Camila Clozato Lara

Estrutura populacional em tamanduá-mirim (*Tamandua tetradactyla* Linnaeus, 1758): variação molecular em regiões genômicas neutras e sob-seleção.



São Paulo

2014

Estrutura populacional em tamanduá-mirim (*Tamandua tetradactyla* Linnaeus, 1758): variação molecular em regiões genômicas neutras e sob-seleção.

Tese apresentada ao Instituto de Biociências da Universidade de São Paulo, como requisito parcial para a obtenção do título de Doutor em Biologia, Genética.

Orientador:

Prof. Dr. João Stenguel Morgante Co-orientadores:

Dra. Nadia de Moraes-Barros

Prof. Dra. Simone Sommer

São Paulo

2014

Clozato, Camila L. Estrutura populacional em tamanduá-mirim (*Tamandua tetradactyla* Linnaeus, 1758): variação molecular em regiões genômicas neutras e sob-seleção. 175 pp.

Tese de Doutorado - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. Estruturação genética 2. Microssatélites 3. Complexo Principal de Histocompatibilidade 4. *Tamandua tetradactyla* I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a). Orientador(a)

Agradecimentos

Foram quatro anos de muitas mudanças. Agradeço ao doutorado a profunda transformação que sofri como pessoa. Pela primeira vez morei sozinha, aprendi a viver sem a família que me acompanhava até então, e que se tornou, seja pela distância ou pela saudade, cada vez mais querida.

Aos poucos, me tornei mais crítica, mais consciente do mundo ao meu redor, e mais grata pela vida que eu tive a oportunidade de experimentar. Em algum grau, tudo isso se relaciona com o doutorado, que direta e indiretamente me expôs a diversas questões. Eu nunca soube, por exemplo, que a pizza de São Paulo era tão melhor assim que a de Minas Gerais! Mas mantenho minha opinião sobre o pão de queijo.

Tenho, portanto, muitos agradecimentos a fazer. O primeiro, ao meu orientador Prof. Dr. João Morgante, que me recebeu em seu laboratório tão gentilmente, e me permitiu compartilhar de seu vasto conhecimento como pesquisador e de sua experiência de vida.

Agradeço às co-orientadoras Dra. Nadia Moraes-Barros e Prof. Dra. Simone Sommer que auxiliaram muito na elaboração das ideias e na discussão formal dos trabalhos, melhorando significativamente a qualidade dos resultados apresentados.

O trabalho não teria sido possível sem a colaboração de instituições e pesquisadores que cederam amostras ou auxiliaram nos trabalhos de campo. Agradeço às equipes dos CETAS no nordeste (BA, AL, PE, PB, CE), ao biólogo Anderson Feijó da Universidade Federal da Paraíba, à equipe do Ecopoint em Fortaleza, a Prof. Dra. Flávia Barros da Universidade Federal do Alagoas, à equipe da Pousada Aguapé no Pantanal, a Claudia Madella, ao Prof. Dr. Eduardo Eizirik e Henrique Figueiró da Pontíficia Universidade Católica do Rio Grande do Sul, à Dra. Gilda Tebet do Museu de Historia Natural Capão da Imbuia em Curitiba, ao Dr. Hélio Boudet do Museu de Biologia Mello Leitão em Santa Teresa, ao Dr. Marcelo Graipel da Universidade Federal de Santa Catarina, Stella Franco e Dr. João Oliveira do Museu Nacional do Rio de Janeiro, ao Dr. Mário de Vivo do Museu de Zoologia da USP, ao Dr. Fabrício Santos da Universidade Federal de Minas Gerais. Agradeço à Flávia Miranda e José Abílio Ohana não só pelas valiosas amostras, mas também pelas frutíferas discussões e profundo interesse compartilhado pelo estudo do tamanduá-mirim.

Agradeço à bolsa de doutorado concedida pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), e pela bolsa do Programa de Doutorado Sanduíche no Exterior (PDSE) realizado no Instituto de Pesquisa em Zoologia e Vida Selvagem (IZW) de Berlim, Alemanha. Agradeço aos amigos do LABEC que me ajudaram em diversos momentos: Rodrigo Francisco, Maria Helena Maia, Anna Carolina Milo, Helena Tadiello, Márcia Pinceratti, Juliana Ferreira, Ana Pavan, Gisele Dantas, Bárbara Bitarello, Gabriela Cardoso, e Sofia Silva, que além de amiga, se tornou mais uma orientadora, me estimulando grandemente quando eu mais precisei.

Agradeço à minha família pelo apoio, carinho e suporte nas horas de neurose! Dona Maria José e Sr. Ermelindo (pra sempre), Renata (alguns conhecem como Vânia), Júlia, Maria Izabel, Marcelo, Rose, Pedro e Wilian, sou grata pela companhia de vocês ao longo desses anos.

Em 2009, na minha dissertação de mestrado, agradeci o apoio e companheirismo do meu então namorado, Caetano Oliveira. Hoje, após uma longa jornada juntos, entre tantos compromissos, viagens, despedidas, mudanças e reencontros, nos tornamos uma família. Embora seja hoje meu marido, eu o agradeço por ser meu eterno namorado, e sou grata pelo mesmo companheirismo e apoio de quando começamos a caminhar juntos, há quase dez anos atrás.

Índice

| Resumo | Página 7 |
|--|------------|
| Abstract | Página 9 |
| Introdução Geral | Página 11 |
| Objetivos | Página 30 |
| <u>Capítulo 1</u> - Amostras históricas e não-invasivas: um estudo de caso de erros de genotipagem em microssatélites isolados <i>de novo</i> para o tamanduá-mirim (<i>Tamandua tetradactyla</i> L., Pilosa) | Página 33 |
| <u>Capítulo 2</u> - Diversidade genética e estrutura populacional de um mamífero amplamente distribuído da América do Sul, o tamanduá- mirim (<i>Tamandua tetradactyla</i> , Pilosa) | Página 60 |
| Capítulo 3 - Diversidade de sequências e sinais de seleção no gene DRB de MHC Classe II do tamanduá mirim (<i>Tamandua tetradactyla</i>) ao longo de diferentes biomas. | Página 101 |
| Conclusão Geral | Página 141 |
| Anexos | Página 146 |

Resumo

Este trabalho teve como objetivo principal descrever a diversidade genética e identificar a estrutura populacional de populações de tamanduá-mirim distribuídas ao longo dos biomas brasileiros através do uso de ferramentas de genética de populações e do acesso a regiões neutras e adaptativas do genoma da espécie. A amostragem de indivíduos de tamanduá-mirim é complicada pela difícil detecção do animal em trabalhos de campo, pela sua complicada captura e manipulação. Assim, fez-se necessário o uso de espécimes de museu e de amostras não-invasivas. A fim de validar o uso das mesmas para a genotipagem confiável de oito locos de microssatélites desenvolvidos para a espécie foi utilizado um método de padronização da qualidade das amostras (Índice de Qualidade, QI) e estimativa dos erros de genotipagem. Foi observada uma qualidade superior das amostras não invasivas (N=19) em relação às peles de museu (N=138). Foi possível também eliminar amostras com desempenho ruim (QI<0.7 e sucesso de amplificação maior que 75%), e garantir a confiabilidade dos resultados de microssatélites das amostras que permaneceram no estudo para análises posteriores.

Devido à grande área de distribuição da espécie de forma contínua, se tornou complicado traçar populações pré-definidas. Assim, a abordagem da genética da paisagem foi a ferramenta mais apropriada para o estudo da estrutura de populações em *T. tetradactyla* (N=176). Comparativamente, duas abordagens foram usadas: agrupamento de indivíduos pelo critério de proximidade geográfica, designado aqui como *a priori* (20 populações), e análises baseadas no indivíduo, sem informação prévia de agrupamento populações), referida no capítulo como *a posteriori* (quatro transectos testados). Foram encontrados níveis de diversidade moderados (média de 11.38 alelos por lóco) e poucas

evidências de estruturação entre populações das localidades amostradas (K=2 na maioria dos testes). Os indivíduos que mostraram maior diferenciação foram originados da Floresta Amazônica. Esta região também demonstrou maior diversidade genética (riqueza alélica e heterozigosidade esperada) que as outras. Por outro lado, populações distribuídas ao longo da Mata Atlântica e regiões adjacentes demonstraram um padrão de isolamento por distância. Populações do Brasil central (Cerrado e Pantanal) não demonstraram diferenciação em relação às demais.

Finalmente, foi estudada a variação genética adaptativa da espécie através da diversidade do gene DRB do complexo MHC. Esta família gênica codifica proteínas envolvidas no reconhecimento de antígeno e ativação da resposta imune adaptativa, e são regulados por seleção natural, especialmente por pressão seletiva dirigida por patógenos. É esperado que a diversidade de patógenos seja distinta nos biomas brasileiros, sendo os ambientes florestais mais biodiversos neste quesito do que ambientes da Diagonal Seca, o que representa pressões seletivas diferentes. Assim, foi investigada a diversidade do gene DRB éxon 2 em indivíduos (N=65) dos diferentes biomas brasileiros através de sequenciamento de nova geração (454 GS Junior), e os resultados de distribuição dos alelos foram comparados com os microssatélites. Foi encontrada uma alta diversidade (60 alelos no nível de aminoácido e 70 alelos no nível de nucleotídeos) e assinaturas claras de seleção positiva no gene (dN/dS=2.94). Maior riqueza alélica e proporção de alelos privados foram encontradas em biomas florestados, especialmente na Floresta Amazônica. Além disso, os marcadores neutros (microssatélites), demonstraram padrões similares ao DRB, revelando a força de eventos demográficos e deriva genética que também moldaram os padrões de diversidade deste gene do MHC.

Abstract

This work aimed to describe the genetic diversity and population structure of populations of the lesser anteater distributed along Brazilian biomes through population genetic tools, accessing neutral and adaptive genomic regions of the species.

Sampling lesser anteater individuals is complicated due to infrequent detection of the animal in field works, its complicated capture and manipulation. Thus, it was necessary to use museum specimens and noninvasive samples. To validate these samples for the reliable genotyping of eight microsatellite loci developed for the species, a standardization method of the quality of samples (Quality Index, QI) and genotyping errors was used. A superior quality of noninvasive samples (N=19) compared to study skins (N=138) was observed. It was also possible to eliminate samples with a bad performance (QI<0.7 and amplification success higher than 75%), and thus guarantee the reliability of microsatellites results for samples kept in further analysis.

Due to the great and continuous distribution area of the species, it becomes difficult to delineate predefined populations. Thereby, landscape genetics approach was a better suited tool for studying the population structure of *T. tetradactyla* individuals (N=176). Comparatively, two approaches were used: grouping of individuals by a geographical proximity criterion, designated here as *a priori* (20 populations), and analysis based on individuals, without previous information of population grouping, referred here as a posteriori (four transects tested). Moderate levels of diversity were found (average of 11.38 alleles per locus), and few evidences for structuring between populations of the sampled localities (K=2 in most tests). The individuals showing the major differentiation were

originated from the Amazon Forest. This region also demonstrated higher genetic diversity (allelic richness and expected heterozygosity) than others. By the other hand, populations distributed in Atlantic Forest and adjacent regions demonstrated a pattern of isolation by distance. Populations from central Brazil (Cerrado and Pantanal) did not show distinction from the others.

Finally, the adaptive genetic variation of the species was studied through DRB gene diversity, from MHC. This gene family codes for proteins involved in the antigen recognition and activation of the adaptive immune response, and are regulated by natural selection, especially by selective pressure driven by pathogens. It is expected that the pathogen diversity is distinct in Brazilian biomes, being florested biomes more diverse than dry central Brazil environments, which represents different selective pressures. Therefore, the diversity of *DRB* exon 2 gene in individuals (N=65) from different Brazilian biomes was investigated through Next Generation Sequencing (454 GS Junior), and the results of allele distributions were compared with microsatellites.

A high diversity was found (60 alleles in amino acid level and 70 in nucleotide level) and clear signatures of positive selection in *DRB* (dN/dS=2.94). Greater allelic richness and private allele number were found in forested biomes, especially in the Amazon Florest. Besides, neutral markers (microsatellites) demonstrated similar patterns to *DRB*, revealing the strength of demographic events and genetic drift in shaping the diversity patterns in MHC.

Introdução Geral

Introdução Geral

A diversidade genética é essencial para a continuidade das espécies na natureza e o componente genético é um dos parâmetros utilizados para avaliar o grau de ameaça em que uma espécie se encontra (Frankham *et al.* 2002). De fato, uma grande meta-análise de espécies de diversos táxons demonstrou que a maioria dos táxons ameaçados apresenta, concomitantemente, menor diversidade genética do que seus táxons próximos não ameaçados (Spielman *et al.* 2004). Embora a baixa diversidade genética não determine necessariamente um grau de ameaça para populações ou espécies, é um dos indicadores de alerta para propósitos futuros de conservação (Noss 1990). Trabalhos que focam em diversidade genética enfatizaram, sobretudo, a variação neutra do genoma, mas hoje em dia podemos aceder com a mesma facilidade à diversidade genética neutra e à adaptativa, como apresentado em seguida.

O estudo da variação genética ao longo da distribuição geográfica de uma espécie foi, e é ainda, extensamente explorado pela disciplina da filogeografia (Avise 2000). Originalmente dedicada ao estudo do DNA mitocondrial como marcador genético, a filogeografia expandiu grandemente nos últimos anos, incorporando análises complexas, mais integrativas, de diversificação de populações ou espécies com modelos estatísticos preditivos de teste de hipótese (Knowles 2009; Hickerson *et al.* 2010). Os marcadores genéticos como ferramenta de estudo também passaram por grande desenvolvimento. Estudos nesta área fizeram uso de polimorfismos de base única (*Single Nucleotide Polymorphism*, SNP) em regiões codificantes e intrônicas, microssatélites, e mais recentemente, de genomas parciais ou completos (Sunnucks 2000; Davey *et al.* 2011). Ainda entre as vertentes originadas da área da filogeografia, surgiu recentemente a genética da paisagem que tem como objetivo fornecer informações sobre a interação entre as características da paisagem (como a presença de barreiras físicas à dispersão, que podem ser desde rios e montanhas até fatores menos óbvios, como gradientes de umidade no ambiente ou a presença de uma rodovia extensa) e os processos micro-evolutivos (como o fluxo gênico, a deriva genética e seleção natural dentro de uma mesma espécie; Manel *et al.* 2003). Os principais enfoques da disciplina da genética da paisagem são a detecção das descontinuidades genéticas, sem a inferência *a priori* de quais e/ou quantas são as populações, e a correlação destas com a paisagem e características ambientais. Esta abordagem pode ser útil na identificação de barreiras crípticas, que podem ser barreiras ao fluxo gênico contemporâneo entre as populações sem nenhuma causa óbvia, ou de regiões de contato secundário entre populações previamente isoladas, revelando assim uma barreira ao fluxo gênico histórica (revisado em Holderegger & Wagner 2008; Storfer *et al.* 2010).

