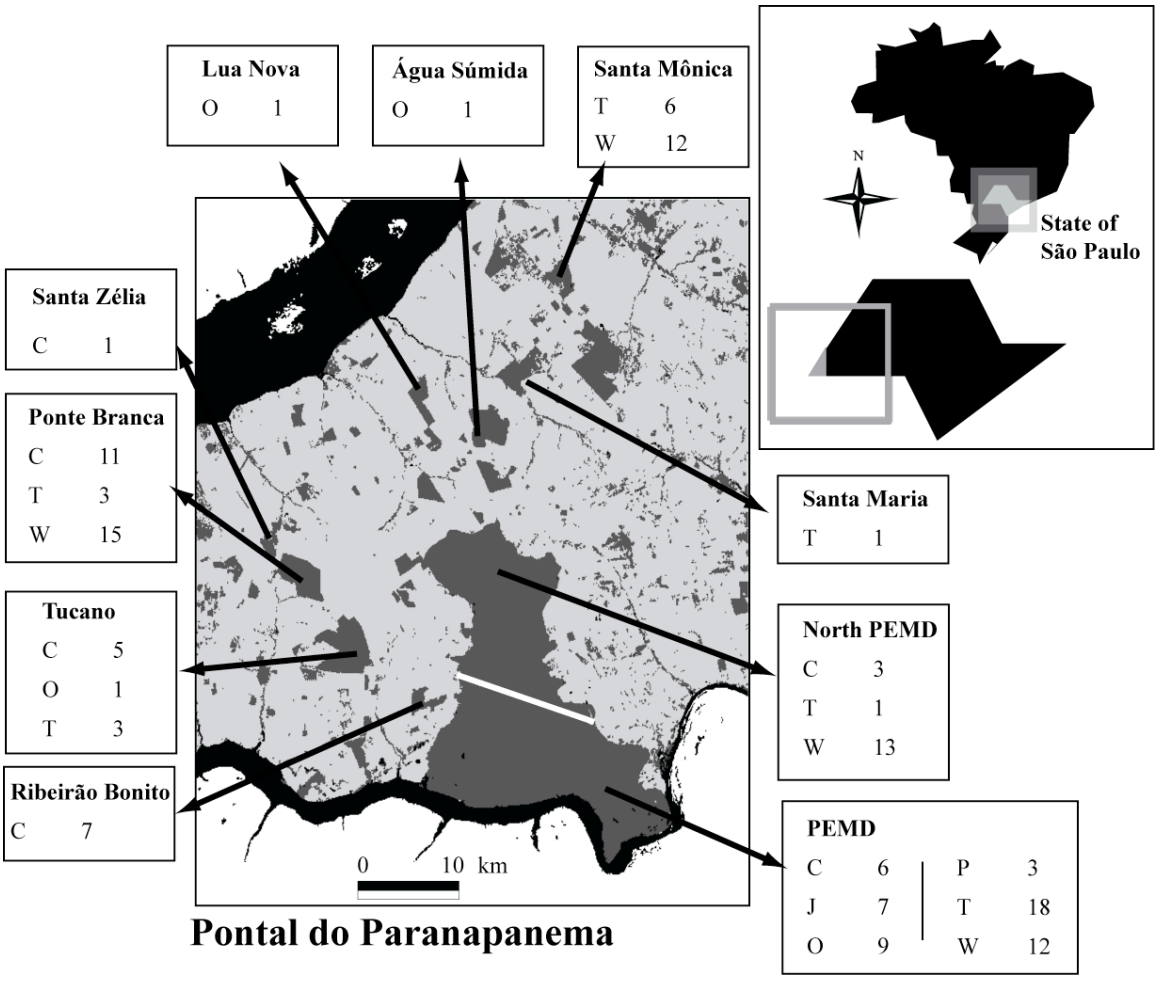

Figure 4.1 Location of study site (inset above right), sampling locations and number of samples per location within the *Pontal*. The map does not include jaguar samples from *Ivinhema* (*N*=4) or from *Anaurilândia* (*N*=1). Forest fragment names are in bold. Solid white line outlines main road crossing the PEMD, dividing it into two main areas, referred to here as North PEMD and PEMD. The first letter of each species is used.

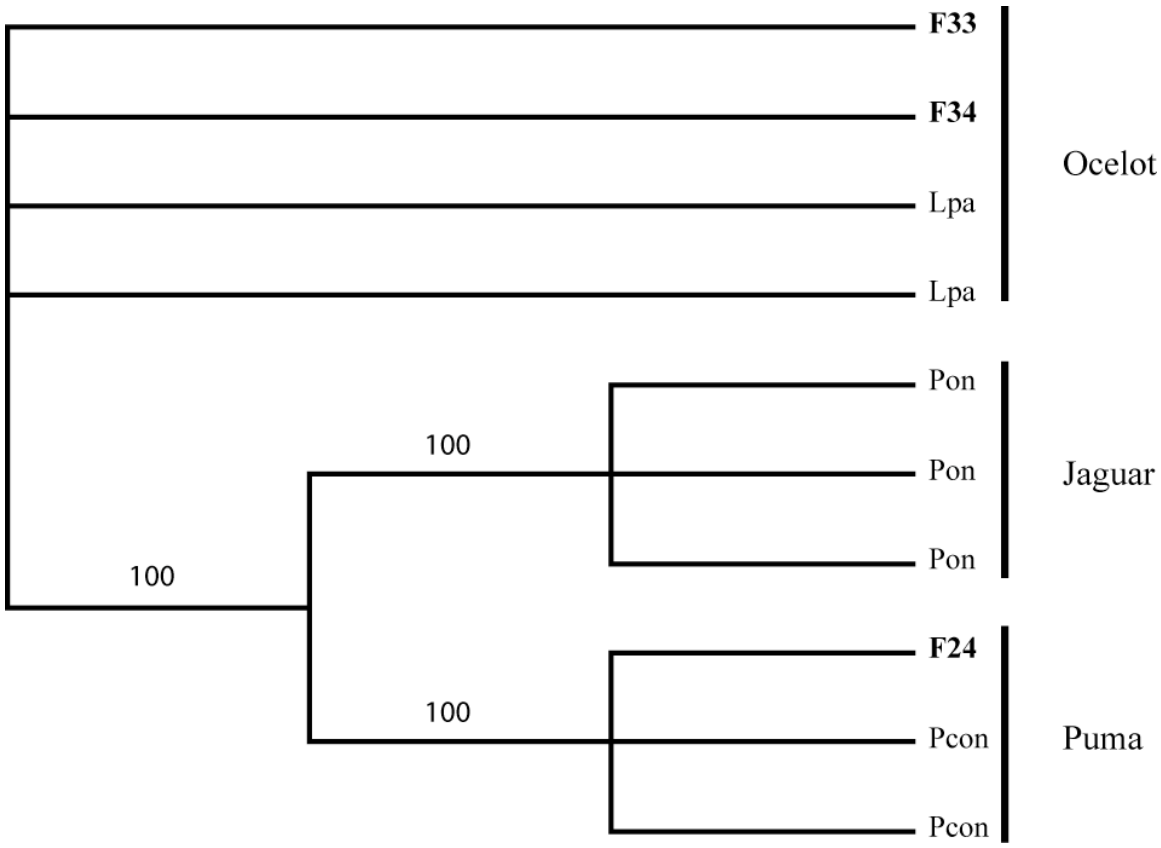

Figure 4.2 Maximum parsimony tree clustering dung samples (bold labels) with reference sequences. Only one tree was found after a heuristic search using PAUP 4.0b10 (Swofford, 2002). Values of bootstrap above ≥50% after 1000 iterations are shown above relevant branches.

Genotyping errors

A total of 9700 PCR reactions were run on the automated sequencer, between tests and genotyping, to produce a total of 143 genotypes at 6-7 loci per species. For each species and sample type we calculated allelic dropout (ADO) and false allele (FA) rates as described in Broquet & Petit (2004, Tables 4.3 and 4.4). No studies, to our knowledge, have published ADO or FA rates for blood samples, so it is hard to gauge what would be acceptable rates. In the case of dung samples, published ADO and FA rates calculated using the same method used here varied between 0-41% and 0-15.3%, with an average

10.42% and 1.92%, respectively (Broquet & Petit, 2004). In the case of blood samples, estimates may be inflated because the choice of samples was skewed towards problematic samples in order to obtain an upper bound for both ADO and FA rates.

Table 4.3 Summary statistics for ADO and FA rates in blood samples

Species	PCRs	Total Number Heterozygotes	PCR Success Rate	ADO	FA
Collared Peccary	118	83	0.96	0.09	0.03
White-lipped Peccary	154	104	0.91	0.05	0.03
Lowland Tapir	244	124	0.93	0.07	0.00
Ocelot	46	28	0.98	0.07	0.02
Jaguar	48	22	1.00	0.05	0.04

Table 4.4 Summary statistics for ADO and FA rates in dung samples

Species	PCRs	Total Heterozygotes	PCR Success Rate	ADO	FA
Lowland Tapir	530	331	0.55	0.21	0.03
Ocelot	28	13	0.93	0.09	0.05

Population genetics summary statistics for whole samples

Collared peccaries

In total, 33 samples of collared peccaries were analyzed over six loci. All six loci were found to be polymorphic, with an average 6.67 alleles per locus (Table 4.5). There were no significant departures from linkage equilibrium. Tests for departures from HWE were significant for one locus after correction for multiple tests (Table 4.5).

Table 4.5 Summary statistics over all samples of collared peccary (*N*=33). He, expected heterozygocity under HWE corrected for sampling bias; Ho, observed heterozygosity; *f*, Weir & Cockerham's (1984) inbreeding coefficient.

Locus	No. Alleles	He	Ho	*f*	*p*-Value	% Complete Genotypes
S0225	3	0.65	0.45	0.307	0.009	1.00

SW444	4	0.58	0.24	0.586*	0	1.00
S0226	3	0.35	0.36	-0.053	0.77	1.00
SW957	8	0.84	0.81	0.038	0.03	0.97
IGFI	9	0.81	0.70	0.146	0.02	1.00
SW857	13	0.88	0.97	-0.105	0.3	0.91
All loci	6.67	0.69	0.59	0.15*	0	0.98

* Denotes significant *f* values after standard Bonferroni correction for multiple tests

White-lipped peccaries

Samples of 52 white-lipped peccaries were analyzed over six loci, with an average 5.5 alleles per locus (Table 4.6). There was no evidence of linkage disequilibrium or significant departures from HWE.

Table 4.6 Summary statistics over all samples of white-lipped peccaries (*N*=52). He, expected heterozygocity under HWE corrected for sampling bias; Ho, observed heterozygocity; *f*, Weir & Cockerham's (1984) inbreeding coefficient.

Locus	No. Alleles	He	Ho	*f*	*p*-Value	% Complete Genotypes
S0225	6	0.67	0.62	0.083	0.16	1.00
SW444	4	0.51	0.57	-0.114	0.82	0.98
S0226	7	0.80	0.81	-0.006	0.42	1.00
SW957	10	0.77	0.86	-0.113	0.9	0.94
IGFI	4	0.34	0.33	0.032	0.53	0.98
SW857	2	0.26	0.23	0.123	0.33	1.00
All loci	5.5	0.56	0.57	0.00	0.649	0.98

* Denotes significant *f* values after standard Bonferroni correction for multiple tests

Lowland tapir

A total of 32 tapir samples were genotyped at six loci. All six loci were polymorphic with an average 7.0 alleles per locus (Table 4.7). There was no evidence for linkage disequilibrium. One locus showed significant departure from HWE.

Table 4.7 Summary statistics over all samples of lowland tapirs (*N*=32). He, expected heterozygocity under HWE corrected for sampling bias; Ho, observed heterozygosity; *f*, Weir & Cockerham's (1984) inbreeding coefficient.

Locus	No. Alleles	He	Ho	*f*	*p*-Value	% Complete Genotypes
Tte12	7	0.73	0.75	-0.021	0.52	1.00
Tba20	10	0.88	0.77	0.124	0.044	0.81
Tba23	6	0.77	0.70	0.089	0.1	0.84
HMS7	5	0.70	0.74	-0.054	0.7	0.97
HTG4	5	0.63	0.42	0.336	0.02	0.97
HTG7	9	0.85	0.52	0.394*	0	0.91
All loci	7	0.76	0.65	0.145*	0	0.92

* Denotes significant *f* values after standard Bonferroni correction for multiple tests

Ocelot

A total of 12 ocelot samples were genotyped at seven loci. All seven loci were found to be polymorphic with an average 5.43 alleles per locus (Table 4.8). All but one locus was in HWE.

Table 4.8 Summary statistics over all samples of ocelots (*N*=12). He, expected heterozygosity under HWE corrected for sampling bias; Ho, observed heterozygocity; *f*, Weir & Cockerham's (1984) inbreeding coefficient.

Locus	No. Alleles	He	Ho	*f*	*p*-Value	% Complete Genotypes
FCA441	3	0.58	0.33	0.440	0.120	1.00
FCA094	9	0.89	1.00	-0.125	1.000	0.83
FCA201	2	0.16	0.17	-0.048	1.000	1.00
FCA290	6	0.59	0.33	0.450*	0.005	1.00
FCA105	7	0.87	0.75	0.147	0.200	1.00
FCA211	6	0.66	0.58	0.120	0.350	1.00
FCA347	5	0.74	0.55	0.277	0.050	0.92
All loci	5.43	0.64	0.53	0.180*	0.008	0.96

* Denotes significant *f* values after standard Bonferroni correction for multiple tests

Jaguar

A total of 12 jaguar samples were genotyped at seven loci (Table 4.9). Locus FCA201 was found to be monomorphic, and was excluded from subsequent analyses. An average

4.0 alleles per locus were found. There was no evidence for linkage disequilibrium or for departures from HWE.

Table 4.9 Summary statistics over all samples (*N*=12) and just for the PEMD (*N*=7) of jaguars. He, expected heterozygosity under HWE corrected for sampling bias; Ho, observed heterozygocity; *f*, Weir & Cockerham's (1984) inbreeding coefficient.

Locus	No. Alleles	He	Ho	*f*	*p*-Value	% Complete Genotypes
Whole						
FCA441	4	0.48	0.58	-0.222	1.00	1.00
FCA094	3	0.69	0.67	0.038	0.33	1.00
FCA290	2	0.46	0.50	-0.082	0.83	1.00
FCA105	5	0.75	0.83	-0.111	0.71	1.00
FCA211	5	0.74	0.55	0.273	0.02	0.92
FCA347	5	0.74	0.67	0.107	0.08	1.00
All loci	4	0.65	0.63	0.000	0.49	0.99
PEMD						
FCA441	2	0.36	0.43	-0.200	1.00	1.00
FCA094	3	0.60	0.71	-0.200	0.85	1.00
FCA290	2	0.53	0.57	-0.091	0.85	1.00
FCA105	4	0.65	0.86	-0.358	1.00	1.00
FCA211	4	0.65	0.57	0.130	0.28	1.00
FCA347	3	0.56	0.71	-0.300	1.00	1.00
All loci	3	0.56	0.64	-0.17	0.83	1.00

* Denotes significant *f* values after standard Bonferroni correction for multiple tests

Inferred structure

Preliminary explorations in STRUCTURE suggested high admixture for all examined species. In particular, using the method outlined by Pritchard *et al.* (2000) to infer *K*, individual membership was assigned to each population at roughly 1/*K* (i.e., *Q*=1/*K* for each cluster), for each *K* tested, and there was high fluctuation in alpha (i.e., admixture parameter) within runs. Pritchard & Wen (2004) indicate that these results are suggestive of lack of structure. However, similar results were observed by Coulon *et al.* (2006), who subsequently found significant structure using a spatially explicit model. Similar results

were obtained with the method developed by Evanno *et al.* (2005). The results below therefore pertain to BAPS.

For both spatial and non-spatial group cluster analyses, we grouped samples according to *a priori* information about the species. We grouped both species of peccaries in accordance with capture sites, which correspond to different herds, with a total of six groups for each species. Lowland tapir samples were grouped following the history of fragmentation of the area. At the onset of fragmentation in the 1950's, three large fragments were created, what is today the PEMD plus the *6RR* fragment, and two larger ones which have since dwindled to what is today better represented by *Ponte Branca* and *Tucano* to the west of PEMD and *Santa Maria* and *Santa Mônica* to the north of PEMD (Uezu, 2007). The ocelots were divided into three groups based on radio-telemetry data from Jacob (2002), one with individuals on the west side of the PEMD, another on the east side and a third outside the PEMD to the north. Finally, radio-telemetry data for jaguars in the region (Cullen, Jr., personal communication) suggests three groups, one in PEMD, another in *Anaurilândia* and a third in *Ivinhema*. We subsequently used BAPS to infer the best partition of these *a priori* groups, assigning a range of initially possible *K* from 1 to 20. The partition with the minimum log marginal likelihood (log(ml)) was considered to be the best partition of groups.

The best partition found five clusters for collared peccaries (Figure 4.3). This partition clustered the groups sampled within the PEMD (PEMD and North PEMD groups), and kept all other groups in separate clusters. The second best partition found four clusters, had a relatively high posterior probability, and was distinct from the first by clustering together the groups *Ribeirão Bonito* and *Tucano* (Figure 4.4). In the white-

lipped peccaries, the best partition found four clusters (Figure 4.3). Two sets of groups were clustered together, the groups found below the road (PEMD1 and PEMD2), and the groups *Ponte Branca* and North PEMD 2. All other groups formed separate clusters. The second best partition was less likely than the first partition, but clustered the groups above the road within the PEMD (North PEMD1 and 2), leaving the *Ponte Branca* group in a cluster of its own (Figure 4.4). Each lowland tapir group formed a cluster of its own with high posterior probability, for a total of three clusters (Figure 4.3). The second best partition grouped the *Ponte Branca*/*Tucano* fragments with the *Santa Maria*/*Santa Mônica* fragments (Figure 4.4). The best partition for the ocelots clustered the two PEMD groups (East PEMD and West PEMD), and kept the North group in a separate cluster (Figure 4.3). The second best partition had a very low relative posterior probability, and clustered all three groups into one population (Figure 4.4). Finally, within the *Pontal* the samples for jaguars were constrained within the PEMD. Spatial clustering of individuals just from the PEMD produced one cluster with all individuals (Figure 4.3). Spatial group clustering of all jaguar samples partitioned the samples into three clusters, separating PEMD, *Anaurilândia* and *Ivinhema* (Figure 4.4). The second best partition joined *Anaurilândia* and *Ivinhema* (Figure 4.4).