Embora o crescimento das áreas da filogeografia e genética da paisagem tenham sido constantes a nível mundial, ainda existe um atraso no conhecimento filogeográfico de táxons da América do Sul. Particularmente em relação a mamíferos, ainda que o Neotrópico apresente uma grande riqueza de espécies, até o ano de 2008 somente cerca de 7% de todos os estudos de estruturação genética foram feitos com mamíferos sul-americanos (Beheregaray 2008). Não obstante, os estudos com táxons da América do Sul podem fornecer conhecimento sobre os processos históricos subjacentes à diversificação nesta região (por exemplo, Turchetto-Zolet *et al.* 2013). O entendimento destes processos é de grande importância, sobretudo considerando a vasta degradação das paisagens neotropicais devido à recente ação antrópica, resultado em severa perda de hábitat (Brown & Brown 1992; Carvalho *et al.* 2009). O estudo da diversidade genética neutra é,

geralmente, a primeira escolha para acessar os padrões de diversidade dos táxons dessas regiões.

Diversidade genética neutra

A teoria neutra da evolução assume que a principal força evolutiva atuando na fixação de novos alelos em uma população é a flutuação aleatória das frequências alélicas em populações finitas, geração após geração, processo conhecido como deriva genética. Assim, a maioria da variação genética intra e interespecífica é neutra com relação à seleção e, portanto, diferentes alelos de um lóco possuem o mesmo valor adaptativo, sem qualquer efeito diferenciado ou direto no *fitness* de cada individuo (descrito em Kimura 1983).

O estudo da variabilidade genética com marcadores seletivamente neutros nos permite observar padrões que podem ser explicados, por exemplo, por processos demográficos sofridos pelas populações de uma espécie ao longo do tempo, tais como migrações (medidas em fluxo gênico entre populações), efeito fundador, alterações do tamanho efectivo populacional, flutações populacionais, subdivisão de populações pelo surgimento de barreiras (vicariância), ou por efeito de isolamento por distância (Hartl & Clark 2006).

Os microssatélites, ou repetições de sequência simples, são marcadores considerados neutros em sua maioria, estão frequentemente presentes em regiões genômicas não-codificadoras ao longo de todo o genoma, e se caracterizam pela sua alta taxa de mutação associada ao alto grau de polimorfismo de tamanho, relacionado ao número variável de seus motivos repetitivos (Tautz 1989). Os marcadores microssatélites, devido ao seu elevado grau de polimorfismo, são marcadores altamente eficientes na

14

resolução de diversas questões a respeito da biologia das espécies (Frankham *et al.* 2002), e têm sido bastante usados para estudos de populações naturais, especialmente aqueles voltados para a conservação de espécies de mamíferos (Garner *et al.* 2005). A abrangente utilidade destes marcadores os torna uma ferramenta metodológica poderosa na inferência sobre padrões de dispersão intraespecíficos, grau de divergência entre populações, o tamanho das mesmas, a diversidade genética remanescente nelas, o grau de endogamia, assim como sobre as relações de parentesco entre indivíduos, individualização de amostras, entre outros. Assim, sua utilização pode contribuir para o melhor conhecimento de alguns aspectos da biologia destas espécies que são difíceis de estudar através de observações diretas (Balloux & Lugon-Moulin 2002, revisado em Selkoe & Toonen 2006).

Diversidade genética adaptativa

Normalmente, a maioria dos estudos acessa apenas a variabilidade neutra, onde se assume que essa variação é indicadora da variabilidade adaptativa. Entretanto, estudos apontam que a diversidade neutra pode não refletir acuradamente a diversidade genômica subjacente (Holderegger *et al.* 2006; Väli *et al.* 2008, Ljungqvist *et al.* 2010).

A variação genética que é controlada pela seleção natural é conhecida como diversidade genética adaptativa. Esta pode ser acessada através de experimentos de genética quantitativa ou por marcadores adaptativos que advêm de lócos de genes funcionais, cuja variação alélica tem consequências no *fitness* de indivíduos (Sommer 2005; Holderegger *et al.* 2006).

Populações que apresentam baixa diversidade em marcadores genéticos neutros podem apresentar uma diversidade também diminuída em regiões genômicas importantes para a adaptação a mudanças ambientais, tais como alterações nas comunidades de patógenos ou mudanças climáticas. Muitos genes são potencialmente responsáveis pela habilidade evolutiva de resposta às mudanças ambientais, como genes de pigmentação ou genes do sistema imune (Höglund 2009). Exemplos de correlações positivas entre diversidade neutra e adaptativa podem ser encontrados na literatura, como foi relatado para a raposa do Ártico (*Vulpes lagopus*), que demonstrou baixa diversidade genética em suas populações continentais e insulares em microssatélites e em genes do complexo maior de histocompatibilidade (Geffen *et al.* 2007; Ploshnitsa *et al.* 2012).

De fato, os genes do complexo principal de histocompatibilidade, MHC (do inglês, *Major Histocompatibility Complex*), são exemplos clássicos de marcadores adaptativos. Constituem uma família multigênica cujos produtos são glicoproteínas de superfície celular que desempenham o importante papel de apresentar peptídeos às células T no sistema imune (Klein 1986). Este complexo de genes é composto, sobretudo, por duas subfamílias: MHC classe I e classe II. Ambas apresentam antígenos e permitem o reconhecimento de moléculas próprias e não-próprias (Acevedo-Whitehouse & Cunningham 2006). Alguns genes das classes I e II apresentam níveis extraordinários de polimorfismos sendo caracterizados por grande número de alelos de frequência intermediária e com alta ocorrência de substituições não-sinônimas dentro das regiões que codificam os sítios de ligação de antígenos (do inglês, *Antigen Binding Site –* ABS). Esse padrão é inconsistente com um cenário de neutralidade, sugerindo a ação de seleção natural do tipo balanceadora na região (Hughes & Yeager 1998).

O uso dos genes do MHC como marcadores genéticos em organismos não-modelo tem crescido muito nos últimos anos. Alguns dos estudos visam compreender o papel da seleção e da deriva genética em moldar a diversidade das populações atuais, através da comparação de padrões revelados pelo uso de marcadores neutros, como microssatélites, e do MHC (Aguilar *et al.* 2005; Bos *et al.* 2008; Agudo *et al.* 2011). Outros estudos propõem-se a correlacionar a variação encontrada no MHC com aspectos da biologia da espécie estudada, como escolha de parceiros e reconhecimento de parentesco (Ditchkoff *et al.* 2001; Landry *et al.* 2001), ou no uso da diversidade do MHC como indicativo de viabilidade de populações de espécies ameaçadas, uma vez que estes genes estão relacionados com a resposta imune e defesa a patógenos (Langefors *et al.* 2001; Wegner *et al.* 2003).

Apesar do crescente interesse dos geneticistas de populações no complexo de genes do MHC, este é ainda pouco estudado em espécies não-modelo e o reconhecimento e caracterização dos seus alelos pode ser bastante complicado. As técnicas para detectar alelos de genes do MHC podem introduzir variações falsas (artefatos) que precisam ser identificadas e tratadas. Alguns fatores impõem dificuldades significativas na genotipagem de MHC: (i) duplicações recentes, (ii) variação no número de lócos em cada espécie e (iii) a presença de pseudogenes (Babik 2010). Apesar da existência de alguns estudos de MHC em espécies não-modelo, correlações claras sobre a diversidade nestes marcadores e seus efeitos no *fitness* de indivíduos/populações ainda não foram completamente esclarecidos (Radwan *et al.* 2010) e as questões pendentes representam material inestimável para futuros estudos.

Pilosa: Importância histórico-evolutiva

A ordem Pilosa (Gardner 2005), linhagem das atuais preguiças arbóreas e tamanduás, juntamente com seu grupo irmão, a ordem Cingulata, os tatus, constitui a magna-ordem Xenarthra, uma das quatro linhagens basais de placentários que englobam as ordens atuais de mamíferos eutérios (Springer *et al.* 2004).

As espécies atuais de xenarthros são remanescentes de uma radiação evolutiva promovida por seu isolamento na América do Sul durante o Terciário (entre o Paleoceno e o Eoceno, cerca de 65 milhões de anos atrás), e influenciada pelas mudanças ambientais no passado (Patterson & Pascual 1972; Delsuc et al. 2004). Seu registro fóssil parece ser restrito ao continente americano (Delsuc et al. 2001). A monofilia da ordem é classicamente reconhecida, todos os pilosos fósseis e viventes exibem uma redução dental e perda do esmalte dentário (característica chave do grupo), até o extremo caso dos tamanduás, que são desdentados (Engelman 1985; Vizcaíno 2009). Apesar de ser composta por apenas 10 espécies, retêm boa parte da história evolutiva dos placentários e são, portanto, uma das prioridades para conservação da diversidade filogenética em vertebrados (Fonseca 2001). Apesar de apresentar tal importância, apenas alguns estudos foram publicados até o momento utilizando ferramentas de genética molecular, abordando, por exemplo, relações filogenéticas dentro deste e entre outros grupos de mamíferos (Arnason et al. 1997; De Jong 1998; Delsuc 2002), diversidade genética entre populações de uma mesma espécie (Moraes-Barros et al. 2002; Moraes-Barros et al. 2006) ou focados em genética da conservação (Collevati et al. 2007; Lara-Ruiz et al. 2008).

18

Tamanduá-mirim: espécie de ampla distribuição geográfica

O tamanduá-mirim, *Tamandua tetradactyla*, também conhecido como tamanduá-decolete, é um membro da família Myrmecophagidae, ordem Pilosa, infra-ordem Vermilíngua (animais com língua fina e alongada) (Engelman 1985). Seu gênero compreende duas espécies: *T. tetradactyla e T. mexicana*, sendo a primeira distribuída ao longo da América do Sul, leste dos Andes, e a última no sudoeste do México até oeste dos Andes. Assim, o tamanduá-mirim possui ampla distribuição geográfica, é encontrado na América do Sul da Venezuela e Trinidad ao norte da Argentina, sul do Brasil e no Uruguai (Superina *et al.* 2010, Figura 1) a elevações de até 2000 metros (Novak *et al.* 1983).



Figura 1- Mapa da distribuição atual de *Tamandua tetradactyla* na América do Sul. Extraído de Superina *et al.* 2010.

Este mamífero de médio porte, possui hábito predominantemente noturno e solitário, se alimenta de uma combinação variável de formigas e cupins dependendo da disponibilidade dessas presas no ambiente (Montgomery & Lubin 1977), com o auxílio de sua língua comprida com muco viscoso, adaptação chave de animais mirmecofágicos. Escansorial, possui uma cauda preênsil bem desenvolvida parcialmente coberta com pelagem. Embora tenha preferência por áreas florestais, é possível também encontrar tamanduás-mirins forrageando em áreas abertas de Cerrado ou vegetação rasteira (Redford, 1994), e ainda em regiões montanhosas de clima tropical (Eisenberg 1989). Os tamanduás têm gestação longa, cerca de 150 dias, e geram somente um filhote, que é alimentado pela fêmea durante seu primeiro ano de vida (Montgomery & Lubin 1977).

O seu padrão de pelagem varia com o espectro da distribuição, sendo possível encontrar espécimes de cor uniformemente clara a dourada, na porção noroeste de sua ocorrência sul-americana, até indivíduos revestidos por um colete preto que cobre seu dorso (Gardner 2008). Esta grande variação de pelagem individual e geográfica gerou a descrição de subespécies (Wetzel 1975). Atualmente estão descritas quatro subespécies para *T. tetradactyla* (Gardner 2008, revisado em Hayssen 2011):

• *T. tetradactyla nigra* Geoffroy St. Hillaire, 1803 (porção norte do Brasil, Guianas e parte da Venezuela e Colômbia):

• *T. tetradactyla quichua* O. Thomas, 1927 (porção oeste, abrangendo Acre e Peru);

• *T. tetradactyla straminea* Cope, 1889 (Brasil central, parte da Bolívia, Argentina e Paraguai);

• *T. tetradactyla tetradactyla* Linnaeus, 1758 (costa atlântica do Brasil).

20

Um estudo morfológico mais amplo analisou a morfologia craniana e coloração da pelagem em tamanduás através de morfometria geométrica. Os caracteres de coloração da pelagem apontaram para a divisão dos indivíduos amostrados em dois grupos: (i) um ao norte do rio Amazonas e oeste do rio Madeira, onde se distribuem animais com alta variação da coloração da pelagem; e (ii) outro grupo ao sul do rio Amazonas e leste do rio Madeira, de animais com padrão de coloração clássicos da espécie (Ohana 2011).

Outros estudos sobre a espécie são restritos principalmente a área da ecologia, ou ainda em estudos de filogenia de mamíferos (Meritt 1975; Engelman 1985; Hay et al., 1994; Rodrigues et al., 2001; Delsuc 2001, 2002; Jimeno 2003). A espécie é considerada pela IUCN como pertencente à categoria "Least Concern" (menor preocupação), justificada pelo fato da espécie ser comum, presente em algumas áreas protegidas, ter grande distribuição geográfica e presumida grande população (Miranda et al. 2014). Órgãos como IBAMA e CITES estão em concordância com esta classificação (IBAMA http://www.ibama.gov.br/fauna/extincao.htm; CITES – www.cites.org, acessado em 11 de Agosto de 2014). Porém, localmente, o panorama parece variar em outra direção. No estado brasileiro Rio Grande do Sul a espécie é considerada vulnerável, pois no local as áreas de florestas foram grandemente devastadas, e as respectivas populações declinaram junto com o habitat (Fontana et al. 2003). No estado de Mato Grosso o animal já é considerado raro em pequenas reservas de Cerrado. Inclusive, é menos visualizado em campo que o vulnerável tamanduá-bandeira, Myrmecophaga tridactyla (Rocha & Dalponte 2006), como também acontece no Parque Nacional das Emas, em Goiás (Rodrigues et al. 2002). Em outro estudo, sobre a situação dos xenartros no Uruguai, o tamanduá-mirim aparece como ameaçado devido à perda de habitat e caça esportiva (Fallabrino & Castiñeira 2006).

Alguns trabalhos realizaram o levantamento de atropelamentos de diversas espécies em grandes rodovias brasileiras, e a espécie é mostrada como um dos mamíferos mais frequentemente atropelados (Casella *et al.* 2006; Pereira *et al.* 2006; Cherem *et al.* 2007; Catzeflis & Thoisy 2012).

Como exposto, existe uma carência de estudos sobre as populações de tamanduámirim: somente estudos morfológicos têm sido insuficientes para o entendimento da estruturação populacional da espécie e a variabilidade genética é virtualmente desconhecida. Neste quesito, há apenas um estudo que realiza a estimativa da diversidade genética de alguns indivíduos de tamanduá-mirim distribuídos no Brasil através do marcador mitocondrial Citocromo Oxidase subunidade I (COI), que descreve alta diversidade haplotípica e ausência de estruturação genética (Moraes-Barros *et al.*, dados não publicados).