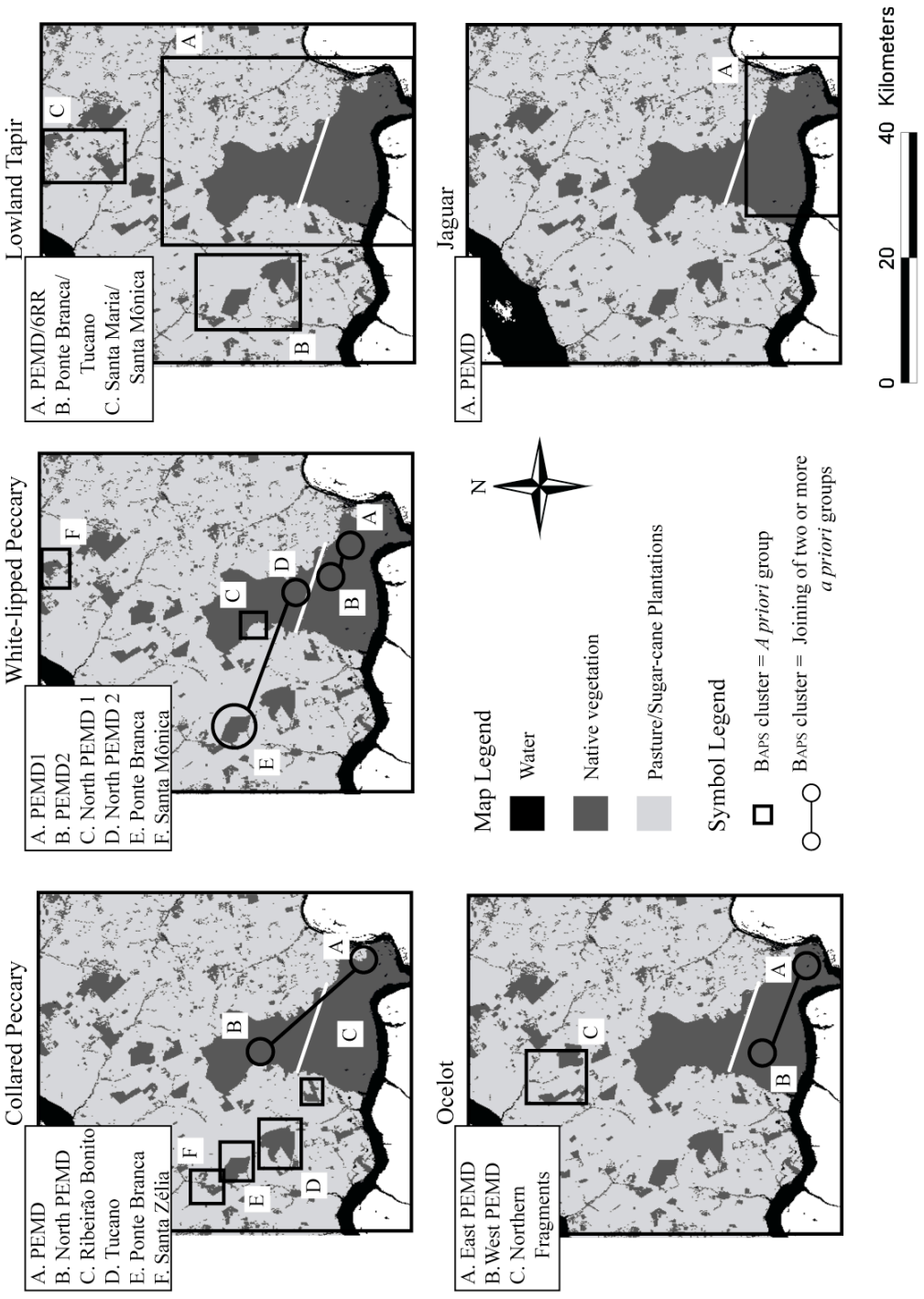

Figure 4.3 Map of groups used in BAPS analysis, and posterior clustering based on the best partition found. Black squares indicate groups that did not cluster with any other group, joined black circles indicate groups that were clustered together. Solid white line indicates the location of the main road crossing the PEMD, separating it in North PEMD and PEMD.

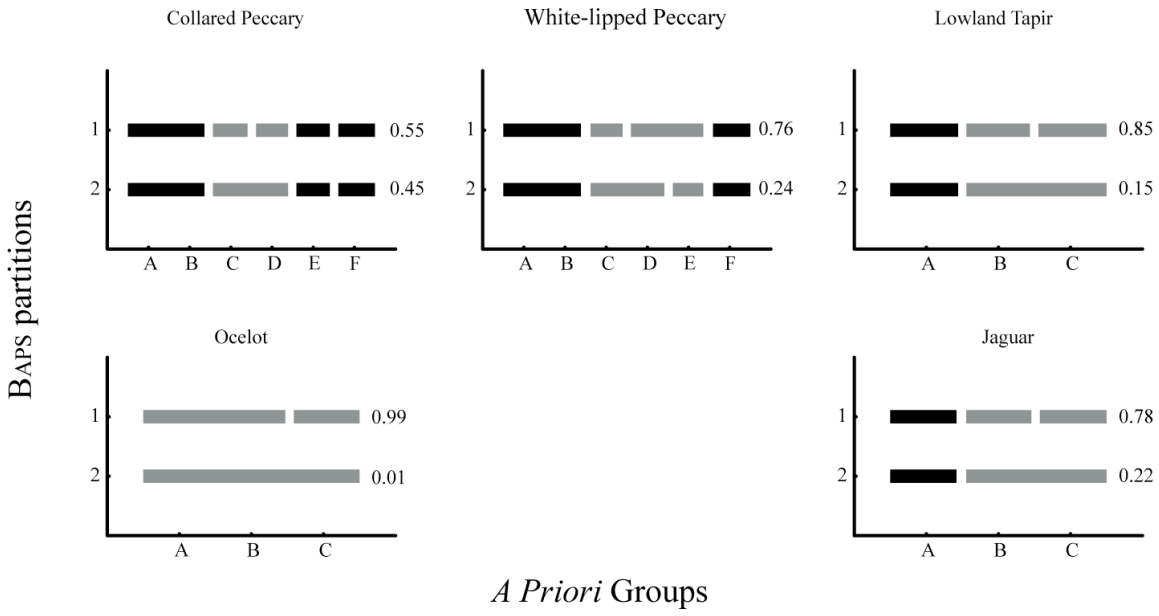

Figure 4.4 Comparison between the best and second best partition of *a priori* groups of individuals found using BAPS for each of the species examined (see text for details). Differences between partitions are highlighted in gray. Species groups follow nomenclature in Figure 4.3, with the exception of jaguar, where A is PEMD; B, *Anaurilândia*; and, C, *Ivinhema*.

Population genetic statistics for inferred populations

Here, we analyzed only the best BAPS partition. Inferred populations for collared and white-lipped peccaries were in HWE (Tables 4.10 and 4.11, respectively). Lowland tapir PEMD populations had two loci with significant heterozygote deficiency (Table 4.12). In ocelots, the PEMD populations displayed significant heterozygote deficiency overall, but not for any locus in particular (Table 4.13). Finally, jaguar populations were in HWE (Table 4.14). There was no evidence for linkage disequilibrium in any of the inferred populations. Tests of the significance of the inferred clustering with AMOVA found significant differentiation for the two species of peccary and tapir populations, but not for ocelot (Table 4.15). AMOVA was significant between the *Ivinhema* and PEMD jaguar populations (Table 4.15).

Table 4.10 Summary statistics for collared peccary populations from the best BAPS partition

Locus	No. Alleles	He	Ho	*f*	*p*-Value	% Complete Genotypes
North PEMD/PEMD (*N*=9)						
S0225	2	0.29	0.33	-0.140	1.000	1.00
SW444	2	0.47	0.22	0.540	0.160	1.00
S0226	2	0.29	0.33	-0.140	1.000	1.00
SW957	8	0.80	0.78	0.034	0.420	1.00
IGFI	7	0.66	0.44	0.340	0.120	1.00
SW857	12	0.92	0.89	0.037	0.230	1.00
All loci	5.5	0.57	0.50	0.112	0.090	1.00
Ribeirão Bonito (*N*=7)						
S0225	2	0.49	0.71	-0.500	1.000	1.00
SW444	2	0.36	0.14	0.620	0.230	1.00
S0226	2	0.14	0.14	0.000	-	1.00
SW957	5	0.82	1.00	-0.230	1.000	1.00
IGFI	5	0.82	0.71	0.140	0.080	1.00
SW857	8	0.94	1.00	-0.071	1.000	0.86
All loci	4	0.60	0.62	-0.007	0.400	0.98
Tucano (*N*=5)						
S0225	3	0.26	0.30	-0.200	1.000	1.00
SW444	2	0.27	0.20	0.270	0.620	1.00
S0226	1	-	-	-	-	-
SW957	5	0.02	0.38	0.180	0.360	0.80
IGFI	3	-0.01	0.40	-0.100	0.770	1.00
SW857	3	-0.08	0.50	-0.600	1.000	0.80
All loci	2.83	0.09	0.36	-0.090	0.580	0.92
Ponte Branca (*N*=11)						
S0225	3	0.48	0.36	0.250	0.220	1.00
SW444	2	0.51	0.27	0.470	0.160	1.00
S0226	2	0.51	0.64	-0.270	0.940	1.00
SW957	6	0.77	0.82	-0.065	0.750	1.00
IGFI	6	0.79	0.82	-0.034	0.730	1.00
SW857	7	0.86	1.00	-0.170	1.000	0.91
All loci	4.33	0.65	0.65	0.030	0.720	0.99

* Denotes significant *f* values after standard Bonferroni correction for multiple tests

Table 4.11 Summary statistics for white-lipped peccary populations from the best BAPS partition

Locus	No. Alleles	He	Ho	*f*	*p*-Value	% Complete Genotypes
PEMD1/PEMD2 (*N*=12)						
S0225	3	0.67	0.75	-0.125	0.780	1.00
SW444	4	0.52	0.55	-0.043	0.740	0.92
S0226	5	0.73	0.67	0.083	0.360	1.00
SW957	9	0.87	0.80	0.090	0.400	0.83
IGFI	2	0.25	0.27	-0.110	1.000	0.92
SW857	1	-	-	-	-	-
All loci	4	0.61	0.61	-0.021	0.600	0.93
North PEMD (*N*=8)						
S0225	3	0.49	0.63	-0.300	1.000	1.00
SW444	3	0.43	0.50	-0.200	1.000	1.00
S0226	4	0.73	0.88	-0.220	0.940	1.00
SW957	3	0.66	0.88	-0.360	0.970	1.00
IGFI	3	0.66	0.63	0.054	0.580	1.00
SW857	2	0.40	0.50	0.270	1.000	1.00
All loci	3	0.56	0.67	-0.126	0.990	1.00
PEMD3/*Ponte Branca* (*N*=20)						
S0225	5	0.53	0.55	-0.050	0.700	1.00
SW444	3	0.51	0.60	-0.180	0.880	1.00
S0226	6	0.76	0.80	-0.060	0.730	1.00
SW957	5	0.78	0.90	-0.150	0.950	1.00
IGFI	2	0.05	0.05	0.000	-	1.00
SW857	2	0.43	0.40	0.073	0.570	1.00
All loci	3.83	0.51	0.55	-0.061	0.920	1.00
Santa Mônica (*N*=12)						
S0225	4	0.66	0.58	0.120	0.400	1.00
SW444	3	0.54	0.58	-0.080	0.720	1.00
S0226	4	0.70	0.92	-0.340	1.000	1.00
SW957	4	0.57	0.82	-0.460	1.000	0.92
IGFI	3	0.49	0.67	-0.380	1.000	1.00
SW857	1			-	-	-
All loci	3.17	0.59	0.71	-0.228	0.980	0.98

*Denotes significant *f* values after standard Bonferroni correction for multiple tests

Table 4.12 Summary statistics for lowland tapir populations from the best BAPS partition

Locus	No. Alleles	He	Ho	*f*	*p*-Value	% Complete Genotypes
PEMD (*N*=19)						
Tte12	4	0.70	0.74	-0.061	0.480	1.00
Tba20	8	0.87	0.86	0.020	0.540	0.74
Tba23	5	0.69	0.68	0.006	0.400	1.00
HMS7	3	0.59	0.68	-0.160	0.880	1.00
HTG4	3	0.58	0.26	0.550*	0.001	1.00
HTG7	8	0.80	0.28	0.660*	0.000	0.95
All loci	5.17	0.71	0.58	0.169*	0.000	0.95
Ponte Branca/*Tucano* (*N*=6)						
Tte12	3	0.55	0.67	-0.250	1.000	1.00
Tba20	5	0.82	0.80	0.030	0.240	0.83
Tba23	3	0.47	0.50	-0.090	1.000	0.67
HMS7	4	0.78	1.00	-0.330	1.000	0.83
HTG4	5	0.82	0.80	0.030	0.670	0.83
HTG7	5	0.67	0.80	-0.230	1.000	0.83
All loci	4.17	0.68	0.76	-0.140	0.780	0.83
Santa Maria/*Santa Mônica* (*N*=7)						
Tte12	6	0.83	0.86	-0.030	0.740	1.00
Tba20	8	0.87	0.57	0.360	0.020	1.00
Tba23	5	0.90	1.00	-0.140	1.000	0.57
HMS7	4	0.69	0.71	-0.030	0.700	1.00
HTG4	4	0.49	0.57	-0.170	1.000	1.00
HTG7	6	0.85	1.00	-0.200	1.000	0.86
All loci	5.5	0.77	0.79	-0.035	0.450	0.90

***** Denotes significant *f* values after standard Bonferroni correction for multiple tests

Table 4.13 Summary statistics for ocelot populations from best BAPS partition

Locus	No. Alleles	He	Ho	*f*	*p*-Value	% Complete Genotypes
PEMD (*N*=10)						
FCA441	3	0.58	0.40	0.320	0.180	1.00
FCA094	7	0.85	1.00	-0.200	1.000	0.80
FCA201	2	0.19	0.20	-0.060	1.000	1.00
FCA290	4	0.55	0.30	0.470	0.018	1.00
FCA105	7	0.86	0.70	0.200	0.140	1.00
FCA211	6	0.67	0.50	0.270	0.140	1.00
FCA347	3	0.68	0.44	0.360	0.040	0.90
All loci	4.57	0.63	0.51	0.194*	0.003	0.96

North (*N*=2)						
FCA441	2	0.67	0.00	1.000	0.330	1.00
FCA094	4	1.00	1.00	0.000	1.000	1.00
FCA201	1	-	-	-	-	-
FCA290	3	0.84	0.50	0.500	0.330	1.00
FCA105	3	0.84	1.00	-0.330	1.000	1.00
FCA211	2	0.67	1.00	1.000	1.000	1.00
FCA347	4	1.00	1.00	0.000	1.000	1.00
All loci	2.71	0.84	0.75	0.362	0.350	1.00

* Denotes significant *f* values after standard Bonferroni correction for multiple tests

Table 4.14 Summary statistics for jaguar populations from BAPS best partition

Locus	No. Alleles	He	Ho	*f*	*p*-Value	% Complete Genotypes
PEMD (*N*=7)						
FCA441	2	0.36	0.43	-0.200	1.000	1.00
FCA094	3	0.60	0.71	-0.200	0.850	1.00
FCA290	2	0.53	0.57	-0.100	0.850	1.00
FCA105	4	0.65	0.86	-0.350	1.000	1.00
FCA211	4	0.65	0.57	0.130	0.270	1.00
FCA347	3	0.57	0.71	-0.300	1.000	1.00
All loci	3	0.56	0.64	-0.170	0.830	1.00
Ivinhema (*N*=4)						
FCA441	3	0.61	0.75	-0.280	1.000	1.00
FCA094	2	0.43	0.50	-0.200	1.000	1.00
FCA290	2	0.25	0.25	0.000	-	1.00
FCA105	4	0.75	0.75	0.000	0.770	1.00
FCA211	3	0.80	0.67	0.200	0.460	0.75
FCA347	4	0.82	0.75	0.100	0.250	1.00
All loci	3	0.61	0.61	-0.030	0.410	0.96

* Denotes significant *f* values after standard Bonferroni correction for multiple tests

Table 4.15 Results of AMOVA for the best BAPS partition for the *Pontal do Paranapanema*. Jaguar (whole) includes samples from *Anaurilândia* and *Ivinhema*

Species	Φ_{ST}	p
Collared Peccary	0.19	0
White-lipped Peccary	0.13	0
Lowland Tapir	0.1	0
Ocelot	0	0.78
Jaguar (PEMD)	0	0.94
Jaguar (Whole)	0.16	0

Discussion

Noninvasive sampling and genotyping

Genetic studies that use mainly noninvasively sampled DNA are increasingly more common (for reviews Broquet *et al.*, 2007; Chapter 2). Most of these studies collect fresh samples or sample in environmental conditions that are favorable to the preservation of DNA (i.e. extremely dry conditions, or extremely cold). The region of the *Pontal* is situated close to the Tropic of Capricorn, and therefore is very warm and humid, which probably significantly reduces the "half-life" of DNA in the field (Hájková *et al.*, 2006). In an attempt to reduce this effect, most of the sampling was carried out during the dry, and cooler, winter and spring months, and during the cooler period of the day. Nevertheless, due to the low density of the species examined, most of the samples found were of sub-optimal quality, being at least more than a day old. In the specific case of collared and white-lipped peccaries, no reliable dung samples were found (which seems to be an issue in other areas, A. Keuroghlian, personal communication). As a result, of the total of 96 samples collected, only 17 yielded usable genotypes (Table 4.2). To address these shortcomings, future work in the region should concentrate sampling in the larger forest fragments where densities of animals are higher, and possibly use scat scenting dogs (Wasser *et al.*, 2004) to improve the rate and quality of samples recovered. In addition, the sampling protocol designed for elephants (Fernando *et al.*, 2003b) was not as efficient with the examined species, in which even with fresh samples good quality DNA was hard to obtain. We are currently testing a different sampling protocol consisting on scraping the whole surface of a fresh dung sample with a cotton swab or plastic spoon. This sampling protocol focuses on sampling only the target species

epithelial cells, allows for better preservation by complete mixing of sample and preserving agent (such as alcohol or RNAlater), and facilitates manipulation in the lab because the sample is totally dissolved in liquid.

In genotyping dung samples, we found that some primer pairs, which worked with blood samples, did not perform well with low quality and/or degraded DNA. In particular, tapir loci Tte1, Tte5, Tte9 and Tba15 (Norton & Ashley, 2004) were dropped from this study due to their inability to amplify target sequences from dung samples. This observation has been corroborated by at least three independent, on-going, studies with two other species of tapir. It is not clear, at the moment, why the primer pairs do not work on these samples. However, we do suggest that new primer pairs should be developed for this species.

Genotyping errors

As indicated by Paetkau (2003), errors do occur, and it is of paramount importance that we try to identify and minimize the sources of these errors. To examine the reliability of the genotypes used, we computed ADO and FA rates for the blood (Table 4.3) and dung samples (Table 4.4). In the case of blood samples, the specified rates are only an estimate based on the repetition of a subset of the available samples. If we use the estimated rates to derive expectations of the number of wrong genotypes for the remaining blood samples, we find that on average we would expect to find another 7.02, 11.4, 3.36, 3.43 and 2.4 ADO type errors in collared peccaries, white-lipped peccaries, lowland tapir, ocelot and jaguar genotypes, respectively, in the remaining genotypes. As mentioned before, these are probably upper bound rates, and we expect the realized number of errors

to be smaller than the estimates above. In the case of FA rates, they were much smaller than ADO, and the anomalies observed were quite easily spotted on the eletropherograms. As such, we are confident that the genotypes from blood samples form a reliable dataset. The dung samples produced much higher rates of ADO, as expected. However, in this case, it is the actual observed value, rather than an estimate. In addition, it is comparable to ADO rates seen in other studies (e.g. Frantz *et al.*, 2003). FA rates were comparatively low, and similar for FA rates observed for blood samples. As such, we are confident that our exhaustive PCR procedure has resulted in an overall reliable database.