Esta tese tem como objetivo principal preencher estas lacunas de conhecimento, contribuindo para o melhor entendimento da relação entre as populações de diferentes partes da distribuição do tamanduá-mirim através do uso de ferramentas de genética de populações e do acesso a regiões neutras e adaptativas do genoma da espécie. Tendo em vista que o tamanduá-mirim ocorre em todo o território brasileiro, nos seus diferentes biomas e ecossistemas, apresentando uma distribuição contínua, se torna complicado traçar populações pré-definidas. Assim, a abordagem da genética da paisagem foi a ferramenta utilizada e mais apropriada para o estudo da variação genética e estrutura de populações em *T. tetradactyla*, que foi realizado através do uso de microssatélites isolados *de novo*. Ainda pelo mesmo motivo (ampla ocorrência), é possível que haja diferenças na variabilidade neutra e adaptativa, uma vez que a espécie se encontra adaptada a tantas paisagens

diferentes e distintas pressões seletivas. Por isso, um marcador de uma região genômica

sabidamente sob seleção (MHC) foi usado para investigar a variabilidade adaptativa.

Referências

- Acevedo-Whitehouse K, Cunningham A (2006) Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology & Evolution* **21**, 433–8.
- Agudo R, Alcaide M, Rico C *et al.* (2011) Major histocompatibility complex variation in insular populations of the Egyptian vulture: inferences about the roles of genetic drift and selection. *Molecular ecology*, **20**, 2329–40.
- Aguilar A, Smith TB, Wayne RK (2005) A comparison of variation between a MHC pseudogene and microsatellite loci of the little greenbul (*Andropadus virens*). *BMC Evolutionary Biology* **5**, 47.
- Arnason, U, Gullberg, A, Janke, A (1997) Phylogenetic analyses of mitochondrial DNA suggest a sister group relationship between Xenarthra (Edentata) and Ferungulates. *Molecular Biology and Evolution* 14(7), 762-768.
- Avise (2000). *Phylogeography. The history and formation of species*. Havard University Press, Massachusetts.
- Babik W (2010) Methods for MHC genotyping in non-model vertebrates. *Molecular Ecology Resources*, **10**, 237–51.
- Balloux F & Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**(2), 155-165.
- Beheregaray L B (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, **17**, 3754–3774.
- Bos DH, Gopurenko D, Williams RN, Dewoody JA (2008) Inferring population history and demography using microsatellites, mitochondrial DNA, and major histocompatibility complex (MHC) genes. *Evolution; international journal of organic evolution*, **62**, 1458–68.

- Brown KS & Brown GG (1992) Habitat alteration and species loss in Brazilian forests. *In*: Whitmore TC & Sayer JA (eds.) **Tropical deforestation and species extinction**. Chapman & Hall, Londres.
- Carvalho FMV, De Marco Júnior P, Ferreira LG (2009) The Cerrado into-pieces: habitat fragmentation as a function of landuse in the savannas of central Brazil. *Biological Conservation*, **142**(7), 1392-1403.
- Casella J, Cáceres NC, Goulart CS, Filho ACP (2006) Uso de sensoriamento remoto e análise espacial na interpretação de atropelamentos de fauna entre Campo Grande e Aquidauana, MS. Anais 1º Simpósio de Geotecnologias no Pantanal, 321-326.
- Catzeflis F, Thoisy, B (2012) Xenarthrans in French Guiana: a brief overview of their distribution and conservation status. *Edentata* **13**, 29-37.
- Cherem JJ, Kammers M, Ghizoni-Jr IR, Martins A (2007) Mamíferos de médio e grande porte atropelados em rodovias do Estado de Santa Catarina, sul do Brasil. *Biotemas* **20** (3), 81-96.
- Collevatti RG, Leite KCE, Miranda GHB, Rodrigues FGH (2007) Evidence of high inbreeding in a population of the endangered giant anteater, *Myrmecophaga tridactyla* (Myrmecophagidae), from Emas National Park, Brazil. *Genetics Molecular Biology* **30**, 112-120.
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews* 12:499-510
- De Jong WW (1998) Molecules remodel the mammalian tree. *Trends in Ecology and Evolution* **13**(7), 270-275.
- Delsuc F, Catzeflis FM, Stanhope MJ, Douzery EJP (2001). The evolution of armadillos, anteaters and sloths depicted by nuclear and mitochondrial phylogenies: implications for the status of the enigmatic fossil *Eurotamandua*. *Proc. R. Soc. Lond. B.* **268**, 1605-1615.
- Delsuc F, Scally M, Madsen O, Stanhope MJ, de Jong WW, Catzeflis F, Springer MS, Douzery EJP (2002). Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. *Molecular Biology Evolution* **19**(10), 1656-1671.
- Delsuc F, Vizcaíno SF, Douzery EJP (2004) Influence of Terciary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans. *BMC Evolutionary Biology* **4**, 11-24.

- Ditchkoff SS, Lochmiller RL, Masters RE, Hoofer SR & Van der Bussche RA (2001) Majorhistocompatibility-complex-associated variation in secondary sexual traits of white-tailed deer (*Odocoileus virginianus*): evidence for good-genes advertisement. *Evolution* **55**, 616-625.
- Eisenberg JF (1989) Order Xenarthra (Edentata). In: *Mammals of the Neotropics The northern Neotropics*. Vol.1. The University of Chicago Press, London, pp. 50-67.
- Engelman GF (1985) The phylogeny of the Xenarthra. In: Montgomery GG (ed.) *The evolution and Ecology of Armadillos, Sloths and Vermilinguas.* Smithsonian Institution Press, Washington.
- Fallabrino A, Castiñeira E (2006) Situacion de los Edentados en Uruguay. Edentata 7: 1-3.
- Fonseca G. (2001) The conservation of Xenarthra will be vital for the preservation of mammalian phylogenetic diversity. *Edentata* **41**, 1-10.
- Fontana CS, Bencke GA, Reis RE (eds.) (2003) *Livro vermelho da fauna ameaçada de extinção do Rio Grande do Sul*. Edi-PUCRS, Porto Alegre.
- Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to conservation genetics*. Cambridge University Press, UK.
- Gardner AL (2005) Order Pilosa. In: Wilson DE, Reeder DM (eds.) Mammal species of the world: a taxonomic and geographic reference. 3 ed. Baltimore, The John Hopkins University Press, p. 98-102.
- Gardner, A L (2008) *Mammals of South America Vol 1*. Baltimore, The John Hopkins University Press.
- Garner A, Rachlow JL, Hicks JF (2005) Patterns of genetic diversity and its loss in mammalian populations. *Conservation Biology* **19**(4), 1215-1221.
- Geffen E, Waidyaratne S, Dalén L, Angerbjörn A, Vila C, Hersteinsson P, Fuglei E, White PA, Goltsman ME, Kapel CMO (2007) Sea ice occurrence predicts genetic isolation in the Arctic fox. *Molecular Ecology* 16, 4241–4255.
- Hartl DL, Clark AG (2006) **Principles of Population Genetics**. Sinauer Associates, Massachusetts.
- Hay MA, Bellem AC, Brown JL, Goodrowe KL (1994) Reproductive patterns in Tamandua (*T. tetradactyla*). J. Zoo. Wildl. Med. 25, 248-258.
- Hayssen V (2011) *Tamandua tetradactyla* (Pilosa: Myrmecophagidae). *Mammalian Species*, **43**, 64–74.

- Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Vitoriano PF, Yoder AD (2010) Phylogeography's past, present and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, **54**(1), 291-301.
- Hoderegger R & Wagner HH (2008) Landscape genetics. *BioScience* 58(3), 199-207.
- HoldereggerR, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology* **21**(6), 797-807.
- Höglund J (2009) Evolutionary Conservation Genetics. Oxford University Press, New York.
- Hughes a L, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annual review of genetics*, **32**, 415–35.
- Jimeno GP (2003) Crianza artificial y manejo reprocdutivo de los Tamanduá (*Tamandua tetradactyla*) en el Jardin Zoológico de Rosario, Argentina. *Edentata* **5**: 24-28.
- Kimura M (1983) The neutral theory of molecular evolution. London. Cambridge University Press, UK.
- Klein J (1986) Natural History of the Major Histocompatibility Complex. Wiley, New York.
- Knowles L (2009) Statistical Phylogeography. J Annual Review of Ecology, Evolution, and Systematics 40, 593-612
- Landry C, Garant D, Duchesne P & Bernatchez L (2001) 'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). *Proc. R. Soc. Lond. B.* 268, 1279-1285.
- Langefors A, Lohm J, Grahn M, Andersen O & von Schantz T (2001) Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas* salmonicida in Atlantic salmon. Proc. R. Soc. Lond. B. **268**, 479-485.
- Lara-Ruiz P, Chiarello A, Santos FR (2008) Extreme population divergence and conservation implications for the rare endangered Atlantic Forest sloth, *Bradypus torquatus* (Pilosa: Bradypodidae). *Biological Conservation* 141, 1332-1342.
- Ljungqvist M, Åkesson M, Hansson B (2010) Do microsatellites reflect genome-wide genetic diversity in natural populations? A comment on Väli *et al.* (2008). *Molecular Ecology*, **19**, 851-855.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* **18**(4), 189-197.

- Meritt J (1975) *Readability*. United Kingdom Reading Association. London: Hodder and Stoughton.
- Miranda F, Fallabrino A, Arteaga M, Tirira DG, Meritt DA, Superina M (2014) *Tamandua tetradactyla*. The IUCN Red List of Threatened Species. Version 2014.2. <www.iucnredlist.org>. Downloaded on 09 August 2014.
- Montgomery GG & Lubin YD (1977). Prey influences on movements of neotropical anteaters. In: Philips RL & Jonkel C (eds.) *Proceedings of 1975 Predator Symposium*. Missoula, University of Montana, pp. 103-131.
- Moraes-Barros N, Morgante JS, Miyaki CY (2002) Genetic diversity in different populations of sloths assessed by DNA fingerprinting. *Braz. J. Biol.* **62**(3), 503-508.
- Moraes-Barros N, Silva JAB, Miyaki CY, Morgante JS (2006) Comparative phylogeography of the Atlantic Forest endemic sloth (*Bradypus torquatus*) and the wide-spread three-toed sloth (*Bradypus variegatus*) (Bradipodidae: Xenarthra). *Genetica* **126**(1-2), 189-198.
- Noss RF (1990) Indicators for monitoring biodiversity: a hierarchical approach. *Conservation Biology*, **4**(4), 355-364.
- Novak RM (1983) Walkers Mammals of the world. The Johns Hopkins University Press, Baltimore e Londres. pp. 147-168.
- Ohana JAB (2011) Variação morfológica do tamanduá-mirim, Tamandua tetradactyla (Linnaeus, 1758)(Pilosa, Vermilingua). Dissertação de Mestrado. Universidade Federal do Pará/Museu Paraense Emílio Goeldi.136 pp.
- Patterson B & Pascual R (1972). The fossil mammal fauna of South America. In: Keast A, Erk FC and Glass BP (ed.) *Evolution, mammals and southern continents*. State University of New York Press, Albany, pp 247-309.
- Pereira APFG, Andrade FAG, Fernandes MEB (2006) Dois anos de monitoramento dos atropelamentos de mamíferos na rodovia PA-458, Bragança, Pará. *Bol. Mus. Para. Emílio Goeldi* 1(3), 77-83.
- Ploshnitsa AI, Goltsman ME, Macdonald DW, Kennedy LJ, Sommer S (2012) Impact of historical founder effects and a recent bottleneck on MHC variability in Commander Arctic foxes (*Vulpes lagopus*). *Ecology and Evolution* 2, 165–80.
- Radwan J, Biedrzycka A, Babik W (2010) Does reduced MHC diversity decrease viability of vertebrate populations? *Biological Conservation* **143**, 537-544.

Redford KH (1994) The Edentates of the Cerrado. *Edentata* 1: 4-10.

- Rocha EC & Dalponte JC (2006) Composição e caracterização da fauna de mamíferos de médio e grande porte em uma pequena reserva de cerrado em Mato Grosso, Brasil. *R. Árvore* 30: 669-678.
- Rodrigues FGH, Marinho-Filho JS, Santos HG (2001) Home ranges of translocated lesser anteaters (*Tamandua tetradactyla*) in the Cerrado of Brazil. *Oryx* **35**, 166-169.
- Rodrigues FHG (2002) Composição e caracterização da fauna de mamíferos do Parque Nacional das Emas, Goiás, Brasil. *Revista Brasileira de Zoologia* **19**(2), 589-600.
- Selkoe K, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**, 615–29.
- Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology*, **2**, 16-34.
- Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences* 101, 15261-15264.
- Springer MS, Stanhope MJ, Madsen O & de Jong WW (2004). Molecules consolidate the placental mammalian tree. *Trends in Ecology and Evolution* **19**(8), 430-438.
- Storfer A, Murphy MA, Spear SF, Holderegger R, Waits L (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496-3514.
- Sunnucks P (2000) Efficient genetic markers for population biology. TREE 15(5):199-203.
- Superina M, Miranda FR, Abba AM (2010) The 2010 Anteater Red List Assessment. *Edentata* **11**(2), 96-114.
- Tautz D (1989) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acid Research* **17**, 6463-6471.
- Turchetto-Zolet C, Pinheiro F, Salgueiro F, Palma-Silva C (2013) Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, **22**, 1193–213.
- Väli U, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, 17, 3808-3817.
- Vizcaíno SF (2009) The teeth of the "toothless": novelties and key innovations in the evolution of xenarthrans (Mammalia, Xenarthra). *Paleobiology* **35**(3), 343-366.

- Wegner KM, Reusch TBH & Kalbe M (2003) Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *Journal of Evolutionary Biology* 16, 224-232.
- Wetzel RM (1975) The species of *Tamandua* Gray (Edentata, Myrmecophagidae). *Proceedings* of the Biological Society of Washington **88**, 95-112.

<u>Objetivos</u>

Objetivos

Objetivo Geral

Caracterizar a diversidade genética e estruturação populacional da espécie *Tamandua tetradactyla* ao longo de sua distribuição, com enfoque nos biomas brasileiros, através de marcadores neutros e não-neutros, a fim de investigar a relação entre a paisagem diversa de ocorrência da espécie e a estruturação genética de suas populações.

Objetivos Específicos

- Amostras de *Tamandua tetradactyla* são de difícil obtenção, e frequentemente são coletadas de forma oportunista, resultando no uso de tipos diferentes de material biológico, como tecido fresco, bulbo de pelos, carcaça de atropelamento, pele de museu, entre outros. Assim, um dos objetivos deste trabalho foi padronizar a qualidade do DNA e a eficiência de sua amplificação a partir dos diferentes tipos de amostras biológicas;
- 2. Usando marcadores neutros do tipo microssatélite, pretendeu-se investigar a estruturação genética dentro da área de ocorrência amostrada da espécie (variabilidade genética e diferenciação geográfica), e quais os seus potenciais geradores, nomeadamente paisagem (diferentes biomas) ou isolamento por distância.
- 3. Utilizando como marcador não-neutro o gene DRB do complexo MHC de Classe II, pretendeu-se constatar os averiguar de seleção natural no gene, e verificar se existem composições alélicas distintas entre indivíduos de tamanduá-mirim de diferentes biomas

brasileiros, partindo do pressuposto que estes diferentes ambientes apresentam pressões seletivas distintas.