Inferred population structure

We infer the genetic structure of a diverse group of mammals within a landscape that has been fragmented within the past 50 years. We use two different approaches to infer population genetic structure, STRUCTURE (Falush *et al.*, 2003; Pritchard *et al.*, 2000) and BAPS (Corander *et al.*, 2003). STRUCTURE results suggest there is no population genetic structure within this landscape for any of the examined species. BAPS, on the other hand, was able to find significant clustering within the *Pontal* for collared and white-lipped peccaries, lowland tapirs and ocelots, and within a larger area (which includes the *Pontal*) for the jaguars. In a similar setting, a model that took the spatial distribution of individuals into account was able to identify population genetic structure in European roe deer that went undetected with STRUCTURE (Coulon *et al.* 2006). And, similarly to the examined landscape in the roe deer study, fragmentation of the *Pontal* is a recent event relative to the number of generations that has transpired in the focal species since

fragmentation. The small number of generation since fragmentation, in addition to the small samples sizes, leads to relatively shallow likelihood topographies for the Bayesian clusters (Mank & Avise, 2005). But, because BAPS takes as input groups of individuals as determined by the sampling design (rather that just individuals themselves), effectively using an informed prior (even though, it assumes all possible clusters as equally likely), and will analytically solve the posterior distributions of allele frequencies and clusters, it is more likely to uncover significant clusters than STRUCTURE (Mank & Avise, 2005). This gain in power should be treated with caution, for it is possible that using informed priors may override the genetic signal, identifying genetic clusters when in fact there are none (Corander *et al.*, 2003; Mank & Avise, 2005). However, in Chapter 3 I show that the model will consistently identify some level of clustering (albeit an underestimate of the actual number of clusters) in cases when time since fragmentation has been too short and there is a significant difference in population sizes among subpopulations. This suggests that, while the results here may be underestimates of the actual number of clusters, there are signals of significant population genetic structure within these populations. Below we examine the inferred structure for each species.

Collared peccary

For collared peccaries, clusters were coterminous with forest fragments (Figure 4.3), suggesting a significant effect of habitat fragmentation. This is further reinforced by the clustering of the two PEMD herds (PEMD and North PEMD; Figure 4.3) that are linked by relatively continuous forest cover (with the exception of the road) but are more distant from each other than are any of the other herds, which are separated by tracks of deforested land. On the other hand, behavioral studies have shown collared peccary herds

to be highly cohesive units, with little evidence of dispersal among herds, or acceptance of immigrant individuals into established herds (Byers & Bekoff, 1981; Keuroghlian *et al.*, 2004). As such, the observed partition could also be related to divisions among herds, either exclusively or in addition to isolation due to fragmentation. The second best partition had high posterior probability, suggesting some relationship between the *Tucano* and the *Ribeirão Bonito* group. Given the species' tolerance of human disturbances (Altrichter & Boaglio, 2004), and the natural corridor that is formed by gallery forests between these two fragments (Chapter 5), it is possible that these two groups are connected in some way. The degree of differentiation between *Ribeirão Bonito* and the PEMD herds, and the difference in size between the *Tucano* and *Ribeirão Bonito* fragments, suggests that individuals from the *Tucano* herd re-colonized the *Ribeirão Bonito* fragment relatively recently. In this case, some degree of connectivity among fragments is implied, suggesting that, for this species at this scale, social barriers may be more important than landscape barriers in determining levels of connectivity. However, we need a better understanding of how new herds are formed to explain the patterns of clustering among the PEMD herds, and the possible cluster *Ribeirão Bonito*/*Tucano*.

White-lipped peccaries

In white-lipped peccaries, a more complex clustering pattern is observed (Figure 4.3). First, the two herds located below the road cluster together. Studies of white-lipped peccaries have shown that, in ideal habitat, the species will roam over large areas (Fragoso, 1998), and herds will be composed of a number of sub-herds that fuse and split over time in a roughly random way (Keuroghlian *et al.*, 2004). Therefore, given the close geographical proximity of these two herds, it is quite possible that they are sub-herds of a

larger herd. Second, North PEMD 2 clusters with the *Ponte Branca* herd. Given that the species is significantly affected by human densities, and will avoid moving through degraded habitat (Altrichter & Boaglio, 2004), this is an unlikely cluster. Yet, as seen in Chapter 3, when isolation is recent, BAPS will sometimes cluster groups based on past connectivity due, most probably, to incomplete lineage sorting and gene flow-genetic drift disequilibrium. A closer look at the second best partition (Figure 4.4) shows the North PEMD 2 clustering with North PEMD 1 instead, and *Ponte Branca* clustering on its own, which is a more likely partition based on the current knowledge of white-lipped peccary biology. If this is indeed the case, this clustering pattern is evidence of past connectivity that has been lost due to habitat fragmentation. Another point against the clustering of North PEMD 2 and *Ponte Branca* herds is the lack of clustering among the herds within the PEMD that are separated by the road. This lack of clustering suggests that the road is acting as an isolation barrier, which is corroborated by preliminary ratio-telemetry data in the region (A. Nava, personal communication). This observation strengthens the idea that these species are highly sensitive to disturbances, and avoid areas of relatively high human densities (Altrichter & Boaglio, 2004). Finally, the separation of *Santa Mônica* and *Ponte Branca* into individual clusters separate from all other herds is consistent with the above observations.

Lowland tapirs

The tapir is the species with the longest generation time among the herbivores, and as such the analysis of the tapir is the most likely to suffer from limitations imposed by the short time (measured in generations) since fragmentation. In addition, most of the lowland tapir samples used in this study were collected in the region south of the road

within the PEMD. This concentration of samples compared to other groups could exacerbate the limits of BAPS even further when coupled with a relatively short time since fragmentation (Chapter 3). Nevertheless, three clusters were found in the optimal partition, assigning each *a priori* grouping to a separate cluster. These clusters were based on past fragmentation history, and therefore the observed clustering may be a reflection of past rather than present landscape connectivity. In addition, given the limitations above, it is possible that these clusters are an underestimate of the true number of clusters. For instance, the overall significant heterozygote deficiency in the PEMD group (Table 4.12) might be evidence of unidentified population sub-structure (Wahlund effect). Although, given that only two loci were found to have significant heterozygote deficiency, null alleles or sampling error (due to the small sample size) might also have caused the deviation from HWE. In any case, lowland tapirs seem to be significantly affected by the landscape fragmentation.

Ocelots

The clustering of both the East and West PEMD groups implies a high degree of gene flow within the PEMD (Figure 4.3). In addition, radio-tracking of an individual captured within the West PEMD area (which was here grouped, *a priori*, as part of the West PEMD) suggests that it is a resident of one of the fragments west of the PEMD (possibly *Ribeirão Bonito* or *Tucano*), which indicates that there could be gene flow among fragments that are relatively close. The two individuals in the Northern Fragments (Figure 4.1), however, did not cluster with the PEMD cluster. While the isolation of this group into its own cluster could be related to sample size, as implied by the second best partition (Figure 4.4), radio-tracking evidence suggests that individuals will not move

long distances unless they are put under special conditions. The individual captured around the Lua Nova fragment (known as "Juninho"), was released on the western boarder of the PEMD, and after a few days journeyed back to his capture site some 30km away, indicating an ability to move large distances in the landscape. But, one juvenile (pre-dispersal age) captured in the West PEMD area was last spotted in the East PEMD area (suggesting it dispersed to a new area within the PEMD), rather than to an area above the road, or outside the PEMD. Finally, in other areas, juvenile ocelots seem to disperse over an average of 10km (Crawshaw & Quigley, 1989). The best partition, coupled with the information available on these species, therefore, suggests that while they have the capacity to move through the landscape, they will avoid it, and will prefer known areas with suitable habitat cover. Finally, the PEMD cluster displayed significant overall heterozygote deficiency, which was due to five loci presenting high, yet insignificant, *f* values (Table 4.13). The Northern Fragments cluster also had a high overall *f* value (albeit non-significant) with high *f* values at three loci (and possible fixation at one locus; Table 4.13). The general lack of heterozygotes in both clusters may be a sign of significant inbreeding in these subpopulations. Further samples would be required to test this hypothesis.

Jaguars

We examined population genetic structure in jaguar populations at two difference scales. First, we examined samples within the *Pontal*, which included only samples within the PEMD (although, there is little evidence of jaguars established in other areas of the *Pontal*). In this case, BAPS was unable to identify more than one cluster, which is consistent with the idea that these animals will move over large distances. In addition, we

also examined genetic structure at a larger spatial scale, including and two extra groups: one with an individual from *Anaurilândia* and another with five individuals from *Ivinhema*. Here, the partition separated each group into isolated clusters, which suggests some degree of isolation. However, the second best partition grouped the individual from *Anaurilândia* with the *Ivinhema* group, which are farther apart than either is from the PEMD. So, contrary to the best partition, the second best partition suggests some degree of landscape connectivity. In this case, as with the tapirs, we believe that the results reflect a historical association rather than contemporary gene flow. At the western tip of the *Pontal do Paranapanema*, where the Paraná and Paranapanema Rivers meet there used to be a large forested area that was inundated with the damming of the two rivers, leading to the dispersal of many of the animals in the area (D. Sana, personal communication). Therefore, it is possible the resident jaguar population was divided, with some going north towards *Anaurilândia*, and some going south towards *Ivinhema*. In addition, the radio-telemetry data from the PEMD show that animals are highly faithful to their territories (much in the same way displayed by the ocelots). As such, even though these animals display a high capacity to disperse in the habitat (as indicated by the link between *Anaurilândia* and *Ivinhema*), they will most likely not disperse very far, or will be unlikely to successfully disperse very far, as suggested by the isolation between the three groups found in the best partition.

Geographical barriers to gene flow

Comparisons among the species suggest that habitat fragmentation may have a significant effect on all species, albeit at different scales. In collared peccaries, there seems to be fine

scale clustering, associated not only with deforested areas, but also possibly with social structure. Meanwhile, for white-lipped peccaries, social barriers seem less important than habitat fragmentation, with landscape features as fine as the road crossing the PEMD having a significant effect. In tapirs, there is still no clear genetic indication of the contemporary effects of fragmentation. But, the current best partition is highly correlated with historical patterns of fragmentation, and it is possible that the species will eventually display finer clustering patterns. A similar pattern was observed in bush-crickets, when landscape changes were at a faster pace than the ability of the evolutionary process to respond (Holzhauer *et al.*, 2006). Ocelots have the ability to move across the landscape, over large distances, yet their reluctance to do so implies that habitat fragmentation is limiting their movements, which is reflected in the best partition found with BAPS. The scale at which ocelot movements become limited is still to be determined, however it is certain that it is contained within the scale of the *Pontal*. Finally, the jaguars are also affected by fragmentation. However, the spatial scale of fragmentation that affects their movement is most probably larger than the *Pontal*. In an analogy, the PEMD, *Ivinhema* and *Anaurilândia* are to the jaguars what the PEMD, *Ponte Branca* and *Santa Mônica* are to the white-lipped peccaries.

Biological correlates of spatial genetic structure

Studies in landscape ecology have shown that animals of differing body size will respond in similar ways to landscape heterogeneity, by displaying similar movement patterns, but at different scales (Crist *et al.*, 1992). The results of inferred spatial clusters in this study seem to suggest that this may be the case for the spatial structuring of genetic variation.

Within each trophic level, body size is correlated with the scale of structuring observed, with the smaller collared and white-lipped peccaries showing finer scale clustering patterns than tapirs; and ocelots showing finer scale clustering patterns than jaguars. While body size is also correlated to number of individuals (negatively) and to generation time (positively), and therefore some differences should arise among the species, the high correlation between body size and median dispersal distance in mammals coupled with the results presented here suggests that even in non-equilibrium situations body size may be a good predictor of degree of genetic structuring within a landscape.

Other studies have shown that the degree of diet specialization (Brouat *et al.*, 2003), and habitat breadth among species with similar diets (Nupp & Swihart, 2000) can be significantly associated with landscape responses in animals. However, little has been shown as to the effect of different trophic levels, even though it affects the scale of dispersal in mammals (Sutherland *et al.*, 2000). The results presented here, nevertheless, indicate that trophic level may be an important factor too (Table 4.15). While herbivores exhibit significant structure within the *Pontal* landscape, the same is not seen in carnivores. Again, differences in population structure among herbivores and carnivores are expected, given differences in population size between the two levels.

In our dataset, nowhere is the relationship between body size and trophic level better illustrated then by the comparison between jaguars and collared peccaries. In this case, jaguars seem to display similar degrees of genetic differentiation as do collared peccaries, but at a much larger spatial scale (Table 4.15).

Conclusions

Most empirical studies in landscape ecology and genetics have focused on understanding the effects of the landscape on a single species. However, few have examined differences among several species in the same landscape. Here, we present comparisons in the population genetic structure of five different mammals species. These species differ in both body size and trophic level, two characteristics that are highly correlated with median dispersal distance in mammals. We were able to identify significant genetic structure among the sampled populations in spite of our small sample size and the relatively short time the habitat has been fragmented (implying that these populations have not reached genetic or population equilibrium). Furthermore, the pattern of genetic structuring correlates well with what was predicted at the beginning, suggesting that body size and trophic level are good predictors of the degree of genetic structuring to be expected within a given landscape.

Chapter 5. Spatial distribution of genetic variation: Inferring landscape connectivity across body size and trophic level

To be submitted as

Gonçalves da Silva A, Cullen Jr. L, Medici EP, Nava A., Valladares-Padua C & Melnick DJ **Are body size and trophic level good predictors of landscape connectivity in mammals?**

Abstract

Landscape connectivity is a measure of the ability of a species' individuals to move in a landscape. Patterns of isolation-by-distance have been used at fine spatial scales to infer landscape connectivity. The assumption is that the pattern of isolation-by-distance will be more evident along preferred dispersal pathways. Here, we use three different measures of geographic distance (Euclidian, and two effective distances) and two different measures of genetic relationship (Moran's *I* and Rousset's *â*), to test patterns of isolation-by-distance in four mammal species across the same fragmented landscape. In particular, we examined ocelots (*Leopardus pardalis*), lowland tapirs (*Tapirus terrestris*), white-lipped peccaries (*Tayassy pecari*) and collared peccaries (*Tayassu tajacu*) in the *Pontal do Paranapenama* region of Brazil. The species differ in average body size and trophic level, which are highly correlated with dispersal distance. Results were consistent between both measures of genetic relationship within each species. In addition, we found significant isolation-by-distance among herbivores, and suggested, but not significant, isolation-by-distance in ocelots. When comparing different measures of geographic distance, the effective distance that placed an equal negative weight to all non-native vegetation classes (pastures and sugar-cane plantations) presented slightly stronger patterns of isolation-by-distance than the other two measures for both species of peccaries. This finding is corroborated by what is known of the species, and by movement patterns of other species in the same landscape. In tapirs, the pattern was equally strong with Euclidian distances and with the effective distance that placed unequal negative weights on non-native vegetation classes (lower for sugar-cane plantations). Nevertheless, these two measures of geographic distance are highly

correlated for tapir samples, and in spite of evidence for the species using non-native vegetation other factors do not allow us to eliminate Euclidian distances. Therefore, it is currently not clear whether tapirs preferentially move through sugar-cane plantations, or are insensitive to landscape features at the studied scale. Finally, the patterns of isolation-by-distance seem to occur over a small portion of the landscape, followed by a region of little to no correlation between genetic relationships and geographic distances. The point of inflection, where gene flow and genetic drift are at equilibrium, is used to infer the spatial scale of effective dispersal. When comparing across species, the results suggest increasing dispersal scale for herbivores with increasing body size. And, results for ocelots suggest similar dispersal scale to tapirs. While there are still some confounding factors, the results corroborate the idea that body size and trophic level may have some effect on the spatial distribution of genetic variation.

Keywords: Landscape connectivity; isolation-by-distance; ecological genetics; dispersal; mammals

Introduction

The ability of a given species to disperse in a fragmented landscape depends on the species' capacity to disperse across less desirable habitat conditions relative to the scale of habitat fragmentation (Wiens *et al.*, 1997). This interaction is usually defined as landscape connectivity (Calabrese & Fagan, 2004; Merriam, 1984).

Landscape connectivity has a significant impact on dispersal, influencing how individuals are distributed within a landscape (e.g. With & Crist, 1995), the structure of dispersal (i.e. sex-biased dispersal, Stow *et al.*, 2001), and what paths are used to disperse among suitable habitat patches (Crist *et al.*, 1992). The effects of fragmentation depend largely on the species ability to detect suitable habitat at a given scale (Zollner & Lima, 1997) on the species' degree of habitat specialization (With & Crist, 1995) and on its life history traits (Sutherland *et al.*, 2000). This intimate association between a species and its landscape has led population geneticists to hypothesize that the landscape context of a species will affect the spatial distribution of genetic variation (for review see Storfer *et al.*, 2007). This hypothesis is based on the assumption that dispersal and gene flow are highly correlated, and therefore landscape connectivity would affect gene flow and ultimately the spatial genetic structure. To test this hypothesis, geneticists have used prior knowledge of the species habitat preferences and detailed maps of study areas to construct most probable dispersal routes. From these, they have built matrices of effective geographic distances (*cf.* Verbeylen *et al.*, 2003) and evaluated how well they explain genetic distances compared to simple Euclidean geographic distances in an isolation-by-distance framework (Wright, 1943). Using this approach, Keyghobadi *et al.* (1999) studied the dispersal patterns of butterflies, and found that genetic distances among

habitat patches were better explained by geographic distances that maximized the species preferred habitat.

For any given species, the exact pattern that arises will depend on whether the scale of the species' range relative to its median individual dispersal distance is of sufficient extent to meet the assumptions of the isolation-by-distance model. The two fundamental assumptions of the isolation-by-distance model are that (1) dispersal is restricted, but still occurs, and (2) dispersal occurs at an equal rate in all directions (i.e. homogeneous). The first assumption (of restricted gene flow) will be violated if there is either a great amount of gene flow across the species' range, or no gene flow at all (Figure 5.1). In either case, there will be no relationship between genetic distance and geographic distance. However, in the first instance, gene flow's homogenizing effect will keep the variance of genetic distance among individuals low. At the other extreme, random loss of genetic variation through genetic drift will cause the variance of genetic distance among individuals to be high. On the other hand, if gene flow is restricted to some degree over the species' range, then a positive correlation between geographic and genetic distance should arise. The specific geographic distance (Euclidian or effective) that better explains genetic differences among individuals throughout the range (homogeneous or heterogeneous). If dispersal occurs preferentially through certain features of the landscape (violating the assumption of homogeneity), then the correlation should be stronger with a geographic distance that captures the heterogeneity of the landscape (i.e. effective distance).

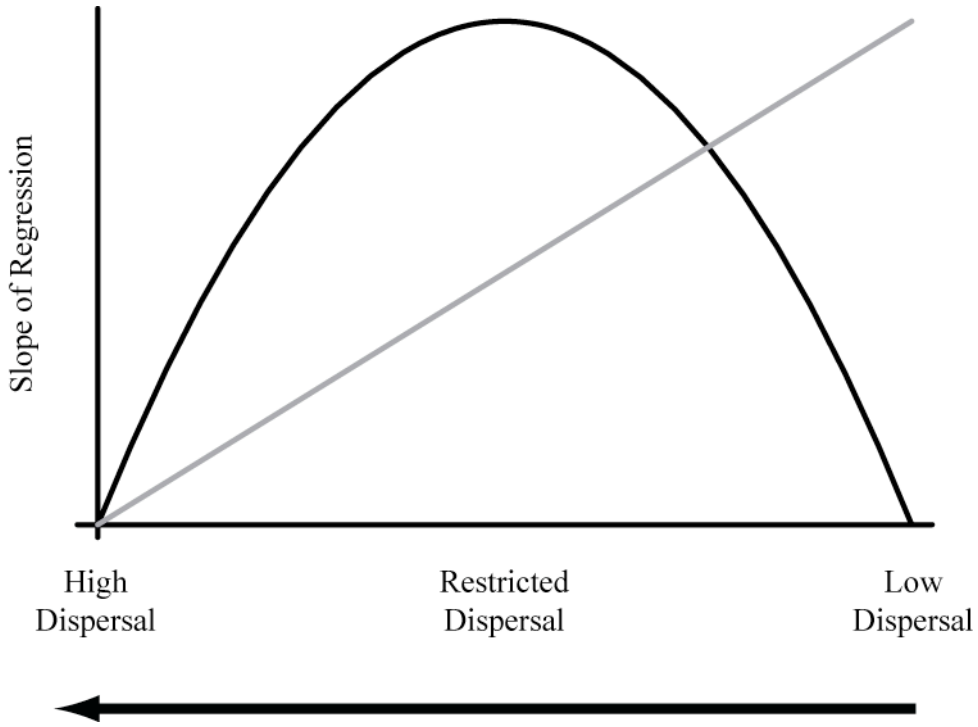

Figure 5.1 Diagram of the relationship between the slope of the regression of geographic distance on genetic distance (black line) and associated increase in variance of genetic distance with decreased gene flow (gray line) and different levels of dispersal. At high dispersal, the variance of genetic distance among all pairs of individuals should be close to zero, and the slope of the regression should be close to zero. At low dispersal, the slope of the regression should also be close to zero, but the variance of genetic distances among all pairs of individuals should be high. At either extreme, the assumption of restricted gene flow is being violated, but in-between these two extremes there is significant isolation-by-distance.

In a study, one can violate the first assumption by sampling an area that is too small or too big relative to the species' dispersal capability, and the second by sampling a landscape that has areas of high risk for dispersing individuals (e.g. fragmented landscape). Therefore, different patterns might emerge for the same species across different landscapes; or among classes of individuals of the same species within a landscape (e.g. sex-biased dispersal). For instance, American martens (*Martes americana*) showed decreased levels of landscape connectivity in a logged area when compared with an unlogged area of similar size (Broquet *et al.*, 2006). In European roe

deer, landscape connectivity was shown to be smaller for females than for males in a fragmented landscape (Coulon *et al.*, 2004). In both these studies, effective distances were able to efficiently capture the effects of the landscape on animal movement. For martens, effective distances were better correlated with genetic distances than Euclidian distances in the logged site, while in the unlogged site both geographic distances were highly correlated, and therefore were equally correlated with genetic distance. In the European roe deer, effective distances were better correlated with genetic distances than they were for Euclidian distances in females, whereas for males, genetic distance was relatively low and constant among most pairs of individuals.

If we can see differences among populations of the same species in different areas and between sexes of the same species in the same area, it is also possible to hypothesize that different patterns may emerge among different species in the same area which are dependent on their dispersal ability. For instance, in plants, dispersal and pollination syndromes are good predictors of the patterns of isolation-by-distance, as they are directly related to dispersal distance (Epperson, 2003). And, in insects, diet breadth and dispersal ability (classified between sedentary and highly mobile) were also good predictors of isolation-by-distance patterns (Peterson & Denno, 1998). In mammals, dispersal distance is highly correlated with body size and trophic level (Sutherland *et al.*, 2000). However, it is still unclear whether they would be good predictors of isolation-by-distance patterns.

Here, we address this gap by comparing isolation-by-distance patterns across four different mammal species (the ocelot, *Leopardus pardalis*; lowland tapir, *Tapirus terrestris*; white-lipped peccary, *Tayassu pecari*; and collared peccary, *Tayassu tajacu*)

with differing body size and trophic levels across the same fragmented landscape to examine putative differences in landscape connectivity. Knowledge of movement patterns for these species is varied. A study of ocelot phylogeography using mtDNA suggests high gene flow among populations with few effective barriers (Eizirik *et al.*, 1998). At smaller spatial scales, ocelots seem to have more restricted movement patterns, preferring to move under the dense cover provided by riparian forests and dense bushes (Jacob, 2002). The lowland tapir, the largest of the three herbivores (Macdonald, 1995), seems to display movement patterns closely associated with water and wooded areas (Padilla & Dowler, 1994). In the *Pontal* region the tapir is known to leave forest fragments to feed in sugar cane plantations, and to search for salt licks in pastures (P. Medici, personal communication). The collared and the white-lipped peccaries differ in their habitat requirements (Fragoso, 1999), though in nature both species can range over wide areas (Carrillo *et al.*, 2002; Judas & Henry, 1999; Keuroghlian *et al.*, 2004). Nevertheless, in more disturbed areas they usually avoid roads and areas of dense human population, especially white-lipped peccaries (Bellantoni & Krausman, 1993). In common, all four species prefer, to different degrees, forested areas to any other habitat type present in the studied landscape.

Based on the algorithm in Sutherland *et al.* (2000) (Equations 1.2-1.4), we estimated median dispersal distances for all four species (Figures 1.4 and 1.5) and compared them to the largest distance between suitable habitat patches. Coupled with the above information on the four species and the studied landscape, we expected all four species to display restricted dispersal over the sampled landscape. However, it is not clear from the models if we should expect the species to violate the assumption of

homogeneous dispersal. Radio-telemetry and observation of tracks (Jacob, 2002; P. Medici, personal communication) suggest that ocelots and tapirs might not violate the assumption of homogeneous dispersal. Observations of peccaries in other landscapes (Altrichter & Boaglio, 2004) suggest that these species might. In addition, we expect that the dispersal scale, as measured by the point where gene flow and genetic drift are at equilibrium, should change with body size and trophic level.

Methods

Study site – The landscape

This study was conducted in the *Pontal do Paranapanema* Region of Brazil, or simply the *Pontal* (Figure 5.2). The study site is contained within an area of roughly 270,000ha that was once the *Grande Reserva do Pontal do Paranpanema* (The Great Reserve of the *Pontal do Paranapanema*, Ditt, 2002, and references therein). This region is located at the western tip of the State of São Paulo (Brazil), and is delimited by the *Paranpanema* River in the South, the *Paraná* River in the West, the *Anhumas* Brook in the North, and the Dividing Ridge of the *Paranapanema-Paraná* River Basins in the East (Ditt, 2002). The area is part of the Atlantic Forest Complex (Morellato, 2000), and its original vegetation cover is specifically classified as Atlantic Forest of the Interior (Ditt, 2002). A GIS database of the *Pontal* classifies vegetation into two broad groups (Uezu, 2007): one with five categories of native vegetation at various succession stages (from pioneer to high canopy mature forest); and another with two categories agriculture (one of sugar-plantations and another of pasture).

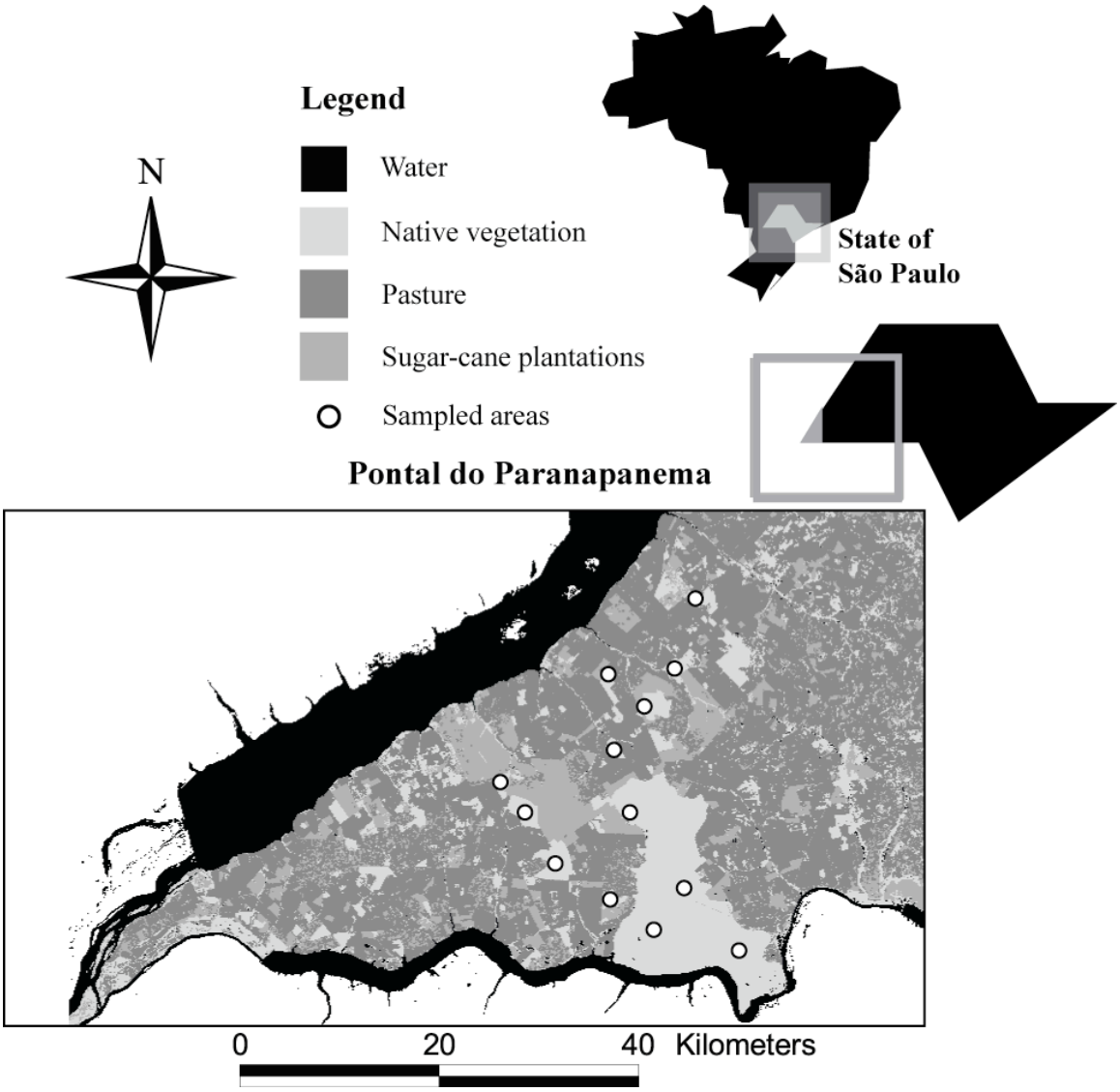

Figure 5.2 The location of the *Pontal do Paranapanema*, and sampled locations

The *Pontal*, and its original forest cover was relatively untouched until the mid-1950's. Since then, it has been reduced to roughly 5% of its original extent (Valladares-Padua *et al.*, 1997), subdividing the forest into hundreds of relatively small forest patches (~345 fragments with areas of 5-2000 ha each) interspersed with agriculture and pasture land (Ditt, 2002). The unique exception is the *Parque Estadual Morro do Diabo* (PEMD), which occupies an area of 37,000ha in the Southeastern corner of the region.

Field collection and laboratory analysis

Both dung and blood samples were collected. Blood samples were kindly provided by researchers from the Institute for Ecological Research (IPÊ), who captured animals for radio-tracking (ocelot and lowland tapir) and conservation medicine (collared and white-lipped peccaries) studies. Dung samples were collected opportunistically in forest fragments in the region. Blood samples were preserved in equal volume of EasyBlood, and 200µl of the mixture was used for extraction using the QIAGEN DNAeasy Tissue Kit (Qiagen, Inc.), following the manufactures instructions, and final DNA was eluted with 200µl of buffer EB (Qiagen, Inc.). Dung samples were collected as fresh as possible (although many were several days old). About 2ml of the surface of the dung was separated into a 5ml sampling tube with a screw top containing ~4ml of RNAlater (Ambion, Inc.). The tube was subsequently vigorously shaken to ensure a good mixture of sample and buffer. Samples were kept at -20 ºC until extraction (between 3-24 months). Extraction of DNA from dung samples followed the protocols described in Chapter 4.

A suite of species-specific and cross-specific microsatellite primers was used to quantify genetic variability (Table 4.1). Additionally, a small 150bp mtDNA sequence was used to discriminate ocelot dung samples from other feline samples (Farrell *et al.*, 2000). Reference sequences were obtained from blood samples. PCR reactions for genotyping were repeated 2-7 times depending on the results, and were carried out in a total volume of 6µl, containing 3µl of QIAGEN Multiplex Master Mix (Qiagen, Inc.), 0.3µM of each primer, and 3µl of DNA extracted from dung or 10-15ng of DNA extracted from blood. PCR was carried out for 35 cycles for blood, and for 45 cycles for

dung samples, preceded by a 15min step at 95 ºC to activate the *Taq* DNA polymerase. One primer from each pair was labeled with a fluorescent marker for automated genotyping. PCR reactions for sequencing followed Farrell *et al.* (2000). Cycle sequencing reactions were carried out directly from PCR fragments using the ABI PRISM® BigDye™ Terminators v3.1 Cycle Sequencing Kit 3.1 (Applied Biosystems) preceded by fragment cleanup using the QIAquick Gel Extraction Kit (Qiagen, Inc.). Cycle sequence reactions were precipitated using the EDTA/Ethanol protocol suggested by the manufacturer of BigDye. Genotypes were obtained on ABI PRISM® 3100 Automated DNA Sequencing machine and processed using GeneScan 3.7 (Applied Biosystems) and Genetyper v3.7 (Applied Biosystems), and sequences were obtained on an ABI PRISM® 3700 Automated DNA Sequencing machine and processed with Sequencher v4.5 (Gene Codes Corporation). All lab work was carried out at the Laboratory of Plant Genetics at the Brazilian Agricultural Corporation – Division of Genetic Resources and Biotechnology (Brasília, Brazil).

Euclidean and effective distances among individuals

Euclidean distances, the shortest "as the crow flies" distance among individuals, were calculated using ArcView 3.3 (ESRI, Inc.) using sample geographical coordinates on a Universal Transverse Mercator (UTM) projection.

To calculate effective distances among individuals, we reclassified a map of vegetation classes to produce what is generally called a friction map (*sensu* Ray, 2005). In a friction map, different features of the landscape are assigned increasing weights as the permeability to animal movement decreases. Following criteria described in Broquet

et al. (2006), we built two different effective distance scenarios, referred here as ED1 and ED2 (Figure 5.4). In ED1, native vegetation was given a weight of 1, plantations a weight of 10, and pastures a weight of 50. In this scenario, we aimed at testing the hypothesis that sugar-cane plantations are more permeable to animal movement than pastures. This scenario is plausible, at least during part of the year, because of the cover provided by mature sugar-cane plantations. In ED2, native vegetation also was assigned a weight of 1, but both plantations and pastures were assigned a weight of 50. In this case, we are testing the hypothesis that both sugar-cane plantations and pastures are equally impermeable to animal movement. This scenario is plausible given the frequent burning of sugar-cane plantations during harvest. Once friction maps were built, least-cost distances among sampled individuals were calculated using the algorithm implemented in the extension PATHMATRIX (Ray, 2005) for ArcView 3.3 (ESRI, Inc.).

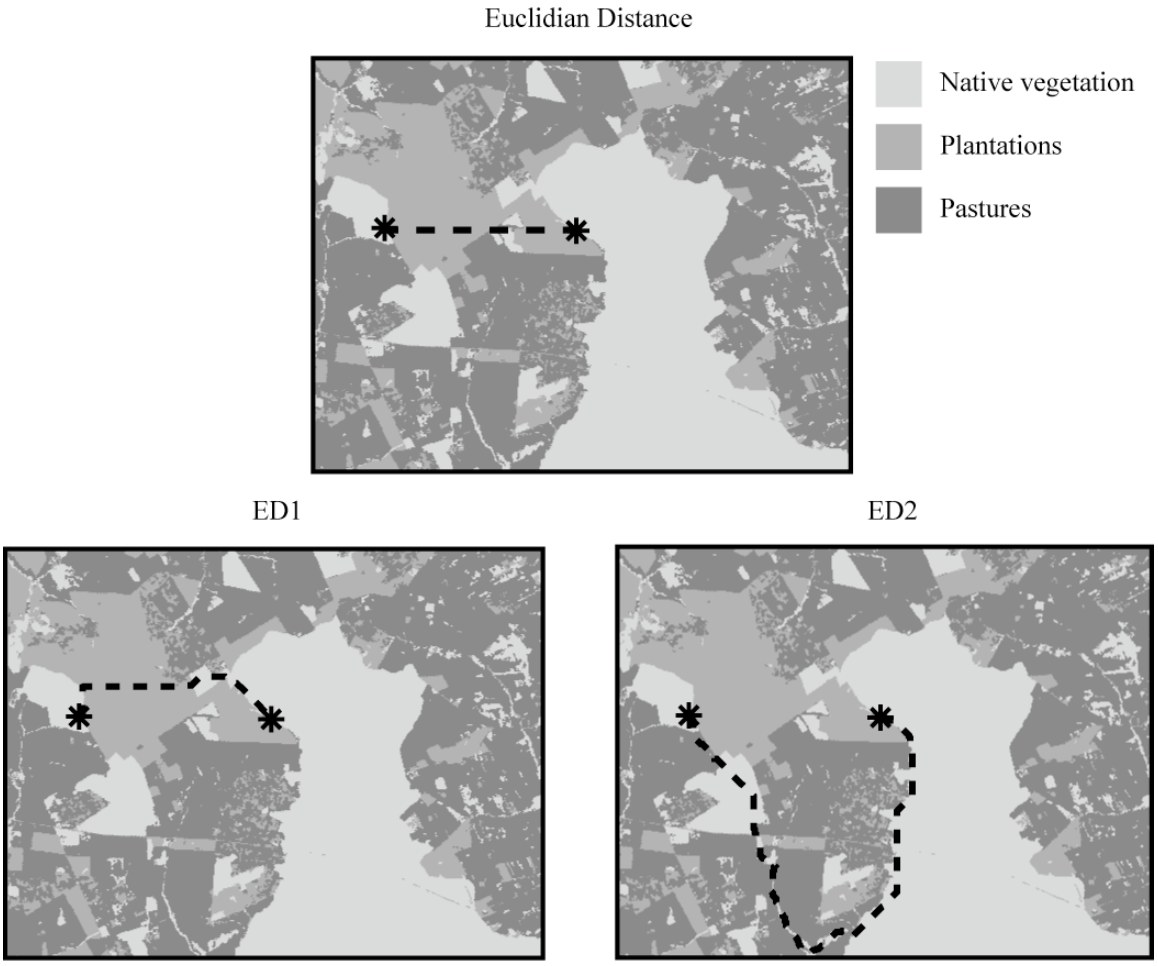

Figure 5.3 Putative dispersal pathways between two individuals with each measure of geographic distance (see text for explanation)

Genetic relationships among individuals

The degree of genetic relationship between all pairs of individuals were estimated using Rousset's (2000) *â*, a measure of genetic distance, and Moran's *I* (Epperson, 2003; Moran, 1950), a measure of genetic similarity. All pairwise genetic relationships were calculated using SPAGEDI (Hardy & Vekemans, 2002). We computed both as they make slightly different assumptions about the data, and Moran's *I* has a lower variance than *â* (Hardy & Vekemans, 2006), even though *â* is used more commonly.

Data analysis

In isolation-by-distance, genetic distances (similarities) are expected to increase (decrease) linearly with the increase in the natural logarithm of geographical distances (Rousset, 1997). To test for this relationship, we examined the slope of the linear regression of genetic relationship onto each geographic distance separately. The significance of the observed slope was assessed by permuting individuals among locations 2000 times. The analysis was carried out using SPAGEDI 1.2 (Hardy & Vekemans, 2002). Significance of regression slopes were assessed by evaluating the proportion of permuted values that were larger than the observed for $\hat{a}$, and smaller for Moran's *I*. This procedure is equivalent to a Mantel's test of matrix correspondence (Hardy & Vekemans, 2006). Geographic distances that were equal to zero were changed to equal to one, so as to include all possible pairwise comparisons in the analysis (Holzhauer *et al.*, 2006).