 Finalmente, o ultimo objetivo foi investigar comparativamente os padrões de diversidade genética encontrados pelos marcadores neutros e não-neutros para tamanduá-mirim, e verificar se houve uma contribuição de eventos demográficos e deriva genética na riqueza e distribuição da diversidade genética adaptativa (estimada indiretamente através do gene *DRB*).

Capítulo 1

Amostras históricas e não-invasivas: um estudo de caso de erros de genotipagem em microssatélites isolados *de novo* para o tamanduá-mirim (*Tamandua tetradactyla* L., Pilosa).

Historical and noninvasive samples: a study case of genotyping errors in newly isolated microsatellites for the lesser anteater (Tamandua tetradactyla L., Pilosa)

Apresentação Capítulo 1

Este capítulo surgiu da necessidade de padronização dos métodos de biologia molecular e critérios de análise para a obtenção de uma genotipagem de microssatélites confiável e comparável entre diferentes tipos de amostras. A amostragem de indivíduos de tamanduá-mirim é complicada pela difícil detecção do animal em trabalhos de campo, pela sua captura e manipulação. Assim, a maioria das amostras biológicas utilizadas em trabalhos de genética é coletada oportunisticamente: carcaças atropeladas, pelos arrancados com o bulbo e peles de museu. Todos esses tipos de amostras biológicas foram usadas no presente trabalho. Dessa forma, fez-se necessário um estudo piloto para estimar os erros de genotipagem provenientes dos tipos de amostras usadas, e estabelecer um padrão de qualidade considerado confiável para realizar os demais objetivos do projeto. Além disso, o estudo caracteriza um conjunto de oito lócos de microssatélites isolados para a espécie *Tamandua tetradactyla*. O trabalho foi publicado no periódico *Molecular Ecology Resources*, volume 14(3), em Maio de 2014.

Historical and noninvasive samples: a study case of genotyping errors in newly isolated microsatellites for the lesser anteater (*Tamandua tetradactyla* L., Pilosa)

CAMILA L. CLOZATO*, NADIA MORAES-BARROS#*, FABRÍCIO R. SANTOS† and JOÃO STENGUEL MORGANTE*

(Artigo publicado: Clozato CL, Moraes-Barros N, Santos FR, Morgante JS (2014) Historical and non-invasive samples: a study case of genotyping errors in newly isolated microsatellites for the lesser anteater. *Molecular Ecology Resources*, 14(3), 531-540.)

Keywords genotyping errors, historical samples, microsatellite loci, noninvasive samples, quality index, *Tamandua tetradactyla*

Abstract

Tamandua tetradactyla (Pilosa), the lesser anteater, is a medium size mammal from South America. Its wide distribution through different landscapes, solitary and nocturnal habits, and the difficulty to capture and contain specimens limit the amount of individuals and populations sampled during fieldworks. These features along with the lack of specific molecular markers for the lesser anteater might be the causes for paucity in population genetic studies for the species. Historical samples from museum specimens, such as skins, and noninvasive samples such as plucked hair can be supplementary sources of DNA samples. However, the DNA quantity and quality of these samples may be limiting factors in molecular studies. In this study we describe nine microsatellite loci for T. tetradactyla and test the amplification success, data reliability and estimate errors on both historical and noninvasive sample sets. We tested nine polymorphic microsatellites and applied the quality index approach to evaluate the relative performance in genotype analysis of 138 historical samples (study skin) and 19 noninvasive samples (plucked hair). The observed results show a much superior DNA quality of noninvasive over historical samples and support the quality index analysis as a practical tool to exclude samples with doubtful performance in genetic studies. We also found a relationship between the age of noninvasive samples and DNA quality, but lack of evidence of this pattern for historical samples.
Introduction

The lesser anteater (*Tamandua tetradactyla*) is a medium sized mammal of the Pilosa order. The species occurs east of Andes, from Colombia, Venezuela, Trinidad to the Guyanas, and south to northern Argentina and northern Uruguay, from the sea level up to 1600 m (Wetzel 1975; Gardner 2005; revised in Hayssen 2011). It is able to occupy different types of habitats, from savannas and wetlands to rain forests. Mostly nocturnal and solitary, it uses the ground and trees for moving and foraging, and eventually may use armadillo's burrows and hollow trees to hide (Montgomery & Lubin 1977). Despite the wide distribution, the characteristics and habits of the species hinder the obtainment of biological samples for molecular studies, especially when populations are addressed. Searching for a lesser anteater in the field is arduous and time consuming, and the use of a trap is not feasible for capture. Once detected, the animal must be properly immobilized with chemicals by a veterinarian to collect blood, a delicate procedure that includes risks to the animal, and that is not always possible due to fieldwork logistic issues.

Therefore, the use of specimens from natural history collections, hereafter referred to as historical samples, as sources of DNA for population genetic studies in the species is a good strategy to achieve an adequate sample size with reasonable geographic coverage, besides adding a time dimension to the analysis, and the possibility to sample from currently extinct populations (Wandeler *et al.* 2007).

Another strategy for sampling difficult free-ranging species is to collect noninvasive samples, such as plucked hair or fecal samples, since there is no need for animal immobilization, and in some cases, no need even for animal visualization, allowing researchers to address questions in

natural populations that would not be possible with traditional methods (Taberlet *et al.* 1999; Waits & Paetkau 2005; Broquet *et al.* 2007; Beja-Pereira *et al.* 2009).

Despite the advantages of both strategies, there are limitations concerning the quantity and quality of DNA samples. In the case of historical samples, the preservation treatments, age and storage conditions seem to be important issues (Wandeler *et al.* 2007; Casas-Marce *et al.* 2010). However, for both types of sample, the presence of inhibitors during the amplification, possibility of contamination, intersample quality variance, and low quantity of DNA are also limitations for their use (Taberlet *et al.* 1996; Hall *et al.* 1997; Wandeler *et al.* 2007).

The most common challenges in genotyping degraded samples are the high rate of amplification failure, presence of false alleles (FA), which are essentially PCR artifacts, and allele dropout (ADO), the non-amplification of an allele from a heterozygote resulting in false homozygotes (Taberlet *et al.* 1999). Moreover, there is also the occurrence of null alleles, which probably happens due to mutations within the primer annealing regions, resulting in non-amplification of some alleles (Shaw *et al.* 1999), also augmenting the amount of homozygotes. In fact, empirical estimates showed that about two thirds of all genotyping errors can originate from the presence of null alleles and imperfect repetitions (Kelly *et al.* 2011).

In order to account for genotyping errors in microsatellite analysis in both sample types, we calculated the quality index (QI) described by Miquel *et al.* (2006) upon a set of newly polymorphic microsatellite loci developed for the lesser anteater. Besides being a powerful tool for the evaluation of genetic diversity and population structure in future studies involving the species, these molecular markers are appropriated to address degraded samples because amplified

loci are usually small DNA fragments, with less than 200 base pairs (bp), easier to be amplified (Wandeler *et al.* 2007).

Even though several authors have addressed the relevance of estimating genotyping errors not only in problematic samples but in all microsatellite studies (Bonin *et al.* 2004, Broquet & Petit 2004, Pompanon *et al.* 2005; Morin *et al.* 2009), once these errors are always present regardless of the method used, few of them have indeed proposed a method capable of a broader use and possibility of comparison between data. Miquel *et al.* (2006) have proposed a standardized method to test the reliability of the genotyping process that allows comparison among samples, loci, protocols and different studies.

In this study we used study skins as representatives of historical samples and plucked hair as representatives of noninvasive samples for comparing the genotyping errors generated in each dataset. By relating the QI of historical and noninvasive samples we aim to determine if their use is reliable for further population genetics studies, considering beforehand the potential misinterpretations to which these sample sets may lead, and test if there are differences regarding their quality during the genotyping procedure.

Materials and Methods

A total of 138 historical samples were used for this study, collected from four different natural history collections (Museu Nacional – MN, Universidade Federal de Santa Catarina – UFSC, Museu de História Natural do Capão da Imbuia – MNHCI, and Museu de Biologia Professor Mello Leitão – MBML) all located in Brazil. The samples were originated from several localities in different Brazilian biomes, collected from 1908 to 2008. All samples consisted of a piece of hide (dried skin) obtained from study skin specimens ranging about 1 cm². Even though bones

and claws are reported to be better sources of DNA than hides (Casas-Marce *et al.* 2010), we did not have access to bone tissue in the collections, and we could not damage any piece of the claws, because those are considered taxonomic markers for *T. tetradactyla*.

A total of 19 noninvasive samples of plucked hair from individualized specimens were used, collected during fieldwork in Mato Grosso do Sul and Ceará states, or donated by the CETAS/IBAMA from Bahia state (a screening center for wild animals), also in Brazil. The samples were collected directly from the animal in the field, transported in sterile tubes at room temperature and stored at 4°C until processed for DNA extraction. They were collected from 2006 to 2012. All biological samples were collected under SISBIO-IBAMA permission number for scientific activities 24001-2/53695225.

Besides historical and noninvasive sample sets, we also used a total of 23 fresh tissue samples of *T. tetradactyla* collected during fieldwork or donated from institutions and researchers, representing populations of Minas Gerais and São Paulo state. These samples were used for the construction of genomic libraries, selections of microsatellites and as a source of comparison for the occurrence of genotyping errors in the previous datasets. All information regarding the samples is encompassed in Table S1 (Supporting Information).

Extraction of genomic DNA was performed with Proteinase K digestion enhanced with DDT 1M, followed by salt precipitation (Sambrook *et al.* 1989). For historical samples, the hide was first cleaned with ultrapure water and ethanol 70% and then hydrated in TE solution for at least 24 hours, which has been reported to improve digestion (Moraes-Barros & Morgante 2007). The entire procedure was performed in a sterile environment dedicated to low quantity DNA samples, exposed to UV light prior to use and segregated physically from laboratories handling PCR

40

products. A blank control was included in each step of all extractions to identify any contamination. Extracts were stored at -20 °C before used.

The isolation of microsatellite polymorphic loci followed the method of enriched genomic libraries (Billotte et al. 1999), with two biotinylated microsatellite probes of motifs (CT)₈, (GT)₈, (GATA)₄ and (GACA)₄. The selected fragments were amplified by PCR, cloned into a pGEM-T vector (Promega), transformed into chimiocompetent Escherichia coli XL1-Blue cells, and cultivated in agar plates containing 100 mg/ml of X-galactosidase and ampiciline for the blue/white selection of positive clones. A total of 272 clones containing inserts were sequenced with the vector's primers (M13F: 5'-CGCCAGGGTTTTCCCAGTCACGAC-3' and M13R: 5'-TCACACAGGAAACAGCTATGAC-3') using ABI Prism Big Dye Terminator Cycle Sequencing Kit (Applied Biosystems) and run in ABI 3100. Forty sequenced clones containing microsatellite motifs were found using the Gramene Project SSR Tool (Ware et al. 2002). From those sequenced clones, 34 primer pairs could be designed. PCR conditions were optimized using fresh tissue samples. These were also used to screen for polymorphism and to have a reference for the general composition of alleles and their range. Software PeakScanner v. 1.0 (Applied Biosystems) was used to inspect the peaks. PCR was performed in 25 µL reactions containing 3.0 mM MgCl₂, 0.2 mM dNTPs, 0.1 pmol of fluorescent labeled primers, 0.1 unit/µL of GO-Taq (Promega) and 1-2 µL of template DNA. Cycling conditions consisted of an initial denaturation at 95°C for 11', 35 cycles at 94°C for 1', 54-60 °C for 1', 70 °C for 1' and a final extension at 60 °C for 60'. DMSO (up to 5%) was occasionally used as adjuvant. For historical and noninvasive samples, bovine serum albumin (BSA, 1.6%) was routinely used as adjuvant. Finally, nine polymorphic microsatellite loci were identified, namely H5, E12, G3, E3, F1R, B2, C10, A9 and A8 (Table 1).

| Locus | Repeat Structure | Primer Sequence (5'-3') | Annealing Temperature (°C) | Size Range (bp) | No. of Alleles | No. of Effective Alleles | Dye Label |
|-------|---------------------|-------------------------|-------------------------------|-----------------|-------------------|--------------------------------|--------------|
| H5 | (TC) ₆ | F-CCCGCAGTATAGAAGCAG | 54-60 | 214-216 | 2 | 1,019 | FAM |
| | | R-CCACGTCACAATCACCT | | | | | |
| E12 | (GT) ₁₀ | F-GGGTTTCCTGTCCCCATTAT | 54-60 | 225-237 | 7 | 4,477 | VIC |
| | | R-GTTCTCCTTGGGAAGCTG | | | | | |
| G3 | (GT) ₁₈ | F-TGGACCCGCCATATAAACAT | 54-60 | 174-222 | 19 | 9,039 | NED |
| | | R-TGGACTAACTGGGCTTCTGC | | | | | |
| E3 | (GT) ₂₃ | F-CACCACGACACCACACTACC | 54-60 | 104-134 | 13 | 1,762 | VIC |
| | | R-TGCTTACGCGTGGACAAAT | | | | | |
| F1R | (CA) ₈ | F-TCCCTAGGGCATCATCGTTA | 54-60 | 208-216 | 5 | 1,299 | VIC |
| | | R-AGCAGCCACGTTCTCAGACT | | | | | |
| B2 | (TG) ₉ | F-CCTTTGGGTCCTGATTGAGA | 54-60 | 191-233 | 23 | 11,184 | NED |
| | | R-AATGGTGGGGGCACTAAGATG | | | | | |
| C10 | $(CA)_6$ | FCTGGCCCTTAGCAGGGTTAT | 58 | 166-176 | 6 | 1,700 | PET |
| | | R-TCTGGTTTCAGGAAGGGTT | | | | | |
| A9 | (AC) ₈ | F-TCCAAGTCTCAGGTCCCAT | 54-60 | 158-194 | 9 | 1,787 | VIC |
| | | R-TGTGAGCCACTGATCGTGTT | | | | | |
| A8 | (TGTC) ₈ | F-ACAGGCTGTTTGAGTGCCA | 56 | 179-199 | 7 | 2,622 | FAM |
| | | R-CCACTGGCACGTTATCGTTT | | | | | |

Table 1 Characteristics of the nine microsatellite loci isolated for *Tamandua tetradactyla*.

The sequences corresponding to each microsatellite locus were deposited in Genbank (http://www.ncbi.nlm.nih.gov/genbank/) under the accession numbers KF746177-KF746185. These loci were tested for linkage disequilibrium (LD) using software Arlequin v. 3.5 (Excoffier & Lischer 2010) with the fresh tissue samples, and no significant and/or consistent LD was found, so they were used for genotyping historical and noninvasive samples. All samples were amplified with 3-5 independent PCR replicates (multiple tube approach). Alleles were considered as "real alleles" if they appeared at least twice in independent PCR replicates (e.g., outliers in microsatellite range were identified taking the fresh tissue samples as references). After assigning a consensus genotype based on simultaneous observation of peaks for all replicates, scores were given for each replicate. When a genotype in one repeat was identical to the inferred true genotype, a score of '1' was assigned. Otherwise, if this genotype presented allelic dropout (ADO), false allele (FA) or even failure of amplification, a score of '0' was assigned. This procedure was done for all PCR replicates. The amount of score '1' amplifications was divided by the total amount of replicates. Then, QIs were calculated for each sample, each locus, and globally, based on the average of obtained values, according to descriptions in Miquel et al. 2006, for the historical and noninvasive sample sets separately. The QI value can vary from zero to one. ADO was considered when homozygote amplification was found in a sample that consistently showed heterozygote genotypes for other different independent amplifications. FA was considered when any allele other than those considered "real alleles" was found.

The package *stats* implemented in R Project for Statistical Computing program (R Development Core Team, 2008) was used to conduct statistical analysis of data. To compare QI values for historical and noninvasive datasets at loci and samples level, a t-test was performed (95% confidence interval). Pearson's r correlation index was used to test the strength of association

between the age of study skin or plucked hair samples (time since it was collected, prepared and stored) and QI values, representing DNA quality. The significance was tested at a 95% confidence interval. Different procedures of study skin treatment and storage can affect their performance for molecular studies. To test if the QI values for historical samples (study skins) had more influence from the age of collection or from the museum they were stored, an ANCOVA test was performed (the four different museums as treatments, age and QI as quantitative variables).

To verify the occurrence of null-alleles in microsatellite genotyping of all sample sets (historical, noninvasive and fresh tissue samples) the software Micro-Checker v. 2.2.3 (Oosterhout *et al.* 2004) was used (set to 95% confidence interval with 1000 randomizations). According to Oosterhout *et al.* 2004, estimations of null alleles are more accurate if the sample set used represents a natural population, since it relies on Hardy-Weinberg equilibrium (HWE) estimates of heterozygotes and homozygotes. Thus, each sample set representing a natural population was used (from historical samples, a subset of Paraná state, n=29; for noninvasive samples, a subset of Mato Grosso do Sul state, n=8; and for fresh tissue samples a subset of Minas Gerais state, n=13) (Table S1). These sample sets representing natural populations were also used for estimating observed and expected heterozygosity (H_0 and H_E), using Arlequin v. 3.5 (Excoffier & Lischer 2010).

Results

Most of the fresh tissue samples amplified successfully for all nine loci (18 out of 23, 78.3%), with a mean amplification success of 97.1%. Due to the high percentage of amplification success and absence of observed ADO and FA in fresh tissue samples, data regarding genotyping errors and QI values will be presented solely to historical and noninvasive sample set. A total of 138

historical samples were analyzed at nine microsatellite loci. Of these, we used only 101 samples that amplified for five or more loci for further analyses. All specimens used in this study, as well as natural history collections of origin, individual ages (years since collection) and QI values were assembled in the Table S1 (Supporting Information). In total, 24 (23.8%) historical samples amplified successfully for all loci. The mean amplification success for all historical samples per locus was 77.4%. All 19 noninvasive samples were used, of which ten (52.6%) amplified successfully for all nine loci. The mean amplification success per locus was 89.5%. In general, historical samples had a reduced amplification success across microsatellite loci and fewer complete individual genotypes than noninvasive samples (Fig.1).



Figure 1 Number of loci amplified, from a maximum of nine to a minimum of five, for historical, N=101, and non-invasive samples, N=19 (left); locus quality index values (QI) for historical and non-invasive sample sets (right).

During the calculation of QI for historical samples, some individuals could not be assigned to any consensus genotype due to the presence of ADO, FA or non-amplification of individual repeats (from 0% in A8 to 16.8% in E3). ADO ranged from 1.3% in A8 and A9 to 4.3% in G3 and C10. FA ranged from 1.7% in H5 and E12 to 7.6% in E3. The error rates were also observed for noninvasive samples, but in a lower rate: individuals with no consensus genotype ranged from 0% to 5.3% in E12 and C10, FA ranged from 0% to 7% in C10 and ADO ranged from 0% to 5.3% in E12 and C10, FA ranged from 0% to 7% in C10 and ADO ranged from 0% to 5.3% in E12 and A9 (Table 2). The same pattern emerged from QI values. For each locus the lowest QI values were observed for historical samples (Table 2, Figure 1). The QI value for each sample ranged from 0.26 to 0.92 in historical and from 0.50 to 0.93 in noninvasive samples. The QI values for samples and loci were both significantly higher for noninvasive samples than for historical samples (samples, p-value = 0.000289; loci, p-value = 0.001217) (Figure 2). The global QI across all samples and loci was 0.58 for historical samples and 0.78 for noninvasive samples, reflecting the overall tendency.



Figure 2 Differences in QI values for loci (p-value = 0.001217) and samples (p-value = 0.000289), between historical and noninvasive datasets.

Table 2 Amplification success, quality index (QI), error rate (Allele Dropout, ADO; False Allele, FA) and samples without a consensus genotype for each locus per sample type (historical/non-invasive). All values are percentages, except for QI.

| Locus | Н5 | E12 | G3 | E3 | F1R | B2 | C10 | A9 | A8 |
|--------------------------|-----------|-----------|-----------|-----------|-----------|-----------|----------|-----------|-----------|
| Amplification Success | 98.0/100 | 71.3/84.2 | 74.3/89.5 | 85.2/94.7 | 72.3/94.7 | 66.3/73.7 | 96.0/100 | 75.2/94.7 | 57.4/73.7 |
| Locus QI | 0.75/0.89 | 0.67/0.77 | 0.59/0.7 | 0.58/0.79 | 0.64/0.81 | 0.64/0.68 | 0.75/0.8 | 0.69/0.79 | 0.61/0.69 |
| ADO | 2.6/0 | 3.9/0 | 3.9/0 | 1 1/18 | 1 5/18 | 2.3/0 | 43/70 | 1.3/0 | 13/18 |
| FA | 1.0/0 | 10/53 | 1 5/6 | 1 2/5 | 13/18 | 3.6/0 | 2.6/0 | 2.3/0 | 0/0 |
| No consensus | 4.9/0 | 39/53 | 12.9/0 | 16.8/0 | 5.6/0 | 10.5/0 | 1 16/53 | 4.9/0 | 0/0 |

There was no significant correlation between the age of historical samples (time since the study skins were prepared and deposited in the collection) and their QI values (r=0.0420774, p-value = 0.6954), while the correlation was negative and significant when analyzing noninvasive samples age (time since the plucked hair was collected and stored, r=-0.5057731, p-value = 0.02716) (Figure 3).



Figure 3 Relationship between the age (years) of samples and their QI, evaluated through Pearson's correlation coefficient r, for historical samples (r=0.0420774, p-value = 0.6954) and non-invasive samples (r=-0.5057731, p-value = 0.02716).

The noninvasive dataset (plucked hair) had an outlier in terms of age: while most of the samples were a few months to two years old, one sample was six years old (TTPR01, Table S1, Supporting Information). To check if this particular individual was biasing the correlation, we also performed the correlation test excluding this sample. The results remained the same, showing even a stronger negative correlation (r=-0.7193593, p-value = 0.0007655, Figure S1, Supporting Information). The ANCOVA test for the influence of the natural history collections in the QI of samples showed that there was no response of the quality index values to the institution from which it was collected (F=1.0086, p-value=0.3931). When plotting QI values against age (years) taking into account the institution of origin of the samples, no linear relationship is found, and samples from the same origin and age achieve a wide range of QI values (Figure S2, Supporting Information).

The software Micro-Checker showed the presence of null alleles in E3 locus within the fresh tissue sample set from Minas Gerais. For historical samples, four of the nine microsatellite loci showed evidence of null alleles in Paraná population: E12, E3, B2 and A9. For the latter, it also suggested stuttering effect, which we judged not to be true, once the visualization of peaks for this locus was very clear and easy to assign. In noninvasive samples, the subset from Mato Grosso do Sul presented null alleles only in locus E3 (Table 3). Estimates of H_E and H_O in each of the populations showed significant deviations from HWE expectations for several loci in all sample sets (Table 3).

Table 3 Summary information of genotyped loci for population subsets of fresh tissue (N=13), historical (N=29) and noninvasive samples (N=8). *Significant p-values on a 0.05 confidence level. **Locus H5 was monomorphic within the population sampled for noninvasive dataset.

| Locus | No. of | Null | H_O | H_E | p-value | |
|----------------------|--------|------|---------|---------|----------|--|
| Fresh Tissue Samples | | | | | | |
| H5 | 2 | No | 0.08696 | 0.08502 | 1.00000 | |
| E12 | 5 | No | 0.52174 | 0.56522 | 0.23394 | |
| G3 | 9 | No | 0.50000 | 0.83827 | 0.00005* | |
| E3 | 4 | Yes | 0.27273 | 0.49049 | 0.00028* | |
| F1R | 2 | No | 0.26087 | 0.23188 | 1.00000 | |
| B2 | 10 | No | 0.50000 | 0.82664 | 0.00120* | |
| C10 | 2 | No | 0.08696 | 0.16232 | 0.13266 | |
| A9 | 5 | No | 0.30435 | 0.31111 | 0.36815 | |
| A8 | 4 | No | 0.20000 | 0.59231 | 0.00010* | |
| Historical Samples | | | | | | |
| H5 | 2 | No | 0.03448 | 0.03448 | 1.00000 | |
| E12 | 7 | Yes | 0.53846 | 0.78054 | 0.05794 | |
| G3 | 9 | No | 0.43478 | 0.86957 | 0.00002* | |
| E3 | 4 | Yes | 0.27778 | 0.57302 | 0.00376* | |
| F1R | 3 | No | 0.16667 | 0.23050 | 0.14900 | |
| B2 | 15 | Yes | 0.65217 | 0.91594 | 0.00000* | |
| C10 | 5 | No | 0.40741 | 0.52760 | 0.00673* | |
| A9 | 4 | Yes | 0.15385 | 0.46078 | 0.00000* | |
| A8 | 3 | No | 0.36842 | 0.39118 | 0.71062 | |
| Noninvasive Samples | | | | | | |
| H5** | 1 | - | - | - | - | |
| E12 | 4 | No | 0.33333 | 0.75758 | 0.04698* | |
| G3 | 5 | No | 0.40000 | 0.84444 | 0.03370* | |
| E3 | 5 | Yes | 0.25000 | 0.53333 | 0.01448* | |
| F1R | 2 | No | 0.28571 | 0.43956 | 0.44130 | |
| B2 | 6 | No | 0.60000 | 0.88889 | 0.14675 | |
| C10 | 3 | No | 0.25000 | 0.24167 | 1.00000 | |
| A9 | 4 | No | 0.50000 | 0.59167 | 0.19607 | |
| A8 | 2 | No | 0.25000 | 0.25000 | 1.00000 | |

Discussion

We have analyzed the genotyping success and errors in 101 historical samples and 19 noninvasive samples using nine newly described polymorphic microsatellite loci for the lesser anteater. This is the first study to describe species-specific genetic markers for *T. tetradactyla*.

Our results show that the quality of plucked hair is legitimately superior to that of study skin samples, as demonstrated by the higher QI global value for loci and samples (Figure 1, Figure 2), higher mean amplification success (89.5%), lower error rates (Table 2), and statistical support. Even though we compared two different sample sizes, we feel confident with the results, since QI calculation is entirely based in weighted averages, and we had significant differences showed by statistical results. Given the clearly distinct tendency found in noninvasive samples, we would expect to observe the same pattern even if this sample size increased.

The mean amplification success observed for historical samples (77.4%) was concordant with what have been reported in literature (40-88% in Rohland *et al.* 2004; 64% in Moraes-Barros & Morgante 2007; 74% in Arandjelovic *et al.* 2009; 43-74% in Polanc *et al.* 2012). However, genotyping errors were high and frequently the consensus genotypes could not be inferred.

The QI values of historical samples varied greatly, and many of them were low. Miquel *et al.* (2006) argued that a threshold above which samples should be considered cannot be defined *a priori*, and depends on the objectives of each study. This value may be different depending on the type of study performed (*e.g.* QI>0.625 in Miquel *et al.* 2006; QI>0.4 in

51

Polanc *et al.* 2012). For a threshold of QI>0.5, 13 historical samples would be excluded, and 88 (87.1 % of total) would pass the criteria for further use. As expected, increasing this threshold would decrease the number of samples for further use, *e.g.*, with QI>0.7, 77 samples would remain, and QI>0.8 only 64 (63.4% of total) samples would remain. Using the latter QI threshold, the selected samples show more complete individual genotypes (less missing data), and therefore, they would more suitable to further population genetics analysis. That shows, as pointed by Miquel *et al.* 2006, that QI is a valuable measure to design experiments from a pilot study and predict how many samples must be collected as a function of the average obtained QI. It is important to notice, however, that the low QI values found in historical samples, as well as the low amplification success can be related to the type of material used, *i.e.*, study skins. These are known to perform worse than other types of material, such as bones and claws (Casas-Marce *et al.* 2010). When there is an alternative to use of dried skins, it should be preferred.

The amplification success for noninvasive samples (89.5%) was also concordant with literature (95.2% in Luikart *et al.* 2008; 90% in Arandjelovic *et al.* 2009; 71-100% in Mondol *et al.* 2009; 72-88% in Boston *et al.* 2012). Concerning the QI threshold, all samples were above QI>0.5, 18 (94.7% of the total) had a QI>0.7 and 15 (78.9%) had a QI>0.8. Over half of the samples had the complete genotype for all microsatellite loci (Figure 1), confirming that a noninvasive sampling approach such as the use of plucked hair is valid for molecular biology studies.

The quality of historical samples depicted by QI values showed not to be correlated with the time since the skins were dried (r=0.0420774, p-value = 0.6954). Even though it seems logical to think that more recent specimens should yield better results than older ones, one

major factor affecting the DNA conservation through the years is actually the type of preservation treatment of the hides (Wandeler *et al.* 2007), as well as the conditions in which they are maintained (presence of humidity, control of temperature, regular care, *etc*). Depending on that, a higher level of contamination and PCR inhibition may be encountered (Hall *et al.* 1997). Therefore, great variance of DNA quality is frequently found among historical samples, not essentially correlated with their age. In fact, among the samples excluded, *e.g.*, on a QI>0.5 threshold, there were old samples (from 1908 to 1950) and relatively new samples (from 2002 to 2008). However, we could not infer a relationship between the effect of the each natural history collection (different preparation and storage) and the quality of the samples. We had no data on how the skins were prepared in each institution (*e.g.*, protocols for skin preparation, if the procedure changed over time, or if conditions of storage were altered), and, therefore, we cannot make statements on this problem. Nevertheless, this issue should be further investigated so that studies using such sample type can account for this effect.

Conversely, the quality of DNA from plucked hair (noninvasive samples) seems to be dependent on time since it was collected. The negative and significant correlation found (r=-0.5057731, p-value = 0.02716, and after exclusion of outlier: -0.7193593, p-value = 0.0007655) indicates that the older the samples get (increasing number of years since collection), the worst the DNA quality gets. This could be explained by the lack of conservation methods applied to this type of material, since plucked hair samples are usually kept in closed recipients without any further protective reagent, such as ethanol, buffers, or others. Also, it is important to notice that the exposure to environmental conditions after collection, such as high temperatures and humidity, can affect the

performance on molecular studies (Kelly *et al.* 2012). Since our samples were all collected the same way, placed in closed recipients since collection and kept at constant temperatures (except for transportation time), we believe our results were not significantly influenced by this factor.