To visualize the effect of geographic distance on genetic distance, we carried out local polynomial regressions for each distance in each species using the loess function in R (R Development Core Team, 2005). A local polynomial regression is a technique that fits regression models to subsets of the data to build an overall function describing the underlying deterministic variation (Cleveland & Devlin, 1988). Because it gives larger weights to data points that are closer than those farther away, it is able to capture finer details in the relationship between geographic and genetic distances than does a simple linear regression (as performed above). Local polynomial regressions are generally used as a tool to visualize the main trends in this relationship (Broquet *et al.*, 2006; Cleveland, 1993; Coulon *et al.*, 2004; Verbeylen *et al.*, 2003). Unlike the traditional linear regression

in which a specific function has to be specified *a priori*, local polynomial regression only requires two variables to be specified: the degree of the polynomial (usually 1 or 2) and a smoothing coefficient (α) that determines the number of neighboring points used during the calculation of each slope. Here we used a polynomial with degree 2, and α=0.75.

The effect of the landscape

Following methods described in Verbeylen *et al.* (2003), we regressed both effective distances onto Euclidean distances using R (R Development Core Team, 2005). Subsequently, we plotted the standard deviation of the residuals along different classes of Euclidean distance to visualize the effect of the heterogeneity of the landscape independently of distance (Broquet *et al.*, 2006).

Results

Summary genetic data

In general, high genetic diversity was observed for all species (Table 5.1). Sample sizes varied from 12 to 52 individuals, and we observed 6-7 polymorphic loci per species. The number of loci analyzed is at the lower limit recommend for this type of analysis (Hardy & Vekemans, 2006). However, for most species we have well above the recommend lower limit of 100 pairwise comparisons (this number can be smaller when using polymorphic loci, Hardy & Vekemans, 2006).

Table 5.1 Summary of genetic diversity for each species in the *Pontal do Paranapanema*

Species	N	No. Loci examined	Average No. Alleles	He
Collared Peccary	33	6	6.67	0.69
White-lipped Peccary	52	6	5.5	0.56
Lowland Tapir	32	6	7	0.76
Ocelot	12	7	5.43	0.64

Geographic distances among individuals

We plotted average ED1 and ED2 distances within classes of Euclidean distances to examine the differences between effective and Euclidean distances. For collared and white-lipped peccaries (Figure 5.4), ED2 leads to larger distances among individuals than ED1 (Table 5.2). For tapir and ocelots (Figure 5.4), the differences are less pronounced (Table 5.2).

Table 5.2 Number of pairwise comparisons, and mean and maximum distances among individuals for each geographic distance examined. Distances are measured in kilometers.

		Euclidean distance		E1		E2	
	Pair	Mean	Max	Mean	Max	Mean	Max
Collared peccary	528	15.8	44.00	18.8	48.8	24.7	54.15
White-lipped peccary	1326	22.46	45.9	29.47	59.7	35.0	81.0
Lowland tapir	496	25.2	53.5	33.9	70.3	33.9	82.95
Ocelot	66	17.67	48.62	20.4	62.8	20.4	56.15

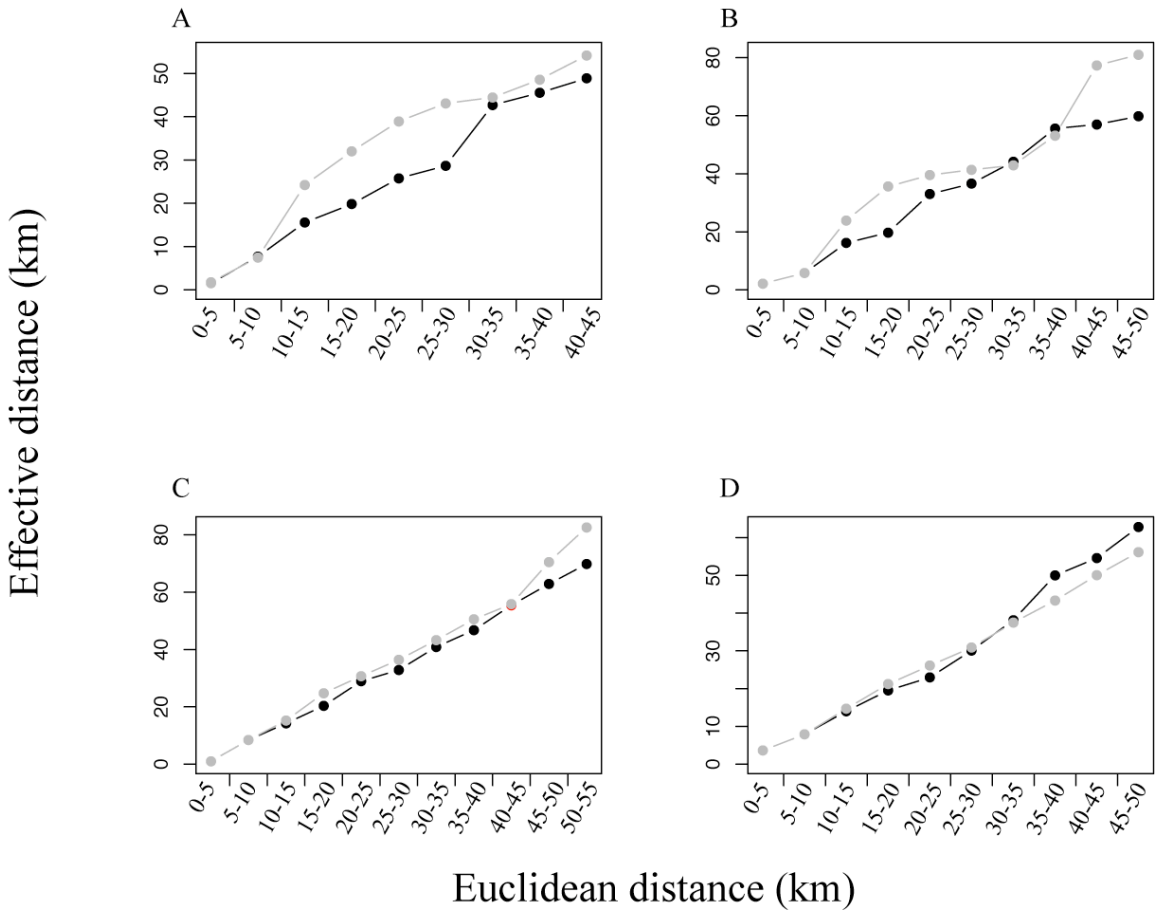

Figure 5.4 Change in effective distances ED1 (black line) and ED2 (gray line) over classes of Euclidian distances calculated among all pairs of individuals for each examined species. A – collared peccary; B- white-lipped peccary; C – lowland tapir; D – ocelot.

The relationship between genetic and geographic distances

Within species, linear regression results were consistent across both measures of genetic relationship. Genetic distances ($\hat{a}$) were significantly correlated with geographic distances for collared and white-lipped peccaries, and lowland tapirs, but not for ocelots (Table 5.3). And, the same pattern was observed for Moran's *I* (Table 5.4).

The amount of variation explained in each model (R^2) was relatively low (Tables 5.3 and 5.4). However, this is to be expected because of high variance associated with measuring genetic distances among individuals (Broquet *et al.*, 2006; Rousset, 2000).

Comparing across measures of geographic distance within each species small differences in the proportion of the variation in genetic relationships explained by each measure of geographic distance can be noticed. For collared and white-lipped peccaries, ED2 was a slightly better predictor of genetic relationships than either Euclidian or ED1. In the case of lowland tapirs, Euclidian and ED1 were slightly better predictors of genetic relationships than ED2.

Local polynomial regressions shows increasing genetic distance at relatively short geographic distances for all three species of herbivores, followed by relative stabilization of genetic distances at larger geographic distances (Figure 5.5). The reverse pattern is observed for Moran's *I* (Figure 5.6). Meanwhile, for ocelots, genetic distances starts relatively high and decrease over short geographic distances, followed by a short increase and stabilization at larger geographic distances (Figure 5.5). Again, the pattern is observed in reverse for Moran's *I* (Figure 5.6). At intermediate and larger geographic distances (>8km) the pattern of genetic differentiation relative to geographic distance resembles what would be expected in isolation-by-distance (i.e. increase in genetic distance until genetic drift becomes stronger than gene flow). The high genetic distance among individuals at short geographical distances, on the other hand, is counterintuitive. A similar pattern was observed for American martens (Broquet *et al.*, 2006), but no explanation was attempted. Nevertheless, the observation is based on a small number of data points, and therefore cannot be considered reliable given the high variance in the measures of genetic distances (and similarity) observed for the other three species in this study, and the high sensitivity of local polynomial regression to outliers (Cleveland,

1993). As such, we are unable to ascertain the biological significance of the high genetic distance at short spatial scales with the current sample size and distribution.

Table 5.3 Summary of linear regression of pairwise genetic distance (*â*, Rousset, 2000) among individuals onto the logarithm of geographic distances

	Euclidean distance			ED1			ED2		
Species	Slope	*p*	R^2	Slope	*p*	R^2	Slope	*p*	R^2
Collared peccary	0.016	0*	0.076	0.015	0*	0.075	0.015	0*	0.077
White-lipped peccary	0.012	0*	0.058	0.011	0*	0.057	0.011	0*	0.06
Lowland tapir	0.032	0*	0.037	0.032	0*	0.037	0.031	0*	0.036
Ocelot	-0.025	0.7	0.008	-0.018	0.64	0.005	-0.013	0.59	0.002

* Denotes significant values at the 0.05 level

Table 5.4 Summary of linear regression of pairwise genetic similarity (Moran's *I*, Moran, 1950) among individuals onto the logarithm of geographic distances

	Euclidean distance			E1			E2		
	Slope	*p*	R^2	Slope	*p*	R^2	Slope	*p*	R^2
Collared peccary	-0.028	0*	0.17	-0.027	0*	0.174	-0.027	0*	0.177
White-lipped peccary	-0.022	0*	0.09	-0.021	0*	0.09	-0.021	0*	0.094
Lowland tapir	-0.052	0*	0.074	-0.054	0*	0.074	-0.052	0*	0.073
Ocelot	0.024	0.83	0.0062	0.021	0.83	0.006	0.024	0.86	0.007

* Denotes significant values at the 0.05 level

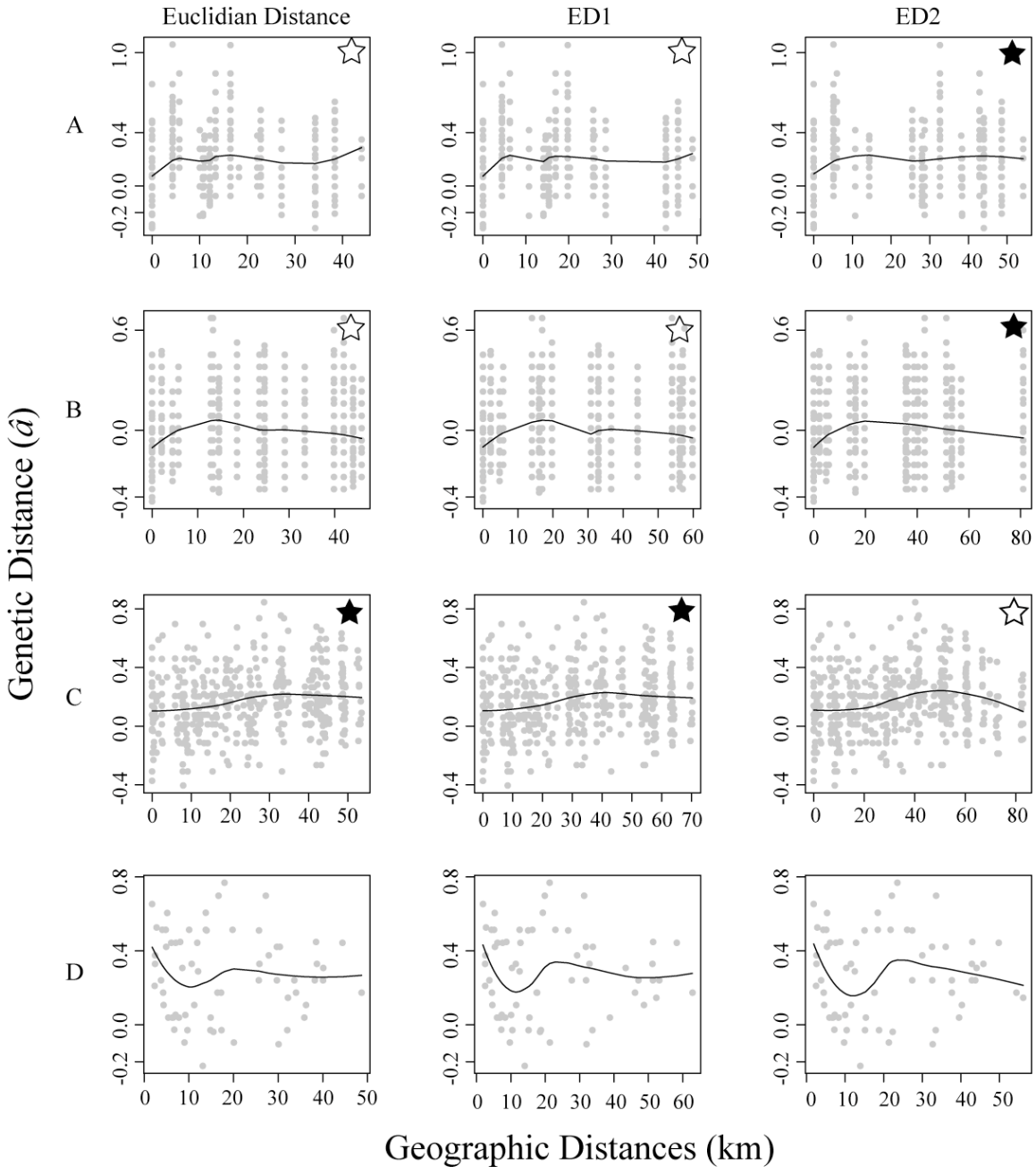

Figure 5.5 Local polynomial regression of individual pairwise genetic distance (*â*) onto geographic distance for each of the examined species (solid line). Individual pairwise genetic distances are plotted as gray points. A – collared peccary; B- white-lipped peccary; C – lowland tapir; D – ocelot. Open stars indicate significant patterns of isolation-by-distance. Solid stars indicate the strongest pattern of isolation-by-distance among the three measures of geographic distance.

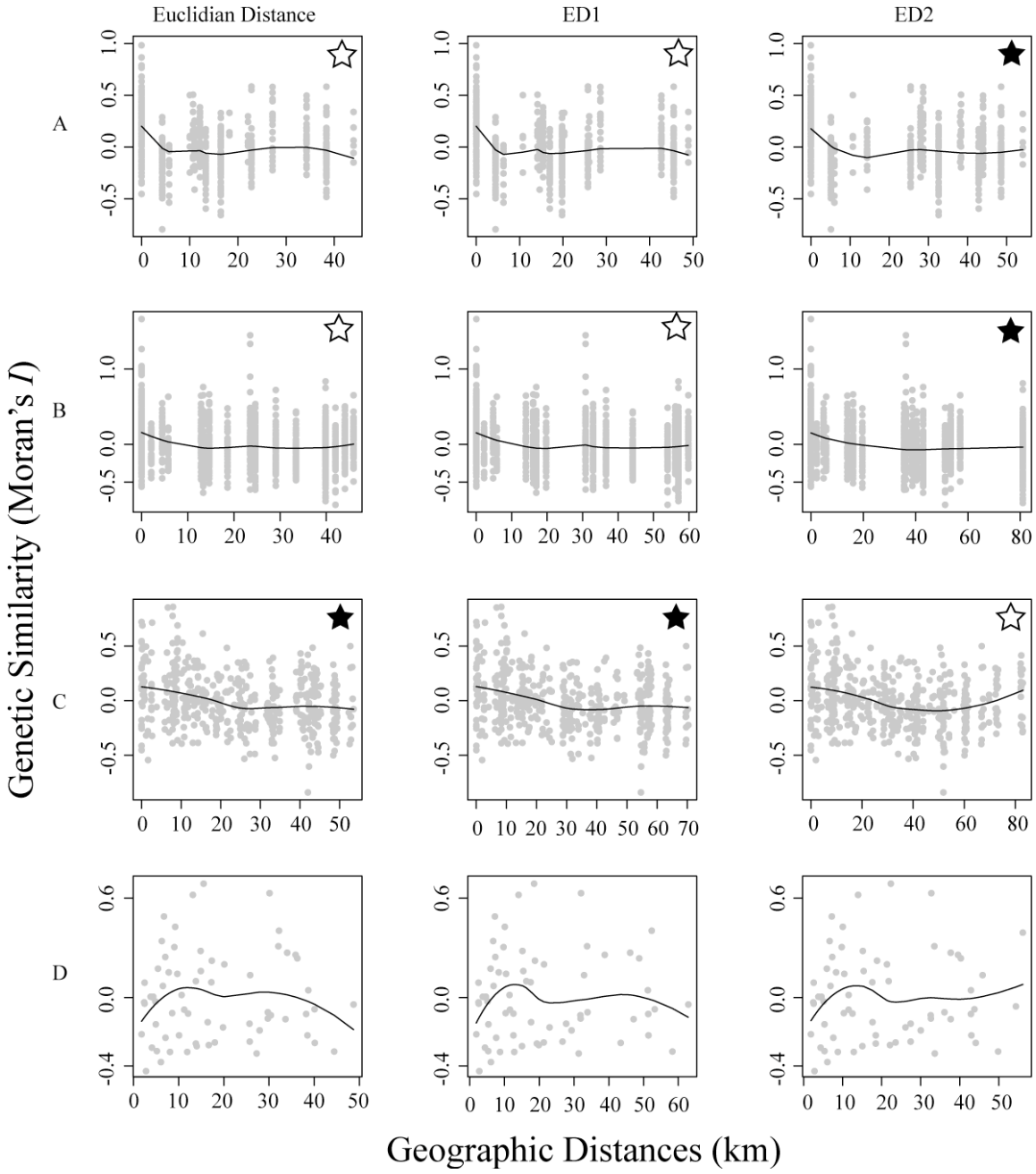

Figure 5.6 Local polynomial regression of individual pairwise genetic distance (Moran's *I*) onto geographic distance for each of the examined species (solid line). Individual pairwise genetic distances are plotted as gray points. A – collared peccary; B- white-lipped peccary; C – lowland tapir; D – ocelot. Open stars indicate significant patterns of isolation-by-distance. Solid stars indicate the strongest pattern of isolation-by-distance among the three measures of geographic distance.

The heterogeneity of the landscape

For all species, scenario ED2 incorporates a higher degree of landscape heterogeneity, as it assumes that animals have equal aversion to moving through both pasture and plantations (Figure 5.7). The ocelot is the only exception, but that is because in scenario ED1 it became "cheaper" to go around a patch of pasture through adjacent plantations, than to go straight through the pasture as in ED2 (i.e. the cost of traveling through plantations is lower than the cost of traveling straight through pastures). If we focus on scenario ED2, we see a steady increase in the standard deviation of the residuals with geographic distance class for both tapirs and ocelots. Meanwhile, for collared peccaries and white-lipped peccaries there is no particular pattern.

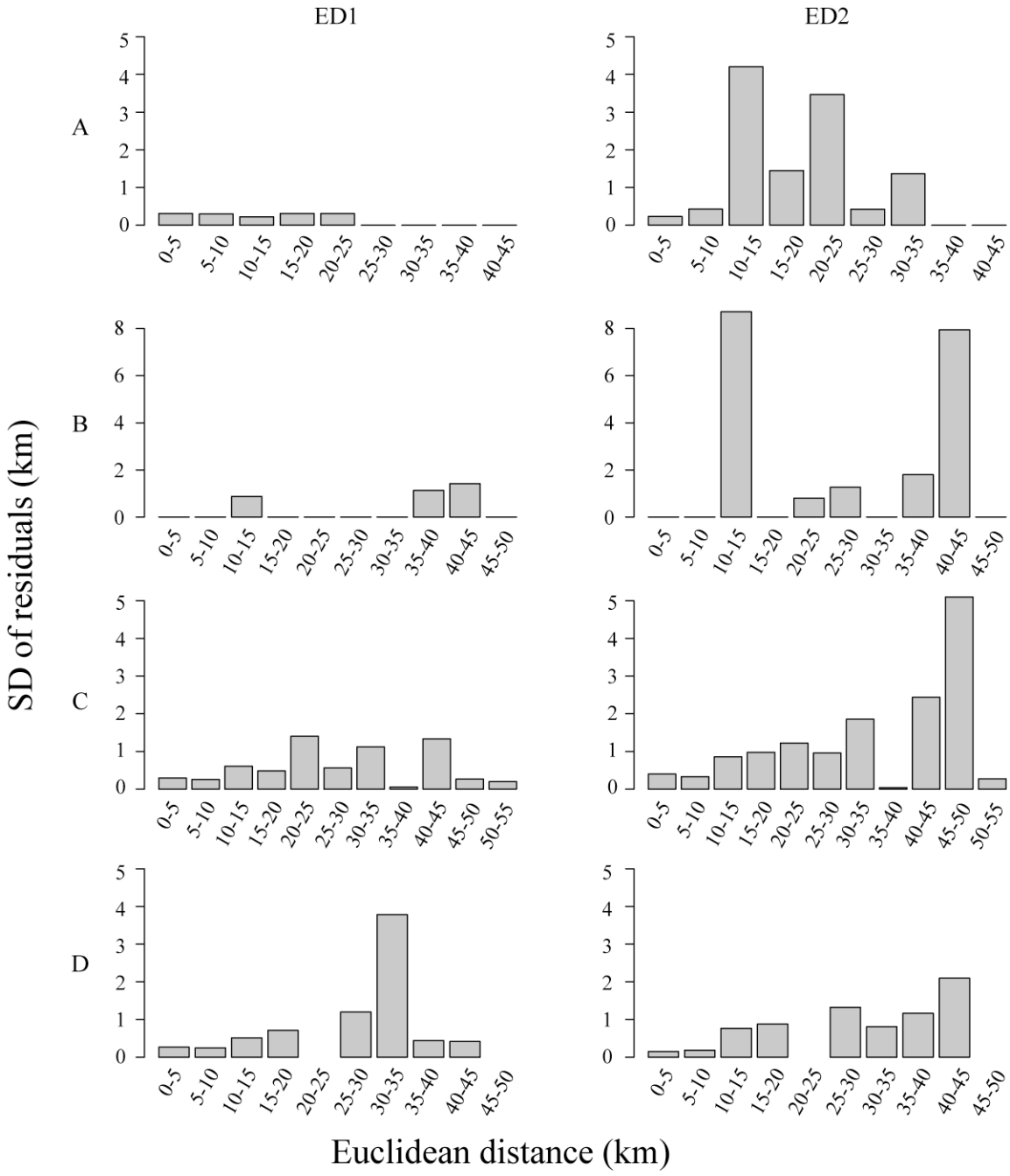

Figure 5.7 Distribution of the standard deviation of residuals of the linear regression of effective distance onto Euclidean distance per class of Euclidean distance. A – collared peccary; B- white-lipped peccary; C – lowland tapir; D – ocelot.

Discussion

Isolation-by-distance

In this study, we examined differences in patterns of isolation-by-distance among four species of non-volant terrestrial mammals within the same fragmented landscape. Among species, geographic distances were significantly correlated with measures of genetic similarity and distance for collared and white-lipped peccaries, and tapirs. For ocelots, no relationship was found. However, visual inspection of the pattern of change in genetic relationships over geographic distance suggests some degree of isolation-by-distance.

In isolation-by-distance, the point of inflection of the curve is the point where gene flow and genetic drift are at equilibrium, and generally marks the geographic scale of effective dispersal (Hutchison & Templeton, 1999). Beyond this point, genetic distances are no longer correlated with geographic distances. In the specific case of Moran's *I*, this point is where the curve reaches zero or becomes negative (Epperson, 2003). Comparing across species, we can see an increase in scale from collared peccaries to lowland tapirs. For collared peccaries equilibrium is reached before 10km, for white-lipped peccaries equilibrium is reached before 20km, and for lowland tapirs it is reached between 20-25km (Figures 5.5 and 5.6). And, if we examine the pattern for distances larger than 8km in ocelots, it suggests an equilibrium point between 20-25km, as well. As such, the results suggest that gene flow reaches equilibrium with genetic drift in each species within the sampled scale. The sampling scale is, therefore, larger than the dispersal scale for the examined species. However, differences in the variance among the observed patterns for each species are hard to assess. Because each locus is assumed to be an independent realization of a stochastic process, measures of genetic relationship

among individuals usually have high variance. Therefore, it is difficult to assess with the current number of loci if variation is increasing at this scale. It is possible that increasing the number of loci could resolve this issue.

Landscape connectivity

To assess differences in landscape connectivity, we compared different measures of geographic distance in their ability to explain variation in measures of pairwise genetic relationship among individuals. The measures of geographic distances used differed in their assumptions about how the landscape affects the movement of individuals. Euclidian distances assume that all features of the landscape are equally permeable to individual movement, and ED1 and ED2 assume that pasture and sugar-cane plantations are less permeable than native vegetation. The different measures of geographic distance reflect different amounts of landscape heterogeneity (Figure 5.7), and are compared by examining differences at the coefficients of determination (R^2). In general, the coefficients of determination were low across the different measures of geographic distance, however, they were still an order of magnitude larger that those found for American martens by Broquet *et al.* (2006). And, even though differences among models were also small, ED2 was a slightly better predictor of genetic relationships in collared and white-lipped peccaries while Euclidian and ED1 were equally better in lowland tapirs (Table 5.2 and 5.3). ED2 measures distances among forest fragments that minimize non-native vegetation (Figure 5.3), in some cases following pathways through gallery forests known to be used by ocelots to move among fragments within the *Pontal* (Jacob, 2002), and pathways known to be preferred by peccaries in other landscapes (Altrichter &

Boaglio, 2004). These characteristics suggest that the small increase in R^2 in ED2 over Euclidian and ED1 distances may in fact be biologically significant. In addition, fragmentation in the landscape is relatively recent (8-10 generations of peccaries), and even though significant isolation-by-distance patterns can originate over relatively small number of generations (≥5, Sokal & Wartenberg, 1983), it is possible that these species are still far from equilibrium and differences observed among the measures of geographic distance might still increase.

In contrast, the high correlation between Euclidean and ED1 distances (Table 5.2; Figure 5.7) makes it hard to assess whether tapirs are insensitive to landscape heterogeneity at this scale, or if they have preferential pathways through sugar cane plantations. This is mainly due to the distribution of samples relative to sugar-cane plantations, resulting in a high correlation between Euclidian and ED1 distances. As such, three scenarios could plausibly explain the observed pattern. First, tapirs are insensitive to heterogeneity in the landscape at this scale, and will move freely throughout the landscape. Second, tapirs are sensitive to the heterogeneity and move preferentially through sugar-cane plantations, a hypothesis that is corroborated by tracks (AGS personal observation). And third, the observed pattern reflects past connectivity, but tapirs move preferentially though areas of native vegetation (ED2). At the moment, it would be difficult to tease apart these different scenarios, and it is possible that it will remain so for a few more generations.

In the case of ocelots, it is possible that outliers are influencing the shape of the local polynomial curve, and the low number of pairwise observations (N=66) coupled with the great variance in measures of pairwise genetic distances have led to regression

slopes that were not statistically significant. Nevertheless, comparison of R^2 over all geographic distances and measures of genetic relationship, suggests that with the current information the best fit is obtained with a model that uses Euclidian distances as predictor of $\hat{a}$ (Tables 5.3 and 5.4). A closer look at the local polynomial curve (Figure 5.5), suggests some degree of isolation-by-distance at geographic distances between 10-20km, followed by a relative stabilization of genetic distances. At distances larger than 20km, where most comparisons are between animals from different forest fragments, the lack of any relationship between genetic and geographic distances could be related to the effects of fragmentation. However, ocelots seem to have the capacity to move large distances through the landscape, as illustrated by the example of an animal captured in a small fragment north of the PEMD and released in the southwest border of the park. After a few days around the release site, the individual returned to its capture location roughly 30km north of the release site (Jacob, 2002). As such, the results presented here corroborate the ecological data in that, even though the animals have the capacity to move in the landscape, they seem to avoid it. But, more conclusive answers require larger samples sizes.

Least-cost distances

To provide a comparative framework, we used the same friction maps for all species. These friction maps were based on GIS maps containing only information on the different classes of vegetation in the landscape. Most ecological work undertaken in the *Pontal* (e.g. Bassi, 2004 and; Jacob, 2002) points to the importance of vegetation cover for the presence and movement of species. However, it is possible that other landscape

factors may be influencing dispersal patterns, and thus levels of connectivity as measured in this study. For instance, areas where forest fragments border pasture land could be associated with a high incidence of disease from spillover (*cf.* Fiorello *et al.*, 2006). As result, dispersing animals could have higher mortality rates in these areas, significantly decreasing connectivity. In addition, human density was not taken into consideration, even though it has an effect on the distribution of white-lipped peccaries (Altrichter & Boaglio, 2004). Therefore, it is possible that more detailed friction maps might result in effective distances that are better able to capture the *Pontal's* landscape heterogeneity, as it is perceived by the species in question. Nevertheless, the level of detail used here was sufficient to capture differences among the examined species, and re-enforces that vegetation types are indeed important to the movement patterns of these species.

â and Moran's I

Most studies of isolation-by-distance at fine spatial scales that focus on the individual as the unit of analysis have used *â* (Rousset, 2000) as a measure of genetic relationship. This measure has its advantages, as it will allow for diverse models of dispersal, and it has a low bias. However, it requires intense sampling to obtain precise results, and therefore usually suffers from high variance. In addition, the underlying model assumes a continuous population, which does not seem to be the case for the populations in this study (Chapter 4). Nevertheless, we justify the use of *â* because we are testing for conditions under which different fragments might be acting as one single population (i.e. landscape connectivity would be highest). And therefore, presumably under such conditions, this measure would be the most adequate.

We also used Moran's *I* (Moran, 1950), which is a measure of autocorrelation among individual genotypes. Moran's *I* has been used in many studies and has well known properties (for review see Epperson, 2003). One of its advantages is that it has a lower variance than $\hat{a}$ (Hardy & Vekemans, 2006), and it has been shown to reflect differences in dispersal capacity in plants (Epperson, 2003).

When comparing both measures, they were consistent across species and measures of geographic distances (Tables 5.2 and 5.3). Nevertheless, for the three species with a larger number of pairwise comparisons, the coefficient of determination was usually higher when using Moran's *I*. Suggesting that in the conditions presented by this study, which are frequently encountered in studies of natural populations, Moran's *I* might be a better measure of pairwise individual genetic relationships. This difference probably stems from the smaller variance associated to Moran's *I* when compared with $\hat{a}$. However, in ideal conditions of sampling intensity and distribution of individuals in the landscape, $\hat{a}$ may be a better measure. There are currently no other studies, to our knowledge, that have compared these two measures.

Conclusions

In this study, we examined landscape connectivity in four different mammal species across the same landscape. As expected from what is known from the species, and estimates of median dispersal distances based on body size and trophic level, we found that all four species did not violate the assumption of restricted dispersal in the isolation-by-distance model. In addition, tapirs may be dispersing homogenously across the landscape, though current data were not able to distinguish between Euclidian distances

and ED1, a measure of geographic distance that sets sugar-cane plantations as more permeable to movement than pastures. It is possible that ocelots may also disperse homogeneously, but more intense sampling would be needed to confirm this. On the other hand, both species of peccaries seem to violate the assumption of homogeneous dispersal, and prefer movement pathways that maximize the use of areas with native vegetation.

Finally, we observed an increase in dispersal scale and a decrease in landscape effects within herbivores that is positively correlated with body size. And, results suggest that a small carnivore (i.e. the ocelot) might have a similar pattern of genetic differentiation over distance as the largest herbivore (i.e. the tapir). This pattern, in turn, indicates that body size and trophic level, as estimators of median dispersal distance, are good predictors of genetic landscape connectivity in mammals. Nevertheless, several confounding factors still need to be addressed. Social structure in peccaries might also be contributing to lower connectivity in the landscape (Dubost, 2001; Keuroghlian *et al.*, 2004), and, the small number of generations since fragmentation probably means that the observed patterns are far from equilibrium.

Chapter 6. Conclusions: Scaling of spatial genetic structure, implication for conservation, and final considerations

In this thesis, I investigated spatial genetic variation in various species of mammals with the intention to uncover significant biological and ecological correlates of spatial genetic structure. To this end, I proposed to examine six species (collared peccary, white-lipped peccary, lowland tapir, ocelot, puma and jaguar) that differed in both body size and trophic level, but were similar in their habitat requirements, within the fragmented landscape of the *Pontal do Parnapanema*, Brazil. Given current knowledge on population genetic theory and landscape ecology, I hypothesized that:

H1.A: Genetic differentiation, as measured by F_{ST}, will decrease with increase in body size within trophic levels

H1.B: Genetic differentiation, as measured by F_{ST}, will be higher for herbivores than for carnivores

H2.A: Landscape connectivity will increase with body size within trophic levels

H2.B: Landscape connectivity will be smaller for herbivores than for carnivores

These hypotheses imply that spatial genetic structure is somewhat scalable. In other words, that the same pattern of differentiation will be seen for different species, but they will occur at different scales. To explore this question, we examined, for each species, the degree of population genetic differentiation using a Bayesian approach (Chapter 4), and the degree of landscape connectivity using an isolation-by-distance framework coupled with a GIS database (Chapter 5).

The nature of the animals investigated and the number of species examined, meant that it would be more cost effective, from a sampling perspective, to work with noninvasively sampled dung. However, this strategy does not come without its limitations (Chapter 2). The major problem with the use of this strategy was related to the

quantity of fresh samples that could be found. In most cases, samples were more than a day old, which severally decreases their potential for extracting suitable DNA for PCR reactions. As a result, our final sample size was significantly affected, and led to the elimination from the analyses of the puma in Chapters 4 and 5. In addition, before addressing the first question, it was necessary to investigate the power of a Bayesian model to identify the correct structure under the limitations imposed by this study (Chapter 3). This was necessary mainly because of the novelty of such models, which lack sufficient empirical tests of their limits. In this case, several limitations were identified *a priori* that could affect the model's power to detect the correct the number of clusters. They were: (1) the short time period since fragmentation relative to generation length in each species; (2) the difference in subpopulation sizes within each species; (3) the differences in dispersal ability; and (4) the sample size that we were able to obtain. The results of the simulations suggested that the main model used here (BAPS, Corander *et al.*, 2003) may have difficulty identifying the correct number of clusters in most situations presented in this study. Nevertheless, it was encouraging that even in the worst case scenario (i.e. five generations since fragmentation, a small and unevenly distributed population, and a small number of markers with relatively low levels of polymorphism) the model was still be able to identify some level of sub-structuring in the data.

The main results of this dissertation (Chapters 4 and 5) are in general agreement with the hypotheses outlined above. In Chapter 4, significant genetic structure was found within each species. It is possible that in some cases, the actual number of clusters was underestimated. Nevertheless, the results are in accordance for what is known about the species within the *Pontal*, and in other areas. A comparison among the species suggests

that landscape fragmentation has a significant effect on the genetic structure of the examined species. But, within the landscape of the *Pontal* the effects vary from species to species following a trend of decreasing genetic structuring with increasing body size within trophic levels **(H1.A)**, and from herbivores to carnivores **(H1.B)**, as was predicted. Similar results were seen in Chapter 5 for landscape connectivity. In this case, while I was unable to examine landscape connectivity in the two larger carnivores (the puma for the lack of samples, and the jaguar for lack of adequate maps), the results for the relatively small ocelots suggest a similar degree of connectivity to what was observed in the large tapirs, which is in agreement with **H2.B**. In addition, collared and white-lipped peccaries not only had a smaller degree of landscape connectivity than the ocelots and tapirs (which is in agreement with **H2.A**), but also showed a slight preference to disperse through forested areas over other features of the landscape.

The results presented here should be interpreted with some caution. Collared peccaries do present the highest degree of structuring, and conversely, the smallest degree of landscape connectivity, but this could be in part related to the barriers imposed by the species' social system. However, the fact that white-lipped peccaries, which do not present such a strict social barrier, and the tapir, which is an essentially solitary species, also display significant genetic structure (Table 4.15) indicates that the degree of genetic structuring displayed by collared peccaries is not solely related social structure. In addition, because of the short number of generations since fragmentation, it is highly probable that some of the examined populations are still far from gene flow-genetic drift equilibrium, but are somewhere in-between the equilibrium of the past, before

fragmentation, and a new equilibrium that may be reached sometime in the future if the current situation persists.

Scaling in spatial genetic structure

In Chapter 1, I calculated the expected genetic differentiation that would arise over increasing median dispersal distances among subpopulations in a stepping-stone model of population structure. Subsequently, I scaled the differentiation to inferred median dispersal distances for each of the six species studied (Figure 1.6; Table 1.2). This exercise implies that the underlying response is scale independent. Therefore, if we were to examine the amount of differentiation in terms of median dispersal units, we should expect to see an increase in differentiation with an increase in median dispersal. To test this, I plotted the F_{ST} values in Table 4.15 against the average distance among individuals (Table 5.2; and 132km for jaguars) scaled to the estimated median dispersal distances in each species (Figure 6.1). The results suggest that there is a general increase in genetic differentiation with an increase in the number of median dispersal units between individuals. Admittedly, the values for jaguars (F_{ST}=0.16 at 0.5 median dispersal units) and ocelots (F_{ST}=0 at 0.5 median units) are very different. But, it should be remembered that jaguars usually have populations that are generally much smaller than ocelot or tapir populations. Therefore, even though there is a suggestion of scaling in genetic differentiation, it may have its limits. The difference between jaguars and ocelots suggests that these limits may be imposed by, among other things, differences in population size.

Intuitively, the scaling process would work as Russian dolls, where as you remove the outer doll, a new doll appears, or as fractal geometric objects, where there is some self-similarity across a certain range of scales (Mandelbrot, 1982). In the specific case examined in this dissertation, the clearest example lies in a comparison of the extremes. Collared peccaries are relatively small herbivores, which at the scale of fragmentation seen in the *Pontal*, have been isolated into smaller habitat patches, displaying some degree of genetic differentiation among them. Nevertheless, dispersal, and thus gene flow, is possible, even if only limited. Conversely, at the scale of the *Pontal*, the jaguar, a much larger animal, was confined to only one fragment that was large enough to sustain a population. However, if examined at a larger scale, several habitat patches appear which are large enough to sustain jaguar populations. At this larger scale, some degree of genetic differentiation is found among the habitat patches, but again there is the possibility of some gene flow among them. And so, as we move from one spatial scale to the other, we see the same patterns of genetic differentiation arising.