In spite of showing better overall quality over historical samples, noninvasive samples should be used right away after collected to guarantee a reasonable performance on molecular studies. However, we would like to emphasize that these results apply to plucked hair samples only, which were the aim of this study. Other sources of noninvasive samples (for example, scats), may behave in a different way, and should be handled in an appropriate manner.

Historical samples showed the presence of null alleles in four microsatellite loci. It is known that null alleles can generate a bias in F_{ST} and genetic distances overestimating these measures as they create false homozygotes (Chapuis & Estoup 2007). Furthermore, studies show that loci affected by null alleles probably do not alter the overall outcome of assignment testing, although it lowers the power of assignment tests and accuracy of F_{ST} (Carlsson 2008). These problems can be addressed by accounting for the downward bias resulting from the null alleles (*e.g.* software Micro-Checker). Adjusting the allele and genotype frequencies of the amplified alleles to account for the excess of homozygotes can allow the use of these loci in analysis that require frequency estimates of natural populations (not individual-based analysis, once genotypes are not modified). Nevertheless, caution should be taken with this approach because it assumes, necessarily, that the population is in Hardy-Weinberg equilibrium (Oosterhout *et al.* 2004). Ideally, microsatellite loci less susceptible to null alleles should be preferred. Noninvasive samples

54

showed null alleles only in one locus, E3, which was also indicated for historical samples and for fresh tissue samples, the latter being the most appropriated source of comparison, for showing the best genotyping results. Moreover, E3 locus also showed the highest occurrence of FA and individuals without an inferred consensus genotype (Table 2). Thus, considering its low QI as well, we judge it would be best to discard this locus for further population genetic analysis, since it could add bias to results. Historical samples showed both the highest amount of null alleles and lowest QI values. It is also possible that the higher occurrence of null alleles in historical samples is an effect of the different samples sizes of datasets, and different characteristics of the populations used for calculation (discussed below), and may be considered an artifact.

Several loci showed significant deviations from HWE in all sample sets, as displayed in Table 3. These results are difficult to evaluate, once the event causing such deviation may come from different sources. Besides technical issues (*e.g.* genotyping errors, addressed in this paper) there may be also violations from HWE assumptions in the populations used for testing, such as cryptic genetic substructure, unstable population size and small sample sizes, which may hinder p-value estimates (Hartl & Clark, 2006).

Conversely, it has been shown that even small levels of laboratory errors, such as very few homozygote individuals for a rare allele, can result in an overestimation of HWE deviations (Morin *et al.* 2009). We have no reason to believe that few individuals in the historical sample set could alter the general results, once the HWE test was also performed excluding the samples with QI below 0.5, 0.7 and 0.8, successively, and the results remained unaltered (data not shown).

Deviations of HWE found in loci among different datasets can also have influenced our estimates of null alleles. The excess of homozygotes observed (Table 3) may reflect violations of assumptions as described above, and/or presence of null alleles, since its detection also takes into account HWE calculation. The populations of each dataset used for HWE estimations may have violated one or more assumptions. For example, the population of historical data set contained individuals from various ages, some overlapping generations in the species (Table S1, Supporting Information). When performing the test again with a group of only 18 samples, all under 10 years old, the results remained equal (the same loci showed HWE deviation, data not shown). This problem was not encountered in noninvasive and fresh tissue datasets. On the other hand, we cannot guarantee the individuals are completely unrelated, once anteaters present lonely behavior, and unless long term observations are done, simply field collection in different localities is not enough to exclude kinship. Therefore, due to these restrictions, the estimation of null alleles may not be comparable between datasets, as these estimates may be overestimated reflecting the dataset characteristics. In this case, QI, ADO and FA estimates may be better sources of comparison between datasets.

Conclusions

In this study we presented a comprehensive comparison of the performance of historical (study skins) and noninvasive (plucked hair) samples through the genotyping of newly described polymorphic microsatellite loci for the lesser anteater. We have reported the superior quality of plucked hair over study skins, and depicted the main error sources in each sample set. We have confirmed that QI measures are valuable to identify problematic samples and loci, and select the best ones for further population genetic analysis. Even

though we detected different error types and the presence of null alleles in historical samples (as well as in noninvasive and fresh tissue, but in a smaller scale) we do not discourage their use in molecular studies, once they represent a valuable source of populations that no longer exist or that have undergone changes through time. We recommend a careful inspection of data, excluding non-satisfactory samples, and applying appropriate methods to judge the results when performing analysis to address population genetic questions.

Acknowledgements

We would like to thank Flávia Miranda, Caetano T. Oliveira, Ecopoint/CE and CETAS/BA for help with noninvasive sample collection, Gilda Tebet from MHNCI, Maurício Graipel from UFSC, Stella Franco and João A. de Oliveira from MN and Hélio Boudet from MBML for help with historical sample collection. We are grateful to Sibelle T. Vilaça for the help with microsatellite development. C.L. We would also like to thank three anonymous reviewers that helped to substantially improve this work. Clozato and N. Moraes-Barros were supported by CAPES, F.R. Santos by FAPEMIG and CNPq, and J. S. Morgante was granted with a research award from FAPESP (08/52207-0).

References

- Arandjelovic M, Guschanski K, Schubert G *et al.* (2009) Two-step multiplex polymerase chain reaction improves the speed and accuracy of genotyping using DNA from noninvasive and museum samples. *Molecular Ecology Resources*, **9**, 28–36.
- Beja-Pereira A, Oliveira R, Alves PC, *et al.* (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources*, **9**, 1279-1301.
- Billotte N, Lagoda PJL, Risterucci AM *et al.* (1999) Microsatellite enriched libraries: applied methodology for the development of SSR markers in tropical crops. *Fruits*, **54**, 277–288.
- Bonin A, Bellemain E, Bronken Eidesen P *et al.* (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261-3273.

- Boston ESM, Puechmaille SJ, Scott DD *et al.* (2012) Empirical Assessment of Noninvasive Population Genetics in Bats: Comparison of DNA Quality from Faecal and Tissue Samples. *Acta Chiropterologica*, **14**, 45–52.
- 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.
- Casas-Marce M, Revilla E, Godoy J (2010) Searching for DNA in museum specimens: a comparison of sources in a mammal species. *Molecular Ecology Resources*, **10**, 502–7.
- Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–31.
- Gardner AL (2005) Order Pilosa. In: *Mammal species of the world: a taxonomic and geographic reference*. 3 ed. (eds Wilson DE, Reeder DM), pp 98-102, The John Hopkins University Press, Baltimore.
- Hall ML, Willcox MS, Jones DS (1997) Association of enzyme inhibition with methods of museum skin preparation. *BioThecniques*, **22**, 928-934.
- Hartl DL, Clark AG (2006) Testing Hardy-Weinberg Equilibrium. In: *Principles of population genetics* (Daniel L Hartl, Andrew G Clark) pp. 70-82. Sinauer Associates, Massachusetts.
- Hayssen V (2011) Tamandua tetradactyla (Pilosa: Myrmecophagidae). Mammalian Species, 43, 64–74.
- Kelly C, Mateus-Pinilla NE, Douglas M *et al.* (2011) Microsatellites behaving badly: empirical evaluation of genotyping errors and subsequent impacts on population studies. *Genetics and Molecular Research*, **10**, 2534–53.
- Kelly MJ, Betsch J, Wultsch C *et al.* (2012) Noninvasive sampling for carnivores. In: *Carnivore Ecology and Conservation, a handbook of techniques* (eds Luigi Boitani and Roger A. Powell), pp 49-54. Oxford University Press, New York.
- Luikart G, Zundel S, Rioux D *et al.* (2008) Low genotyping error rates and non-invasive sampling in bighorn sheep. *Journal of Wildlife Management*, **72**, 299-304.
- Miquel C, Bellemain E, Poillot C *et al.* (2006) Quality indexes to assess the reliability of genotypes in studies using noninvasive sampling and multiple-tube approach. *Molecular Ecology Notes*, **6**, 985–988.
- Mondol S, Karanth KU, Kumar NS *et al.* (2009) Evaluation of non-invasive genetic sampling methods for estimating tiger population size. *Biological Conservation*, **142**, 2350-2360.
- Montgomery GG, Lubin YD (1977). Prey influences on movements of neotropical anteaters. In: *Proceedings of 1975 Predator Symposium*. (eds Philips RL, Jonkel C), pp 103-131, University of Montana, Missoula.
- Moraes-Barros N de, Morgante JS (2007) A simple protocol for the extraction and sequence analysis of DNA from study skin of museum collections. *Genetics and Molecular Biology*, **30**, 1181–1185.

- Morin PA, Leduc RG, Archer FI *et al.* (2009) Significant deviations from Hardy-Weinberg equilibrium caused by low levels of microsatellite genotyping errors. *Molecular Ecology Resources*, **9**, 498-504.
- Oosterhout C Van, Hutchinson WF, Wills DPM *et al.* (2004) Micro-Checker: Software for Identifying and Correcting Genotyping Errors in Microsatellite Data. *Molecular Ecology Notes*, **4**, 535–538.
- Polanc P, Sindičić M, Jelenčič M *et al.* (2012) Genotyping success of historical Eurasian lynx (*Lynx lynx L.*) samples. *Molecular Ecology Resources*, **12**, 293–298.
- Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–859.
- R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
- Rohland N, Siedel H, Hofreiter M (2004) Nondestructive DNA extraction method for mitochondrial DNA analyses of museum specimens. *BioTechniques*, **36**, 814-818.
- Shaw PW, Pierce GJ, Boyle PR (1999) Subtle population structure within a highly vagile marine invertebrate, the valued squid *Loligo forbesi*, demonstrated with microsatellite DNA markers. *Molecular Ecology*, **8**, 407-417.
- Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution*, **14**, 323–327.
- Taberlet P, Griffin S, Goossens B *et al.* (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, **24**, 3189–94.
- Waits LP, Paetkau D (2005) Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, **69**, 1419-1433.
- Wandeler P, Hoeck PE, Keller LF (2007) Back to the future: museum specimens in population genetics. *Trends in Ecology and Evolution*, **22**, 634–642.
- Ware DH, Jaiswal P, Ni J, Yap IV, Pan X, *et al.* (2002) Gramene, a tool for grass genomics. *Plant Physiology*, **130**, 1606–1613.
- Wetzel RM (1975) The species of *Tamandua* Gray (Edentata, Myrmecophagidae). *Proceedings of the Biological Society of Washington*, **88**, 95-112.

Capítulo 2

Diversidade genética e estrutura populacional de um mamífero amplamente distribuído da América do Sul, o tamanduá-mirim (*Tamandua tetradactyla*, Pilosa).

Genetic diversity and population structure of a wide-distributed mammal from South America, the lesser anteater (*Tamandua tetradactyla*, Pilosa).

Apresentação Capítulo 2

Este capítulo apresenta dados de marcadores genéticos nucleares biparentais do tipo microssatélites genotipados em indivíduos ao longo de grande parte da distribuição de tamanduá-mirim, abrangendo os maiores biomas da América do Sul (Floresta Amazônica, Mata Atlântica, Cerrado, Pantanal e Caatinga).

Com estes marcadores neutros exploramos as assinaturas de eventos demográficos e a influência da deriva genética nas populações da espécie, refletidas na forma de diversidade genética e estrutura populacional.

Devido à dificuldade de assinalar populações em uma espécie amplamente distribuída e cuja amostragem é esparsa e oportunística, duas abordagens foram usadas para entender a estruturação genética: agrupamento de indivíduos pelo critério de proximidade geográfica, designado aqui como *a priori*, e análises baseadas no indíviduo, sem informação prévia de agrupamento populacional, referida no capítulo como *a posteriori*.

Genetic diversity and population structure of a wide-distributed mammal from South America, the lesser anteater (*Tamandua tetradactyla*, Pilosa).

CAMILA L. CLOZATO¹, NADIA DE MORAES-BARROS^{1,2}, JOÃO S. MORGANTE¹

Keywords: Atlantic Forest; Amazon Forest; genetic structure; isolation by distance; landscape genetics; microsatellites; South America; *Tamandua tetradactyla*.

(Artigo em preparação)

Abstract

The lesser anteater, Tamandua tetradactyla, is a widespread medium sized mammal of South America that inhabits a variety of different landscapes. In this study we investigated the extant genetic structure of the species and the main evolutionary processes that shaped its diversity. This is the largest genetic study performed for this vagile mammal to date. We used eight polymorphic microsatellite loci and an extensive sampling across the distribution of the lesser anteater, crossing major Brazilian biomes: Amazon Forest, Cerrado, Pantanal, Caatinga and Atlantic Forest. We used two approaches: i) a geographical criterion to assemble individuals into populations, and test differentiation among a set of predefined populations; and ii) an individual-based, spatially explicit Bayesian clustering model to test the major genetic breaks within the sampling area in different geographical transects. We found moderate levels of genetic diversity, and low structuring among geographical localities. Individuals/populations that showed greatest distinctiveness were originated from the Amazon Forest. This corresponded to the first detected split in the genetic composition of the species throughout the sampled distribution. This region also showed the highest genetic diversity, which, coupled with studies of morphological evidences, indicates a center of diversification for the species. Populations distributed across the coast, in Atlantic Forest, displayed a pattern of isolation by distance. Central Brazil biomes, Cerrado and Pantanal, showed no distinction from the surrounding populations, and are likely a past connection between adjacent forested biomes.

Introduction

In natural populations, individuals are often distributed discontinuously across demes, and they are frequently exchanged between adjacent or nearby populations. Particularly for widespread species, the existence of geographical (e.g. mountains, rivers) or environmental barriers (e.g. humidity) favors the differentiation of its populations and the formation of geographical delimited lineages or even new species (Kimura & Weiss 1964; Manel et al. 2003). However, in the absence of such barriers, even widely distributed species may be prevented from constituting a panmitic unit when individuals have limited migration ability, and then a pattern of isolation by distance is likely to be observed (Wright 1943, Malecót 1948). The detection of genetic discontinuities (whether they are generated by barriers or by isolation by distance) and its correlation with landscape features can be addressed by classical population genetics methods, which imply defining populations and measuring genetic differentiation between them. However, when a species is widely distributed in a continuous space, using the individual as the operational study unit may be advantageous to avoid potential bias in identifying populations in advance, such as it is applied in landscape genetics methods (Manel et al. 2003; Manel and Holderegger 2013; Cushman et al. 2012).