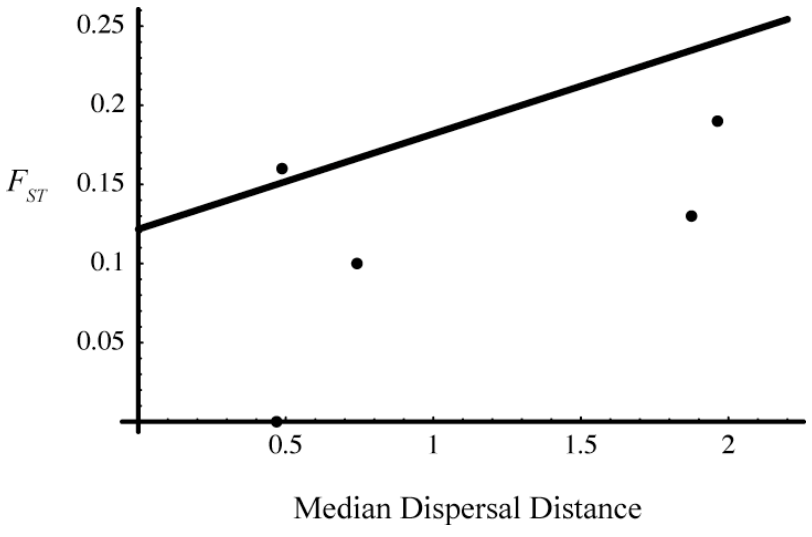

Figure 6.1 The relationship between median dispersal distance and observed F_{ST} values

Implications for conservation of species and landscapes

In general

The potential for the use of landscape genetics in conservation is obvious (Manel *et al.*, 2003). The ability to identify geographical barriers to gene flow, cryptic structure, hybridization zones, dispersal pathways, and recent immigrants is of great importance for the preparation of species and landscape action plans. However, many of the methods, techniques and measures used in landscape genetics require a high density of samples, and a good number of highly polymorphic loci to be effective. For many species this is not possible. Even though the techniques available to develop new markers are becoming simpler and there are possibilities of outsourcing the development of new markers (Selkoe & Toonen, 2006), in many cases, the target species may occur in such low numbers that obtaining meaningful numbers of samples may not be possible. In these cases, it may be preferable to study a similar species occurring in the same landscape, but one that is more abundant. If, for instance, collared peccaries were less abundant than white-lipped peccaries, a study of white-lipped peccaries would be able to uncover shared landscape barriers to gene flow in both species, as they seem to perceive the landscape in the same general spatial scale.

Furthermore, the interpretation of genetic information should always take into account ecological and historical knowledge about the species and the landscape. In the case of this dissertation, I was somewhat privileged to have had an enormous amount of information and knowledge about the species' ecology in the *Pontal* and the history of the landscape at my disposal. This knowledge was fundamental for the interpretation of

the observed genetic patterns. In this dissertation, it becomes apparent how critical this knowledge is in situations of disequilibrium between evolutionary forces. The theory and methods developed to study population genetics generally assume equilibrium among evolutionary forces. With respect to neutral markers, at the temporal and spatial scales usually studied in landscape genetics, this usually implies equilibrium between gene flow and genetic drift. But, in the context of habitat fragmentation, a usual theme in conservation, this is rarely the case. In many areas, fragmentation is relatively recent, having occurred in the past half century. For most of the large, charismatic megafauna that are the focus of many conservation efforts, this is a very short time period. Too short for a new balance to have been achieved, which means that most point observations of genetic data are probably a snapshot of a system in transition, which still has a deep signature of past rather than contemporary patterns of dispersal, population size, etc. Therefore, due to this temporal lag in genetic patterns, it is essential to interpret the observed genetic patterns in light of current and past landscape patterns.

Finally, little work has been undertaken to understand the effects of a single landscape on multiple species. In landscape ecology, there are a few studies of this nature with vertebrates (Nupp & Swihart, 2000; Uezu, 2007), but most are with *Eleodes* beetles (e.g. Crist *et al.*, 1992) or with other insects (e.g. Wiens *et al.*, 1995). The reason to undertake a comparative study is to understand underlying processes that may be common to more than one species. In the case of landscape ecology, it has been shown that many of the patterns associated with animal response to landscape structure are scalable. In other words, the relative differences observed for insects of differing body sizes should hold for larger animals, as long as the same parameters are measured at the

appropriate scale (Wiens, 1989). This dissertation is a step in the direction of showing that processes in population genetics are also scalable. Therefore, there may be much to gain in studying genetic differentiation in relation to landscape heterogeneity in smaller scaled systems, with views to applying this knowledge to conservation. In these smaller systems, there are none of the problems associated with sample sizes, or the quality of the DNA collected. The history is known, and the process of genetic differentiation could be studied through time, allowing for insights into the rates at which equilibrium is reached under different conditions, and how much disequilibrium affects our ability to infer the current genetic structure.

In the Pontal do Paranapanema

In the specific case of the *Pontal*, this dissertation adds to the body of knowledge that habitat fragmentation is affecting the local populations of large mammals. But, it suggests that fragmentation is acting at different scales in each species, and therefore different approaches may be needed, in order to effectively conserve the landscape and the species.

In the case of the herbivores, habitat corridors would be effective, but only as a first step in a larger plan to increase the area of contiguous habitat (which is reflected in the current action plan for the region). This conclusion is supported by at least three lines of evidence. First, white-lipped peccary results (coupled with preliminary telemetry data; A. Nava personal communication) suggest that this species is reluctant to cross the main road going through the PEMD. In addition, the fact that the species is reluctant to cross disturbed areas implies that it would probably not use a corridor between habitat patches, unless it was wide enough. Second, the collared peccary clustering (coupled with the

knowledge of the species' rigid social structure), suggests that even if animals do move from one fragment to another, they may not be able to enter a new group and reproduce. If, new groups are formed by the splitting of larger groups (as may have happened with the population of *Tucano* and *Ribeirão Bonito*), it would be best to have sufficient space for these new groups to colonize. As it is now, it is unclear that the smaller fragments have enough space to support more than one herd. And, finally, in the case of tapirs, it is not clear yet what is the exact effect of fragmentation. However, the small size of the majority of the forest fragments in the region suggests that the subpopulations that could be supported in them would be small, and might be prone to frequent extinction and re-colonization. The source of individuals for re-colonization would most probably be from the PEMD, and as such these fragments would act as a sink for genetic diversity. Unfortunately, there are no larger fragments close enough to which tapirs would be capable of dispersing, to which a corridor could be built, and therefore more than any other species, I think tapirs require larger areas of contiguous habitat to ensure their long-term sustainability in the region.

In the case of the carnivores, corridors at this scale may not be as useful either. Ocelots are clearly capable of moving through the existing landscape matrix. But, as in the tapirs, most of the habitat fragments may be too small to sustain populations large enough for the long-term persistence of the species in the region. And, again, there are no large fragments close enough to which the species could disperse and maintain an exchange of genetic diversity. In the case of the jaguars, the *Pontal*, should be seen as a stepping stone in a conservation action plan spanning a much larger area (corroborating the existing plan). The scale of the *Pontal* is too small to support a long-lasting jaguar

population, but the species should be able to disperse to larger fragments that are located farther away. However, it still remains unclear what the carrying capacity of these fragments may be, if the fragments are already saturated, and thus whether incoming animals would be able to establish and breed.

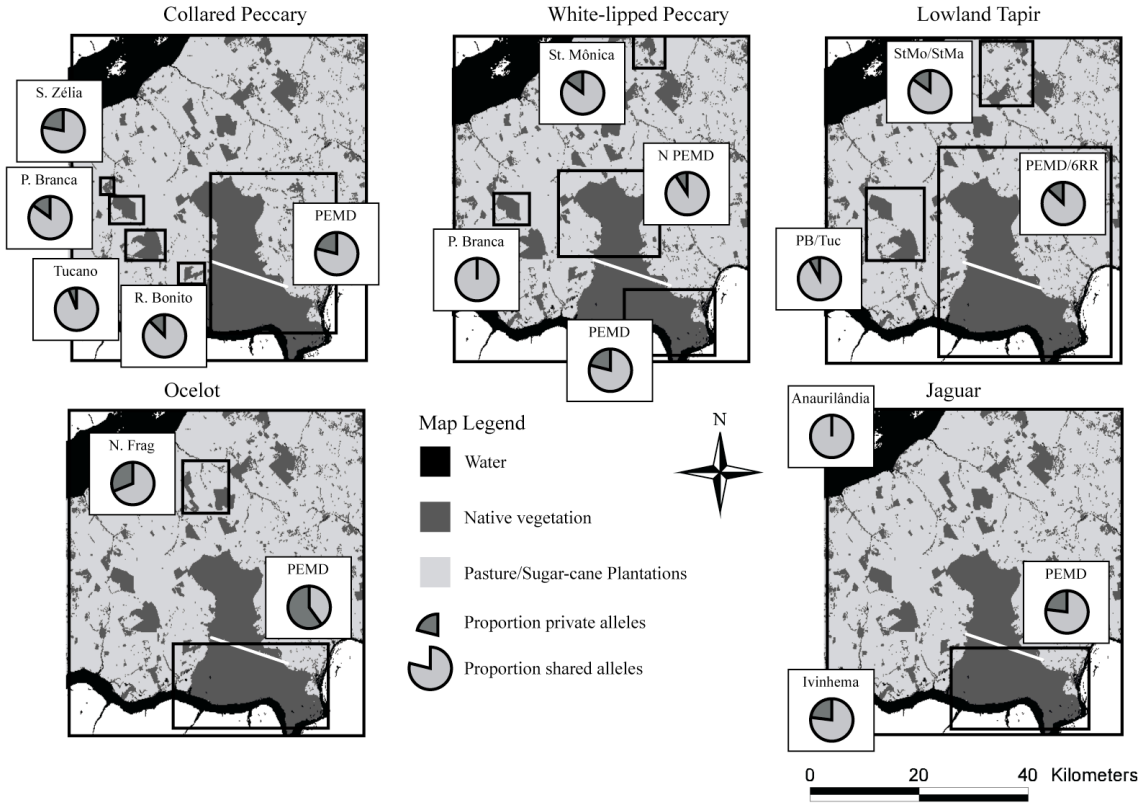

Figure 6.2 Map of private alleles in the *Pontal do Paranapanema* for the five examined species. Subpopulations were assigned according to results in Chapter 4. In the case of jaguars, the geographic locations of *Anaurilândia* and *Ivinhema* are not represented at this scale.

In terms of genetic diversity, the average observed heterozygosity is still relatively high in all species, with the exception of the ocelots, which seem to have a

significant overall deficit of heterozygotes. In addition, there are private alleles associated with the different fragments in each species (Figure 6.2). This suggests that some genetic variation is still retained in the extant subpopulations. Furthermore, the location of private alleles, if used as a criteria for prioritizing conservation, suggests that most sampled subpopulations, in each species, are crucial for the conservation of the landscape as whole. However, we have no measure of the rate of loss of genetic variation at the moment, but it is possible that because the process of fragmentation happened over a period of a few decades, the rate of loss of genetic variation will increase in the near future. Nevertheless, neutral genetic variation is not always correlated with adaptive genetic variation (Reed & Frankham, 2001), and it would be interesting to study adaptive variation in this landscape. In particular, MHC loci may prove to be quite interesting, given the close association of native herbivores with native livestock, and native carnivores with domesticated cats and dogs.

Final considerations

This study is one of the first to examine the comparative effect of fragmentation on several species of mammals within the same landscape. Despite its limitations, the results of this dissertation are still significant. There is an indication that body size and trophic level may be good correlates of spatial genetic structure in mammals. Also, there is an indication that scaling in spatial genetic structure is to be expected in natural systems, indicated by the similar effect of fragmentation at different scales for different species. However, it raises new questions, for instance, what characteristics both of the landscape and the organism that set the limits of the scale domain? Here, we saw that population

size might be a factor, but what about social structure, level of heterogeneity, and habitat specialization? This study also contributes to the body of knowledge showing that landscapes can have significant effects over the distribution of genetic variation. Future research, in this area, may include examining the effects of landscapes on adaptive variation, and the effects that dynamic features of the landscape, such as periodic flooding or pronounced seasonality, may have on the distribution of genetic variation. This study also makes a contribution to the use of Bayesian models to infer population structure. The simulations (Chapter 3) suggest that at least BAPS can be a powerful tool. Yet, further research is needed into the possible overriding effects of using informed priors relative to the amount of genetic information available. Finally, this study corroborates the notion that landscape genetics can produce meaningful information for conservation, provided that it is not the sole source of information, and that the genetic information is interpreted within an ecological and historical context.

Literature Cited

Alberts SC, Altmann J (1995) Balancing costs and opportunities: Dispersal in male baboons. *American Naturalist* **145**, 279-306.

Altrichter M, Boaglio GI (2004) Distribution and relative abundance of peccaries in the Argentine Chaco: associations with human factors. *Biological Conservation* **116**, 217-225.

Andreassen HP, Ims RA, Steinset O (1996) Discontinuous habitat corridors: Effects on male root vole movements. *Journal of Applied Ecology* **33**, 555-560.

Archibald AL, Haley CS, Brown JF*, et al.* (1995) The PiGMaP consortium linkage map of the pig (*Sus scrofa*). *Mammalian Genome* **6**, 157-175.

Balloux F (2001) EASYPOP (Version 1.7): A Computer Program for Population Genetics Simulations. *Journal of Heredity* **92**, 301-302.

Bandelt HJ, Lahermo P, Richards M, Macaulay V (2001) Detecting errors in mtDNA data by phylogenetic analysis. *International Journal of Legal Medicine* **115**, 64-69.

Bassi C (2004) *O efeito da fragmentação sobre a comunidade de mamíferos nas matas do Planalto Ocidental, Estado de São Paulo, Brasil*, Universidade de São Paulo.

Bellantoni ES, Krausman PR (1993) Habitat use by collared peccaries in an urban-environment. *Southwestern Naturalist* **38**, 345-351.

Bellemain EVA, Swenson JE, Tallmon D, Brunberg S, Taberlet P (2005) Estimating population size of elusive animals with DNA from hunter-collected feces: Four methods for Brown Bears. *Conservation Biology* **19**, 150-161.

Birch DE, Kolmodin L, Wong J*, et al.* (1996) Simplified hot start PCR. *Nature* **381**, 445-446.

Bowling AT, Eggleston-Stott ML, Byrns G*, et al.* (1997) Validation of microsatellite markers for routine horse parentage testing. *Animal Genetics* **28**, 247-252.

Broquet T, Ménard N, 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.

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

Broquet T, Ray N, Petit E, Fryxell J, Burel Fo (2006) Genetic isolation by distance and landscape connectivity in the American marten (*Martes americana*). *Landscape Ecology* **21**, 877-889.

Brouat C, Sennedot F, Audiot P, Leblois R, Rasplus JY (2003) Fine-scale genetic structure of two carabid species with contrasted levels of habitat specialization. *Molecular Ecology* **12**, 1731-1745.

Brown JH, Lomolino MV (1998) *Biogeography*, 2nd edn. Sinauer Associates, Sunderland, Mass.

Brown JH, West GB (2000) *Scaling in biology* Oxford University Press, New York.

Byers JA, Bekoff M (1981) Social, spacing, and cooperative behavior of the collared peccary, *Tayassu tajacu*. *Journal of Mammalogy* **62**, 767-785.

Calabrese JM, Fagan WF (2004) A comparison-shopper's guide to connectivity metrics. *Frontiers in Ecology and the Environment* **2**, 529-536.

Carrillo E, Saenz JC, Fuller TK (2002) Movements and activities of white-lipped peccaries in Corcovado National Park, Costa Rica. *Biological Conservation* **108**, 317-324.

Chou Q, Russell M, Birch DE, Raymond J, Bloch W (1992) Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Research* **20**, 17-1723.

Cleveland WS (1993) *Visualizing Data* Hobart Press, Murray Hill, NJ.

Cleveland WS, Devlin SJ (1988) Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* **83**, 596-610.

Coltman DW, Pilkington JG, Pemberton JM (2003) Fine-scale genetic structure in a free-living ungulate population. *Molecular Ecology* **12**, 733-742.

Constable JL, Ashley MV, Goodall J, Pusey AE (2001) Noninvasive paternity assignment in Gombe chimpanzees. *Molecular Ecology* **10**, 1279-1300.

Corander J (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* **20**, 2363-2369.

Corander J, Marttinen P (2006) Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology* **15**, 2833-2843.

Corander J, Marttinen P, Sirén J, Tang J (2006) *BAPS: Bayesian Analysis of Population Structure. Manual v. 4.14* University of Helsinki.

Corander J, Sirén J, Arjas E (in press) Spatial modeling of genetic population structure. *Computational Statistics*.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* **163**, 367-374.

Coulon A, Cosson JF, Angibault JM*, et al.* (2004) Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: An individual-based approach. *Molecular Ecology* **13**, 2841-2850.

Coulon A, Guillot G, Cosson JF*, et al.* (2006) Genetic structure is influenced by landscape features: Empirical evidence from a roe deer population. *Molecular Ecology* **15**, 1669-1679.

Crawshaw PG, Quigley HB (1989) Notes on ocelot movement and activity in the Pantanal region, Brazil. *Biotropica* **21**, 377-379.

Crist TO, Guertin DS, Wiens JA, Milne BT (1992) Animal movement in heterogeneous landscapes: An experiment with *Eleodes* beetles in shortgrass prairie. *Functional Ecology* **6**, 536-544.

Crockett CM, Pope TR (1993) Consequences of sex differences in dispersal for juvenile red howler monkeys. In: *Juvenile primates: life history, development, and behavior* (eds. Pereira ME, Fairbanks LA), pp. 104-118. Oxford University Press, New York.

Crow JF, Kimura M (1970) *An introduction to population genetics* Harper & Row Publishers, New York, NY.

Dálen L, Kvaloy K, Linnell JDC*, et al.* (2006) Population structure in a critically endangered arctic fox population: Does genetics matter? *Molecular Ecology* **15**, 2809-2819.

Dallas JF, Marshall F, Piertney SB, Bacon PJ, Racey PA (2002) Spatially restricted gene flow and reduced microsatellite polymorphism in the Eurasian otter *Lutra lutra* in Britain. *Conservation Genetics* **3**, 15-29.

Dawson KJ, Belkhir K (2001) A Bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genetical Research* **78**, 59-77.

D'Eon R, Glenn SM, Parfitt I, Fortin MJ (2002) Landscape connectivity as a function of scale and organism vagility in a real forested landscape. *Conservation Ecology* **6**, 10. [online] URL: http://www.consecol.org/vol16/iss12/art10/

Ditt EH (2002) *Fragmentos florestais no Pontal do Paranapanema* Annablume Editora, São Paulo, SP.

Dobzhansky T (1937) *Genetics and the Origins of Species.* Columbia University Press, New York, NY.

Dobzhansky T, Wright S (1941) Genetics of natural populations. V. Relations between mutation rate and accumulation of lethals in populations of *Drosophila pseudoobscura*. *Genetics* **26**, 23-51.

Dubost G (2001) Comparison of the social behaviour of captive sympatric peccary species (genus *Tayassu*): Correlations with their ecological characteristics. *Mammalian Biology* **66**, 65-83.

Eizirik E, Bonatto SL, Johnson WE*, et al.* (1998) Phylogeographic patterns and evolution of the mitochondrial DNA control region in two Neotropical cats (Mammalia, Felidae). *Journal of Molecular Evolution* **47**, 613-624.

Eizirik E, Kim JH, Menotti-Raymond M*, et al.* (2001) Phylogeography, population history and conservation genetics of jaguars (*Panthera onca*, Mammalia, Felidae). *Molecular Ecology* **10**, 65-79.

Endler JA (1973) Gene flow and population differentiation: Studies of clines suggest that differentiation along environmental gradients may be independent of gene flow. *Science* **179**, 243.

Epperson BK (2003) *Geographical Genetics* Princeton University Press, Princeton, NJ.

Epps CW, Palsboll PJ, Wehausen JD*, et al.* (2005) Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. *Ecology Letters* **8**, 1029-1038.

Ernest HB, Boyce WM, Bleich VC*, et al.* (2003) Genetic structure of mountain lion (*Puma concolor*) populations in California. *Conservation Genetics* **4**, 353-366.