Many studies have focused on understanding the partition of genetic diversity in widespread mammals and its correlation with major biogeographical patterns in the north hemisphere (*e.g.* Worley *et al.* 2004; Pilot *et al.* 2006). Some studies have also been done in the African continent (*e.g.* Simonsen *et al.* 1998; Wilson & Strobeck 1999), but few attempted to investigate these issues in South America (*e.g.* Moraes-Barros *et al.* 2006; Tchaicka *et al.* 2007; Martins *et al.* 2009). Even though the Neotropics present high species

richness, astonishing diversity, and most importantly, presents a very diverse array of landscapes, only about 7% of all phylogeography studies were performed with South American mammals (Beheregaray 2008), and the region remains markedly understudied (Ceballos & Ehrlich 2009; Jenkins et al. 2013; Turchetto-Zolet et al. 2013). Apart from the Great American Biotic Interchange (Webb 2006) and the Andean uplift (Vuilleumier & Monasterio 1986), which affected the biota in a continental scale, other key factors were important to shape the species diversity and differentiation throughout America. Some recognized elements are: i) the complex geological development of the Amazon basin and posterior forest expansion, highly influenced by the Andean uplift, that affected a number of taxa and was responsible for a number of diversification events (Hoorn et al. 2010); ii) the open vegetation area called "the major South American Disjunction" (Brieger 1969) crosses South America throughout central Brazil, and is constituted by Caatinga biome, Cerrado (Brazilian savanna-like vegetation), and the wetlands Pantanal and Chaco. The major South American Disjunction presents lower number of mammalian endemism (Redford & Fonseca 1986). and may act as a barrier for some taxa, or as a historical connection between forested biomes for others, presenting a current significant coverage of gallery forests (Costa 2003); and iii) the formation of Pleistocenic refuges, forest remnants during glacial periods in the Quaternary were supposedly responsible for the maintenance, diversification and speciation of several species (Haffer 1969; Carnaval et al. 2009). This important composition of features is thought to have affected in some way most of the living taxa in South America.

The lesser anteater, *Tamandua tetradactyla* (Linnaeus 1758), also known as Southern Tamandua, is a medium-sized mammal of the family Myrmecophagidae, order Pilosa. It is part of one of the most ancient lineages of placental mammals, the Xenarthra group (Murphy *et al.* 2001). The species is mostly nocturnal and solitary, and feeds on a combination of ants and termites with the help of its long tongue (Montgomery & Lubin 1977). Although lesser anteaters show preference for forested areas, it is possible to find it foraging in open grassland savanna-like areas such as Cerrado, wetlands such as Pantanal, in mountain tropical regions and transitional forests (Eisenberg 1989; Mares *et al.* 1996). Therefore, the species has a wide geographic distribution: it occurs in South America east of the Andes, from Venezuela and Trinidad until the north of Argentina, and south of Brazil and Uruguay, in elevations up to 2000 meters (Novak *et al.* 1983; Wetzel 1985; Gardner 2008).

The species present a variety of morphological differences across its range. It is possible to find specimens with a full pale tan fur, full black fur, and individuals with the shape of a dorsal sleeveless black vest, which can also present a number of gradual variations (Wetzel 1985; Hayssen 2011). The great morphological variation in the species led to the description of four subspecies, which are delimitated by geographical boundaries (Wetzel 1985; Gardner 2008). Home range size reported for the lesser anteater in Venezuela was 375 ha (Montgomery 1985) and up to 400 ha in Argentina (Parera 2002), while in Brazil it was only of 100 ha (Rodrigues *et al.* 2008). Individuals tend to adapt quickly to new environments, as demonstrated by a translocation study (Rodrigues *et al.* 2001). Little information regarding ecological features of the species is available, and no long-term studies of population dynamics have been performed, nor has phylogenetic/phylogeography research. Given the continuous range distribution and the limited knowledge about its social organization, the definition of populations for genetic studies is difficult.

As the species is distributed along the most part of South America and presents marked morphological variation, it likely presents some level of genetic structure due to historical barriers to gene flow. Alternatively, it is also possible that diversification among populations could be the result of isolation by distance, considering its adaptation to different environments and the relatively large home range size.

In this study, we investigated individuals across the continental distribution of the lesser anteater, using eight polymorphic microsatellite loci to access major patterns of genetic diversity and structure in this wide distributed mammal. First, we used a geographical criterion to assemble individuals into populations, and test differentiation among a set of predefined populations. Then, we used an individual-based, spatially explicit Bayesian clustering model to test the major genetic breaks within the sampling area in different geographical transects comprising distinct landscape configurations, an approach used in landscape genetics studies (Manel *et al.* 2003; Cushman *et al.*, 2013). Finally, we attempted to depict where isolation by distance or barriers to gene flow played a major role in the genetic structure of the species.

Materials and Methods

Sample Collection

A total of 176 individuals of the lesser anteater distributed along the species range were used for this study (Figure 1). Of these, 114 were obtained from fieldwork and collaborations with other researchers and/or institutions. All biological samples were collected under SISBIO-IBAMA permission number for scientific activities 24001-2/53695225. Samples varied in type: tissue, blood and plucked hair were used. Since the species is characterized by solitary behavior, and low density in the wild (Arita et al. 1990), samples were collected opportunistically, and thus individuals, instead of populations, were considered. The remaining 62 samples were study skins collected in natural history collections in Brazil. The viability for the use of historical samples, as well as plucked hair samples mentioned above (*i.e.*, noninvasive samples), were evaluated previously through genotyping errors estimates and Quality Index (QI) values (Chapter 1) for a description of the QI method, please see Miquel *et al.* 2006). Only museum samples with QI values >0.7and amplification of at least 75% of all microsatellite loci were used for this study. For a complete list of the samples used and its origin, please refer to Table S1, Supporting Information.

Geographical locations were recorded as GPS coordinates during fieldwork or deduced from the approximate location where sampleswere collected (*e.g.*, road killed samples). Since the sampling scale is large, errors in coordinates were considered negligible (but see *Clustering Analysis*). Nevertheless, individuals with doubtful collection site donated from researchers or institutions were excluded beforehand.

DNA Extraction and Microsatellite Analysis

Extraction of total genomic DNA was performed with Proteinase K digestion followed by salt precipitation (Sambrook *et al.* 1989). Additional technical details for the extraction of historical samples are described in Clozato *et al.* 2013, as well as the description of the eight microsatellites used in this study for sample genotyping. Polymerase chain reactions (PCR) consisted of 25 μ L containing 3.0 mM MgCl₂, 0.2 mM dNTPs, 0.1 pmol of fluorescent labeled primers, 0.1 unit/ μ L of GO-Taq (Promega) and 1-2 μ L of template DNA. Cycling conditions consisted of an initial denaturation at 95°C for 11', 35 cycles at 94°C for 1', 54-60 °C for 1', 70 °C for 1' and a final extension at 60 °C for 60'. DMSO (up to 5%) was occasionally used as adjuvant. For historical and noninvasive samples, bovine serum albumin (BSA, 1.6%) was routinely used as adjuvant. All samples were amplified at least twice for each locus (multiple tube approach), to minimize genotyping mistakes. Fragments were analyzed for variation in length in an ABI 3100 (Applied Biosystems) automated sequencer. Software PeakScanner v. 1.0 (Applied Biosystems) was used to inspect the peaks (intensity, height and size).

A priori clustering of individuals

To obtain a first picture of the distribution of genetic diversity through the sampling areas, a geographical approach was used: we grouped neighboring individuals into clusters and considered these as population units. This approach is here defined as an *a priori* analysis. We chose to group individuals in a maximum of 200 km in a straight line, respecting a minimum of three samples per population. Due to these criteria, some individuals were excluded from this initial analysis, either because they were geographically isolated, or

because a minimum number of neighboring individuals was not achieved. A total of 20 populations were defined (names were based on Brazilian federal states, except for Peru population): PA1, Pará 1; AM1, Amazônia 1; AM2, Amazônia 2; PER1, Peru (country) 1; MA1, Maranhão 1; PICE1, Piauí and Ceará 1; CE1, Ceará 1; PB1, Paraíba 1; PE1, Pernambuco 1; AL1, Alagoas 1; BA1, Bahia 1; MT1, Mato Grosso 1; MT2, Mato Grosso 2; MS1, Mato Grosso do Sul 1; MS2, Mato Grosso do Sul 2; SCMS1, Santa Catarina and Mato Grosso do Sul 1; SP1, São Paulo 1; MG1, Minas Gerais 1; ES1, Espírito Santo 1; PRSC1, Paraná and Santa Catarina 1. Figure 1 demonstrates the geographical location of these populations by their codes, and Appendix S1 lists individuals belonging to each *a priori* population.



Figure 1 – Sample sites for the species *Tamandua tetradactyla*. Some individuals were grouped into populations based in geographical vicinity (with a minimum of three samples). Twenty populations were assembled (dashed circles), depicted in the map by their code (*e.g.*, AL1 – the codes are available in the text in Materials and Methods section, and individuals in each population are described in Appendix S1). The remaining individuals were not taken into account in *a priori* population analysis, only in individual-based *a posteriori* analysis. Brazilian biomes are represented by shades of grey decreasing from west (Amazon Forest) to east (Atlantic Forest). Circles represent fresh samples and triangles represent historical samples.

This *a priori* geographical clustering analysis was also used to estimate values of observed and expected heterozygosity (H_O and H_E) and investigate if the populations considered are under Hardy-Weinberg equilibrium (HWE). Allelic richness and values of F_{IS} were estimated with the software FSTAT 2.9.3 (Goudet 1995). These estimates were performed with (i) the total sample set; and (ii) excluding the historical samples to test for bias due to the possibility of temporal structuring of populations. Values of pairwise population genetic differentiation estimated by F_{ST} were also calculated, tested with 10000 permutations for significance. Analysis of molecular variance (AMOVA) was performed to test if a hierarchical grouping could explain most of the genetic variation. For these analyses, the software Arlequin v. 3.5 (Excoffier & Lischer 2010) was used. Bonferroni correction for multiple tests (Rice 1989) was applied with alpha set to 0.05 to determine significance.

A posteriori clustering of individuals

In order to investigate the major genetic breaks along the sampled geographical area without assuming pre-defined populations (which may lead to confounding results, given the extended area of study and difficulty to define natural populations), we used an individual-based, spatially explicit approach, referred here as *a posteriori* clustering of individuals. To understand the genetic variation across the sampling range, we pooled individuals in five different subsets: 1) "All Individuals", comprising all sampled individuals; 2) "Major Upper Transect", individuals from east in the Atlantic Forest (Brazilian northeast coast) to west in the Amazon biome (*circa* 4861 km in straight line); 3) "Lower Transect", individuals from east of the Atlantic Forest in the Brazilian southeast to west in the biome Pantanal (wetlands), reaching a small area in the Amazon Forest (*circa* 2291 km) and 4) "East Transect", individuals within the entire range of the Brazilian east coast, from north to south, representing the Atlantic Forest, and reaching the inner biomes from east to west, comprising Caatinga and a part of Cerrado (*circa* 3106 km in straight)
line). The visual representation of the spatial area comprising these subsets is illustrated in Figure 3. These areas work as transects that cross different landscapes and Brazilian biomes (Atlantic Forest, Cerrado, Pantanal, Amazon Forest and Caatinga), allowing the observation of changes in the genetic composition in these different environments.

Two Bayesian model-based, individual-based clustering algorithms were applied to assign individuals into clusters: the R package Geneland v. 4.0.3 (Guillot *et al.* 2005) and the software Strucutre 2.3.3 (Pritchard *et al.* 2000).

Geneland uses a spatial model to cluster a sample into a certain number of groups in a way that each group is genetically homogeneous, *i.e.*, the clusters inferred represent populations in Hardy-Weinberg equilibrium (HWE), or with minimized HWE deviations. Parameters such as allele frequencies and *a posteriori* maximum number of *K* (inferred maximum number of clusters, K_{MAX}) are estimated simultaneously by MCMC algorithm. We used the uncorrelated frequency model (less prone to departure from model assumptions, such as isolation-by-distance), and performed 15 multiple independent runs allowing K_{MAX} to vary from 1 to 10, with 100000 MCMC iteractions and 1000 thinning, allowing geographical coordinates uncertainties to vary at 0.5 km. We checked for convergence observing if the general patterns found for all runs were similar. Each sample subset (groups described above) was run, and the posterior probability value was used to choose the run and the resulting K_{MAX} . We used the spatial D-model because it is more conservative and prevents from inferring spurious populations (Guillot *et al.* 2005).

Likewise, Structure is a model-based clustering method for inferring population structure from a set of individuals. It assumes the dataset is divided in K populations, and individuals

73

are assigned probabilistically to clusters, which are also assumed to be in Hardy-Weinberg equilibrium. Structure is not a spatially-explicit software. Although it is able to account for the origin of individuals as population flags, no prior information of sampling localities was included in our analysis. We set 5 runs per value of K, varying from 1 to 10, with 100000 iterations and a burn in of 1000, with uncorrelated frequencies. We used the method suggested by Evanno *et al.* (2005) to obtain the most likely value of K for each subset, calculated based on the second order rate of change of the likelihood (Delta K). Results were compared with the outcome of the log likelihood for each K, L(K), a method described in Pritchard *et al.* (2000). The outcomes of both softwares (Geneland and Structure) were displayed in assignment graphics of membership probabilities for each individual of resulting clusters for the chosen value of K_{MAX}.

Since the dataset is composed of both fresh and historical samples, we performed all runs excluding the historical samples of each subset to check for bias of different samples sources on the value of K_{MAX} estimated by both programs. We also performed the entire analysis excluding two microsatellite markers that showed some HWE deviations in the *a priori* clustering (loci E13 and G3, see results), and observed if there were any changes in the clustering results biased by these markers.

FSTAT 2.9.3 (Goudet 1995) was used to calculate expected heterozygosity (H_E) and allelic richness per cluster, *i.e.*, the number of alleles per locus, corrected for differences in sample size using rarefaction to estimate the allelic richness expected for a minimum of individuals per cluster. Allelic richness and expected heterozygosity (H_E) were used as overall predictors of genetic diversity.

Isolation-by-distance and long-term historical divergence

Given the wide distribution range of the species and, similarly, of our sampling, a pattern of isolation-by-distance is expected. Therefore, we tested the statistical significance of correlation between genetic and geographical distances with Mantel tests (Mantel 1967) for each of the subset groups. Again, our approach was based solely on individuals. Since Mantel tests are prone to Type I errors (Mermains 2012; Cushman *et al.* 2013), especially when large geographical distributions are tested, we tested for isolation by distance in "All Individuals" subset, and then in all other subsets separately (as they represent partitions of the entire sampling), to check if the correlation remained (following the strategy suggested in Kuchta & Tan 2005). Moreover, we used Partial Mantel tests (Smouse *et al.* 1986) based on multiple matrix regression to disentangle the effects of isolation by distance and isolation by barriers (clusters formed the third matrix created as a covariate), while controlling for the effect of geographical distance.

We used the *ecodist* package (Goslee & Urban 2007) in R (R Development Team 2008) to test the correlation of a matrix of raw genetic distance and geographical distance (Euclidean distances between each pair of individuals measured in log-kilometers) with 1000 replications and to perform Partial Mantel tests.

Results

A priori clustering of individuals

A total of 116 individuals from 20 populations were used for *a priori* clustering of individuals, and 176 were used for *a posteriori* clustering analysis (Figure 1). All loci were polymorphic, exhibiting four to 24 alleles per locus, with an average of 11.38 (\pm 7.44).