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

Excoffier L, Heckel G, Box P, Contact NPG (2006) Computer programs for population genetics data analysis: A survival guide. *Nature Reviews Genetics* **7**, 745-758.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

Farrell LE, Roman J, Sunquist ME (2000) Dietary separation of sympatric carnivores identified by molecular analysis of scats. *Molecular Ecology* **9**, 1583-1590.

Felsenstein J (1975) A pain in the torus: Some difficulties with models of isolation by distance. *The American Naturalist* **109**, 359-368.

Fernando P, Vidya TNC, Rajapakse C, Dangolla A, Melnick DJ (2003a) Reliable noninvasive genotyping: Fantasy or reality? *Journal of Heredity* **94**, 115-123.

Fernando P, Vidya TNC, Payne J*, et al.* (2003b) DNA analysis indicates that Asian elephants are native to Borneo and are therefore a high priority for conservation. *Plos Biology* **1**, 110-115.

Ferreira ME, Grattapaglia D (1998) *Introdução ao uso de marcadores moleculares em análise genética* EMBRAPA, Brasília.

Fiorello CV, Noss AJ, Deem SL (2006) Demography, hunting ecology, and pathogen exposure of domestic dogs in the Isoso of Bolivia. *Conservation Biology* **20**, 762-771.

Fisher RA (1930) *The genetical theory of natural selection.* Clarendon Press, Oxford, UK.

Fragoso JMV (1998) Home range and movement patterns of white-lipped peccary (*Tayassu pecari*) herds in the northern Brazilian Amazon. *Biotropica* **30**, 458-469.

Fragoso JMV (1999) Perception of scale and resource partitioning by peccaries: Behavioral causes and ecological implications. *Journal of Mammalogy* **80**, 993-1003.

Frantz AC, Pope LC, Carpenter PJ*, et al.* (2003) Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Molecular Ecology* **12**, 1649-1661.

Garnier JN, Bruford MW, Goossens B (2001) Mating system and reproductive skew in the black rhinoceros. *Molecular Ecology* **10**, 2031-2041.

Gerloff U, Schloetterer C, Rassmann K*, et al.* (1995) Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild living bonobos (*Pan paniscus*). *Molecular Ecology* **4**, 515-518.

Goossens B, Chikhi L, Jalil MF*, et al.* (2005) Patterns of genetic diversity and migration in increasingly fragmented and declining orang-utan (*Pongo pygmaeus*) populations from Sabah, Malaysia. *Molecular Ecology* **14**, 441-456.

Greenwood PJ (1980) Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour* **28**, 1140-1162.

Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics* **170**, 1261-1280.

Guillot G, Mortier F, Estoup A (2005b) GENELAND: A program for landscape genetics *Molecular Ecology Notes* **5**, 712-715.

Hájková P, Zemanová B, Bryja J*, et al.* (2006) Factors affecting success of PCR amplification of microsatellite loci from otter faeces. *Molecular Ecology Notes* **6**, 599-562.

Haldane JBS (1932) *The causes of evolution* Longman Press, New York, NY, USA.

Haldane JBS (1948) The theory of a cline. *Journal of Genetics* **48**, 277–284.

Hall LS, Grigg GG, Moritz C*, et al.* (2004) Biogeography of fruit eating bats in Southeast Asia. *Sarawak Museum Journal* **LX**, 191-284.

Hanski I (1998) Metapopulation dynamics. *Nature* **396**, 41-49.

Hardy OJ, Vekemans X (2002) SPAGEDi: A versatile computer program to analyze spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618-620.

Hardy OJ, Vekemans X (2006) *SPAGeDi 1.2: A program for the spatial pattern analysis of genetic diversity: User's manual*, Bruxelles, Belgium.

Harestad AS, Bunnel FL (1979) Home range and body weight – A re-evaluation. *Ecology* **60**, 389-402.

Hartl DL, Clark AG (1997) *Principles of Population Genetics*, 3rd edn. Sinauer Associates, Inc., Sunderland, MA.

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

Holzhauer S, Ekschmitt K, Sander A-C, Dauber J, Wolters V (2006) Effect of historic landscape change on the genetic structure of the bush-cricket *Metrioptera roeseli*. *Landscape Ecology* **21**, 891-899.

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* **53**, 1898-1914.

Huxley JS (1938) Clines: An auxiliary taxonomic principle. *Nature* **142**, 219.

Huxley JS (1942) *Evolution: The modern synthesis* Harper and Bros., New York, NY.

Iyengar A, Babu VN, Hedges S*, et al.* (2005) Phylogeography, genetic structure, and diversity in the dhole (*Cuon alpinus*). *Molecular Ecology* **14**, 2281-2297.

Jacob AA (2002) *Ecologia e conservação da jaguatirica (Leopardus pardalis) no Parque Estadual Morro do Diabo, Pontal do Paranapanema, SP* Masters Thesis, Universidade de Brasilia.

Judas J, Henry O (1999) Seasonal variation of home range of collared peccary in tropical rain forests of French Guiana. *Journal of Wildlife Management* **63**, 546-552.

Keuroghlian A, Eaton DP, Longland WS (2004) Area use by white-lipped and collared peccaries (*Tayassu pecari* and *Tayassu tajacu*) in a tropical forest fragment. *Biological Conservation* **120**, 411-425.

Keyghobadi N, Roland J, Strobeck C (1999) Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology* **8**, 1481-1495.

Kimura M (1953) Stepping stone model of population. *Annual Report of the National Institute of Genetics* **3**, 62–63.

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561-576.

Kreader CA (1996) Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Applied and Environmental Microbiology* **62**, 1102-1106.

Kurose N, Masuda R, Tatara M (2005) Fecal DNA analysis for identifying species and sex of sympatric carnivores: A noninvasive method for conservation on the Tsushima Islands, Japan. *Journal of Heredity* **96**, 688-697.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics* **7**, 295-302.

Lawson Handley LJ, Perrin N (2007) Advances in our understanding of mammalian sex-biased dispersal. *Molecular Ecology* **16**, 1559-1578.

Li H, Reynolds JF (1995) On definition and quantification of heterogeneity. *Oikos* **73**, 280-284.

Lidicker, Jr., W. Z., Patton JL (1987) Patterns of dispersal and genetic structure in populations of small rodents. In: *Mammalian Dispersal Patterns* (eds. Chepko-Sade BD, Halpin ZT), pp. 144–161. The Chicago University Press, Chicago, Il.

Lorenzini R, Lovari S (2006) Genetic diversity and phylogeography of the European roe deer: The refuge area theory revisited. *Biological Journal of the Linnean Society* **88**, 85-100.

Lukas D, Bradley BJ, Nsubuga AM*, et al.* (2004) Major histocompatibility complex and microsatellite variation in two populations of wild gorillas. *Molecular Ecology* **13**, 3389-3402.

Macdonald DW (1995) The Encyclopedia of Mammals, p. 895. Facts on File, Inc., New York, NY.

Malécot G (1948) *Les mathématiques de l'hérédité.* Masson.

Malécot G (1968) *The mathematics of heredity.* Freeman, San Francisco, CA.

Mandelbrot BB (1982) *The fractal geometry of nature* Freeman, San Francisco, CA.

Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: Matching biological questions with appropriate techniques. *Trends in Ecology & Evolution* **20**, 136-142.

Manel S, Schwarts MK, Luikart G, Taberlet P (2003) Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.

Mank JE, Avise JC (2005) Individual organisms as units of analysis: Bayesian-clustering alternatives in population genetics. *Genetical Research* **84**, 135-143.

Maruyama T (1972) Rate of decrease of genetic variability in a two-dimensional continuous population of finite size. *Genetics* **70**, 639-651.

Mathews NE, Porter WF (1993) Effect of social structure on genetic structure of free-ranging white-tailed deer in the Adirondack Mountains. *Journal of Mammalogy* **74**, 33-43.

Mayer AL, Petren K, Shelton A*, et al.* (2002) Scaling natal dispersal distances: Confounding factors. *Conservation Ecology* **6**, r8. [online] URL: http://www.consecol.org/vol6/iss1/resp8/

Mayr E (1942) *Systematics and the Origin of Species* Columbia University Press, New York, NY.

McCracken GF, Bradbury JW (1977) Paternity and genetic heterogeneity in the polygynous bat, *Phyllostomus hastatus. Science* **198**, 303–306.

McIntyre NE, Wiens JA (1999) Interactions between landscape structure and animal behavior: The roles of heterogeneously distributed resources and food deprivation on movement patterns. *Landscape Ecology* **14**, 437-447.

Melnick DJ (1987) The genetic consequences of primate social organization: A review of macaques, baboons and vervet monkeys. *Genetica* **73**, 117-135.

Melnick DJ, Hoelzer GA (1992) Differences in male and female macaque dispersal lead to contrasting distributions of nuclear and mitochondrial-DNA variation. *International Journal of Primatology* **13**, 379-393.

Menotti-Raymond M, David VA, Lyons LA*, et al.* (1999) A genetic linkage map of microsatellites in the Domestic Cat (*Felis catus*). *Genomics* **57**, 9-23.

Merriam G (1984) Connectivity: a fundamental ecological characteristic of landscape pattern *Proceedings of the First International Seminar on Methodology in*

Landscape Ecological Research and Planning. Theme I: Landscape Ecological Concepts **1**, 5-16.

Moran PAP (1950) Notes on continuous stochastic phenomena. *Biometrika* **37**, 17-23.

Morellato PC (2000) Introduction: The Brazilian Atlantic forest. *Biotropica* **32**, 786-792.

Morin PA, Chambers KE, Boesch C, 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.

Nagylaki T (1989) Gustave Malécot and the transition from classical to modern population genetics. *Genetics* **122**, 253-268.

Norton JE, Ashley MV (2004) Genetic variability and population genetic structure among Bairds tapirs. *Animal Conservation* **7**, 211-220.

Nsubuga AM, Robbins MM, Roeder AD*, et al.* (2004) Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. *Molecular Ecology* **13**, 2089-2094.

Nupp TE, Swihart RK (2000) Landscape-level correlates of small-mammal assemblages in forest fragments of farmland. *Journal of Mammalogy* **81**, 512-526.

Olden JD, Schooley RL, Monroe JB, Poff NL (2004) Context-dependent perceptual ranges and their relevance to animal movements in landscapes. *Journal of Animal Ecology* **73**, 1190-1194.

Pääbo S, Gifford JA, Wilson AC (1988) Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acids Research* **16**, 9775-9787.

Padilla M, Dowler RC (1994) Tapirus terrestris. *Mammalian Species* **481**, 1-8.

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

Palomares F, Godoy JA, Piriz A, O'Brien SJ, Johnson WE (2002) Fecal genetic analysis to determine the presence and distribution of elusive carnivores: Design and feasibility for the Iberian lynx. *Molecular Ecology* **11**, 2171-2182.

Parsons KM, Dallas JF, Claridge DE*, et al.* (1999) Amplifying dolphin mitochondrial DNA from fecal plumes. *Molecular Ecology* **8**, 1766-1768.

Peakall R, Ruibal M, Lindenmayer DB (2003) Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, *Rattus fuscipes*. *Evolution* **57**, 1182-1195.

Pearse DE, Crandall KA (2004) Beyond *FST*: Analysis of population genetic data for conservation. *Conservation Genetics* **5**, 585-602.

Peterson MA, Denno RF (1998) The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. *American Naturalist* **152**, 428-446.

Piertney SB, MacColl ADC, Bacon PJ, Dallas JF (1998) Local genetic structure in red grouse (*Lagopus lagopus scoticus*): Evidence from microsatellite DNA markers. *Molecular Ecology* **7**, 1645-1654.

Piggott MP, Banks SC, Taylor AC (2006) Population structure of brush-tailed rock-wallaby (*Petrogale penicillata*) colonies inferred from analysis of faecal DNA. *Molecular Ecology* **15**, 93-105.

Piggott MP, Bellemain E, Taberlet P, Taylor AC (2004) A multiplex pre-amplification method that significantly improves microsatellite amplification and error rates for faecal DNA in limiting conditions. *Conservation Genetics* **5**, 417-420.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics* **6**, 847-846.

Pope TR (1992) The influence of dispersal patterns and mating system on genetic differentiation within and between populations of the red howler monkey (*Alouatta seniculus*). *Evolution* **46**, 1112-1128.

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

Pritchard JK, Wen W (2004) Documentation for STRUCTURE software: Version 2. *Department of Human Genetics, University of Chicago, Chicago (http://pritch. bsd. uchicago. edu)*.

Prugh LR, Ritland CE, Arthur SM, Krebs CJ (2005) Monitoring coyote population dynamics by genotyping faeces. *Molecular Ecology* **14**, 1585-1596.

R Development Core Team (2005) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ray N (2005) PATHMATRIX: A geographical information system tool to compute effective distances among samples. *Molecular Ecology Notes* **5**, 177-180.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): Population genetics software for the exact tests and ecumenicism. *Journal of Heredity* **86**, 248.

Reed DH, Frankham R (2001) How closely related are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* **55**, 1095-1103.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223-225.

Rosenbaum HC, Weinrich MT, Stoleson SA*, et al.* (2002) The effect of differential reproductive success on population genetic structure: Correlations of life history with matrilines in humpback whales of the Gulf of Maine. *Journal of Heredity* **93**, 389-399.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**, 58-62.

Russello MA, Gladyshev E, Miquelle D, Caccone A (2004) Potential genetic consequences of a recent bottleneck in the Amur tiger of the Russian far east. *Conservation Genetics* **5**, 707-713.

Sawyer S, Felsenstein J (1983) Isolation by distance in a hierarchically clustered population. *Journal of Applied Probability* **20**, 1-10.

Scandura M (2004) *The use of microsatellites in the study of social stracture in large mammals: Italian wolf and fallow deer as case studies*, University of Bielefeld.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN ver. 2001: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Schooley RL, Wiens JA (2003) Finding habitat patches and directional connectivity. *Oikos* **102**, 559-570.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**, 615-629.

Seutin G, White BN, Boag PT (1991) Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology* **69**, 82-90.

Shrader AM, Owen-Smith N (2002) The role of companionship in the dispersal of white rhinoceroses (*Ceratotherium simum*). *Behavioral Ecology and Sociobiology* **52**, 255-261.

Slatkin M (1973) Gene flow and selection in a cline. *Genetics* **75**, 733-756.

Slatkin M (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* **47**, 264-279.

Slatkin M, Maruyama T (1975) Genetic drift in a cline. *Genetics* **81**, 209-222.

Sokal RR, Wartenberg DE (1983) A test of spatial autocorrelation analysis using an isolation-by-distance model. *Genetics* **105**, 219-237.

Storfer A, Murphy MA, Evans JS*, et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity* **98**, 128-142.

Storz JF (1999) Genetic consequences of mammalian social structure. *Journal of Mammalogy* **80**, 553-569.

Stow AJ, Sunnucks P, Briscoe DA, Gardner MG (2001) The impact of habitat fragmentation on dispersal of Cunningham's skink (*Egernia cunninghami*): Evidence from allelic and enotypic analyses of microsatellites. *Molecular Ecology* **10**, 867-878.

Strahler AN, Strahler AH (1989) *Elements of physical geography*, 4th edn. John Wiley, New York, NY.

Sugg DW, Chesser RK, Stephen Dobson F, Hoogland JL (1996) Population genetics meets behavioral ecology. *Trends in Ecology & Evolution* **11**, 338-342.

Sumner J, Jessop T, Paetkau D, Moritz C (2004) Limited effect of anthropogenic habitat fragmentation on molecular diversity in a rain forest skink, *Gnypetoscincus queenslandiae*. *Molecular Ecology* **13**, 259-269.

Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology & Evolution* **15**, 199-203.

Surridge AK, Suárez SS, Buchanan-Smith HM, Smith AC, Mundy NI (2005) Color vision pigment frequencies in wild tamarins (*Saguinus* sp.). *American Journal of Primatology* **67**, 463-470.

Sutherland GD, Harestad AS, Price K, Lertzman KP (2000) Scaling of natal dispersal distances in terrestrial birds and mammals. *Conservation Ecology* **4**, 16. [online] URL: http://www.consecol.org/vol14/iss11/art16/

Swofford DL (2002) PAUP: Phylogenetic analysis using parsimony and other methods. 4.0 beta. Sunderland, MA: Sinauer Associates.

Taberlet P, Camarra JJ, Griffin S*, et al.* (1997) Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology* **6**, 869-876.

Taberlet P, Griffin S, Goossens B*, et al.* (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**, 3189-3194.

Taberlet P, Luikart G (1996) Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society* **68**, 41-55.

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

Turner MG (1989) Landscape ecology: The effect of pattern on process. *Annual Review of Ecology and Systematics* **20**, 171-197.

Uezu A (2007) *Composição e estrutura da comunidade de aves na paisagem fragmentada do Pontal do Paranapanema*, Universidade de São Paulo.

Valladares-Padua C, Cullen, Jr., L., Padua SM*, et al.* (1997) Resgatando a Grande Reserva do Pontal do Paranapanema: Reforma agrária e conservação de biodiversidade. In: *Congresso Brasileiro de Unidades de Conservação*, pp. 783-791. Universidade Livre do Meio Ambiente, Rede Nacional Pró Unidades de Conservação, Instituto Ambiental do Paraná, Curitiba, PA.

Vasemägi A (2006) The adaptive hypothesis of clinal variation revisited: Single-locus clines as a result of spatially restricted gene flow. *Genetics* **173**, 2411-2414.

Verbeylen G, De Bruyn L, Adriaensen F, Matthysen E (2003) Does matrix resistance influence red squirrel (*Sciurus vulgaris* L. 1758) distribution in an urban landscape? *Landscape Ecology* **18**, 791-805.

Vigilant L (2002) Technical challenges in the microsatellite genotyping of a wild chimpanzee population using feces. *Evolutionary Anthropology: Issues, News, and Reviews* **11**, 162-165.

Wasser SK, Davenport B, Ramage ER*, et al.* (2004) Scat detection dogs in wildlife research and management: Application to grizzly and black bears in the Yellowhead Ecosystem, Alberta, Canada. *Canadian Journal of Zoology/Revue Canadienne de Zoologie* **82**, 475-492.

Wasser SK, Houston CS, Koehler GM, Cadd GG, Fain SR (1997) Techniques for application of faecal DNA methods to field studies of Ursids. *Molecular Ecology* **6**, 1091-1097.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Weiss GH, Kimura M (1965) A mathematical analysis of the stepping stone model of genetic correlation. *Journal of Applied Probability* **2**, 129-149.

Wiens JA (1989) Spatial scaling in ecology. *Functional Ecology* **3**, 385-397.

Wiens JA, Crist TO, With KA, Milne BT (1995) Fractal patterns of insect movement in microlandscape mosaics. *Ecology* **76**, 663-666.

Wiens JA, Schooley RL, Weeks RD (1997) Patchy landscapes and animal movements: Do beetles percolate? *Oikos* **78**, 257-264.

With KA, Crist TO (1995) Critical thresholds in species responses to landscape structure. *Ecology* **76**, 2446-2459.

With KA, King AW (1999) Dispersal success on fractal landscapes: A consequence of lacunarity thresholds. *Landscape Ecology* **14**, 73-82.

Wolfram Research, Inc. (2004) MATHEMATICA Edition: Version 5.1. Wolfram Research, Inc., Champaign, IL.

Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**, 97-159.

Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.

Yao Y-G, Bravi CM, Bandelt H-J (2004) A call for mtDNA data quality control in forensic science. *Forensic Science International* **141**, 1-6.

Zollner PA, Lima SL (1997) Landscape-level perceptual abilities in white-footed mice: Perceptual range and the detection of forested habitat. *Oikos* **80**, 51-60.