Number of individuals (N) in a priori populations varied from 3 to 15 (Table S2, Supporting Information). Allelic richness was normalized for a population of three individuals and varied from 1.00 to 2.00. F_{IS} values varied from -0.50 to 0.75. When considering all populations, two loci presented significant HWE deviations in more than two populations: E12 (populations PA1, MT1, PRSC1, MS1) and G3 (populations PA1, AM1, PRSC1, MS1). Locus B2, C10 and A9 also showed deviations in two or less populations. In total, five populations (PA1, MT1, PRSC1, MS1, AM2) out of 20 presented some level of HWE deviation in one or more loci (Table S2). Consequently, we performed the same analysis, but considering only fresh samples, *i.e.*, excluding historical samples. Since a minimum of three individuals in the population was required, some of the previous populations were excluded, and 16 remained. Loci E12 and G3 presented fewer significant deviations, while B2 loci presented more (Table S3). There were other deviations in this analysis that were not present previously, which may be the outcome of a reduced N per population. In total, five populations presented some level of HWE deviation in one or more loci (MT1, MS1, MS2, PICE, AM2). Although considering most deviations were due to population assembly itself and the small number of individuals included, instead of intrinsic loci problems, or due to temporal sampling differences, in *a posteriori* analysis (below) we decided to test the exclusion of loci E12 and G3, and we performed analyses with and without historical samples, to check if these markers or samples influenced the partition of genetic clusters.

Several significant F_{ST} values were obtained (58 out of 400 comparisons, *i.e.*, 14,5%), and they were present both between close as well as distant geographical pairs of *a priori* populations (Table S4). The highest significant F_{ST} value was found between two adjacent populations (0.44917, p=0.00175, between BA1 and ES1). Population BA1 also showed a significant and relatively high genetic distance from a close population to the south, MG1 (0.32561, p=0.00977). An additional pair of neighboring populations (MT2 and MS1) showed high significant F_{ST} (0.34541, p=0.00488). Populations PICE1 and PA1 presented significant F_{ST} values against most of its surrounding populations. The lowest F_{ST} value (0.02124, p=0.56836, not significant) occurred between the populations PRSC1 and MT1. A heatmap of significant F_{ST} values between 20 inferred populations is represented in Figure 2.



Figure 2 – Heatmap of pairwise population F_{ST} values in *a priori* analysis. Higher values are represented by darker shades of grey. Only statistically significant values were displayed (p-value <0.05). A complete list of F_{ST} values and respective p-values is available in Table S4.

An AMOVA test was performed based on populations from different biomes and by significant results of pairwise population F_{ST} values. Based on the later, we grouped populations of Amazon biome plus PICE1 population [PER1; AM1; AM2; MT2; PA1; MA1;PICE1]; three groups of populations of the Atlantic Forest northeast, with BA1, and PE1 isolated as different groups due to high differentiation from surrounding demes, and one group comprising the remaining northeast populations [AL1;PB1;CE1]. Finally, a group composed of Atlantic Forest southeast and central Brazil, including also the population from Pantanal [ES1;MG1;SP1;PRSC1; SCMS1; MT1, MS2, MS1]. Even though the hierarquical grouping based on significant F_{STS} revealed a higher percentage of variation explained by these groups than by biomes, both approaches showed a massive variation between individuals within populations, and within individuals. Table 1 summarized these results.

Table 1 – Analysis of molecular variance (AMOVA) of two different groupings of populations, and the percentage of the genetic variation explained by these groups. Pop.=populations, Ind.=individuals.

| | | % Variation Among: | | | | |
|--|---|--------------------|--------------------------|------------------------|------|--|
| Groups | Populations in each group | Groups | Pop. within groups | Ind. within pop. | Ind. | |
| By Biomes: [Amazon] [AtlanticForest][Caatinga] [Cerrado][Pantanal] | [PER1; AM1; AM2; MT2; PA1; MA1][ES1;MG1;SP1;PRSC1; SCMS1;AL1;PB1;BA1;PE1] [PICE1;CE1][MT1;MS2][MS1] | 1,58 | 7,74 | 31,74 | 58,9 | |
| By breaks indicated by significant F_{ST} s | [PER1; AM1; AM2; MT2; PA1; MA1;PICE1][PE1][BA1] [AL1;PB1;CE1][ES1;MG1;SP1; PRSC1; SCMS1; MT1, MS2, MS1] | 4,39 | 5,53 | 31,53 | 58,6 | |

A posteriori clustering of individuals

Geneland and Structure gave relatively concordant results in the majority of the tests performed. The maximum K number of populations (K_{MAX}) chosen from the Average Posterior Probability in Geneland and from the Delta K in Structure were mostly equivalent (k=2). L(K) resulted in a bigger K_{MAX} when testing "All Individuals" (k=4) and "East Transect" (k=3) (Table 2). For every test performed in each of the subsets, with and without historical samples, the K_{MAX} was mostly equal to two clusters. A difference appeared in the subset Major Upper Transect, when considering only fresh samples: Geneland could not distinguish different clusters, while Structure indicated two. This could be the influence of diminishing the number of individuals, especially from the Amazon Forest. Furthermore, all Geneland and Structure runs were tested for the absence of markers E12 and G3, and no alteration regarding the partition of genetic structure could be observed related to these loci, so they were kept in the analysis as well.

Table 2 – Values of K_{MAX} obtained from Geneland and Structure for each subset used for analysis (N, number of samples per subset) inferred after the criterion used by each software to choose the best maximum number of clusters (Average Posterior Probability for Geneland, Delta K and L(K) for Structure). In each case, values are shown for all samples in the subset (Historical+Fresh) followed by values for fresh samples only (Fresh).

| Subsets | | Geneland | Structure | | |
|---|-----|------------------|---------------------------|-----------------|--|
| | | K _{MAX} | K _{MAX} (DeltaK) | $K_{MAX}(L(K))$ | |
| All Individuals (Historical+Fresh) | 176 | 2 | 2 | 4 | |
| All Individuals (Fresh) | | 2 | 2 | 4 | |
| Major Upper Transect (Historical+Fresh) | | 2 | 2 | 2 | |
| Major Upper Transect (Fresh) | | 1 | 2 | 2 | |
| Lower Transect (Historical+Fresh) | | 2 | 2 | 2 | |
| Lower Transect (Fresh) | | 2 | 2 | 2 | |
| East Transect (Historical+Fresh) | | 2 | 2 | 3 | |
| East Transect (Fresh) | | 2 | 2 | 2 | |

Most assignment graphics showed strong consistency across all tests, with and without historical samples, especially for Geneland outputs. Structure revealed stronger mixture among individuals in all transects, detecting more migrants. Overall, Structure's assignment graphics showed a tendency towards the indicated K_{MAX} , except for Lower Transect, which did not show a clear genetic split, although K=2 was most likely. All results described in the next paragraphs regarding graphic assignments are based on Figure 3.

In general, all subsets were partitioned into two genetic clusters. For Geneland, the graphical output of the individual membership probabilities for each subset shows a clear shift in genetic composition from one cluster to another. In subset one (All Individuals), Geneland showed a cluster formed by individuals from the Amazon Forest biome. Most of these differentiated individuals corresponded to populations MT2, AM2 and PER1

described in the *a priori* analysis. However, the other cluster inside this biome is more related to all other individuals.



Figure 3 – In the left, visual representation of subsets: 1, "All Individuals"; 2, "Major Upper Transect", from east in Atlantic Forest (Brazilian northeast coast) to west in the Amazon biome (*circa* 4861 km in straight line); 3, "Lower Transect", from east in Atlantic Forest (Brazilian southeast region) to west in the Pantanal biome (*circa* 2291 km); 4, "East Transect", samples within the range of the Atlantic Forest and Cerrado biomes from north to south (*circa* 3106 km in straight line). In the right, assignment graphics of individuals from population membership probabilities represented by two shades of grey. Each line represents one individual. Above graph shows Geneland results, followed by Structure results below.

Structure could detect the same shift in the genetic composition, but it identified much more mixture then Geneland, especially in individuals from Pantanal, and scattered individuals from the Atlantic Forest biome (light-grey lines in "All Individuals"). Structure also detected differentiation inside Amazon Forest biome, although individuals seemed to be more genetically homogenous than in Geneland results. L(K) results from Structure maintain the overall clustering pattern, but it further splits Amazon individuals in two separate groups, as well as the remaining individuals in two indistinguishable clusters.

The Major Upper Transect, subset two, showed a split of two clusters, one including samples from Caatinga, Cerrado and Atlantic Forest northeast, and the other including individuals from the Amazon biome. Geneland showed a more pronounced separation of clusters, while Structure also identified both groups, but detecting more migrants (*e.g.*, migrants in Cerrado and Atlantic Forest).

In Geneland, subset three (Lower Transect) clustered the majority of individuals together (Pantanal, Cerrado and Atlantic Forest), and differentiated only three individuals belonging to the Amazon biome (these correspond to population MT2 in *a priori* analysis). In Structure this transect was not differentiated, despite resulting in K=2.

East transect showed a gradual shift in the genetic composition from north to south in Geneland results, and a less differentiated pattern in Delta K's Structure result.

Summarizing the results of Geneland and Structure, it is possible to infer a genetic differentiation of Amazon Forest individuals, evidenced by three subsets (All Individuals, Major Upper and Lower Transects).

Allelic richness comparisons between biomes revealed a slightly higher diversity in the Amazon populations (AM), followed by the central biomes Cerrado and Pantanal (CE/PAN), with lower allelic richness in the Atlantic Forest biome (AF) (Figure 4).



Figure 4 – Plots of predictors of diversity: mean allelic richness (black line) and expected heterozygosity (grey line) for microsatellite loci, in biomes Amazon Forest (AM), Cerrado and Pantanal biomes (CE/PT) and Atlantic Forest (AF).

Although there is a trend of decreasing allelic richness and expected heterozygosity (H_E) from the Amazon to the other biomes, Student's two sample t-test of difference between means of allelic richness and H_E for Amazon and Atlantic Forest clusters were not significant (between AM and AF+CE/PAN: allelic richness, t=0.1952, p-value=0.424, H_E , t=0.28, p-value=0.3918; between AM and AF: allelic richness, t=0.2783, p-value=0.3926, H_E , t=0.2872, p-value=0.3891).

Isolation-by-distance and long-term historical divergence

A positive and significant correlation between genetic and geographical distance (Mantel's test) was observed in all subsets (Figure 5).



Figure 5 – Mantel tests of genetic and geographical (Log 10) distance matrices for each subset tested in a posteriori analysis. A positive significant correlation is observed in all subsets, with a higher value in Major Upper Transect.

When testing for correlation inside each transect separately, Major Upper Transect presented the highest correlation slope with significant values (0.2557, p<0.0001), while Lower Transect presented the lowest (0.0696, p=0.0007). Correlation inside Cluster 2 composed by differentiated Amazonian individuals in "All Individuals" was not significant, although still positive. The same pattern occurred for Major Upper Transect Cluster 2. Both clusters from East Transect kept positive and significant correlation values.

Partial Mantel tests controlling for geographical distance revealed a significant relationship between genetic distances and the presence of barriers inside Major Upper Transect clusters (0.1251, p=0.0008), while no evidence for the existence of barriers was observed for East Transect clusters (Table 3).

Table 3 – Mantel tests performed, scale, maximum continental distance, correlation value (r), and p-value. All tests were performed between raw genetic distances against geographical distances between each pair of individuals measured in log-kilometers. Partial Mantel tests included a third covariate represented by individuals assigned to clusters by Geneland, and the analysis was run controlling for the geographical distances. *Gen*=genetic distance, *Bar*=tested barrier between clusters, *Dist*=geographical distance. Asterisks (*) indicate statistically significant p-values. Lowe Transect Cluster 2 was not tested due to small sample number (N=3).

| Test | Scale | Distance | r | P-value |
|------------------|---|---------------|---------|----------------|
| | | (km) | | |
| Mantel | All Individuals | - | 0.1411 | 0.0001* |
| Mantel | All Individuals Cluster 1 (dark grey) | - | 0.2521 | 0.0001* |
| Mantel | All Individuals Cluster 2 (light grey) | - | 0.0542 | 0.6702 |
| Mantel | Major Upper Transect | 4861 | 0.2557 | 0.0001* |
| Mantel | Major Upper Transect Cluster 1 (dark grey) | - | 0.1650 | 0.0001* |
| Mantel | Major Upper Transect Cluster 2 (light grey) | - | 0.0621 | 0.2803 |
| Mantel | Lower Transect | 2291 | 0.0696 | 0.0007* |
| Mantel | Lower Transect Cluster 1 (dark grey) | - | 0.0291 | 0.0001* |
| Mantel | Lower Transect Cluster 2 (light grey)** | - | - | - |
| Mantel | East Transect | 3106 | 0.1489 | 0.0001* |
| Mantel | East Transect Cluster 1 (dark grey) | - | 0.1189 | 0.0001* |
| Mantel | East Transect Cluster 2 (light grey) | - | 0.0783 | 0.0007* |
| Partial Mantel – | East Transect | 3106 | -0.0740 | 0.2503 |
| Gen~Bar/Dist | (Cluster 1x Cluster 2) | | | |
| Partial Mantel – | Major Upper Transect | 4861 | 0.1141 | 0.0008* |
| Gen~Bar/Dist | (Cluster 1x Cluster 2) | | | |

Discussion

We used a widespread sampling, distributed across most of *Tamandua tetradactyla's* range, crossing all major Brazilian biomes to investigate the global pattern of genetic diversity and structure in the species. This is the largest genetic study performed for this vagile mammal in South America to date. We used two distinct approaches to inspect the genetic structure pattern in the species using microsatellite data: an *a priori* approach, in which individuals were clustered based on geographical proximity, and *a posteriori* approach, performed with individual-based, spatially explicit, model-based softwares. Both approaches led to similar results in general patterns of genetic structuring.

In *a priori* population assembly, deviations from HWE in some populations at different loci may be the result of low sampling per locality/population. We believe that genotyping errors (null alleles and allelic dropout) were less likely the cause of these deviations, since we observed the same pattern with and without historical samples, which have degraded and low DNA quantity, and tend to be more prone to these errors than fresh samples (Pompanon *et al.* 2005). Using the same reasoning, we excluded the possibility that temporal structuring could affect our results. It is also possible that some premises of HWE were violated, since we grouped individuals by geographic proximity in this approach. It is possible that within a given so-called "population", individuals were not actually panmitic, and migration could also affect allele frequencies in that "population" (Hartl & Clark 2006). Therefore, this outcome suggests that these "*a priori* populations" may not all reflect subtle local structuring in natural populations of the lesser anteater, although it does show a general pattern of structure. Combining results from *a priori* and *a posteriori* clustering methods, moderate levels of genetic diversity were found, depicted from the number of alleles per locus and per cluster. However, low genetic structure was found, represented mainly by a genetic split between Amazon Forest individuals and the rest of the sampling distribution. In fact, a study with a smaller sampling (61 individuals) describing the Citochrome Oxidase subunit I (COI) mitochondrial DNA diversity of the lesser anteater found no geographical structuring of the genetic diversity, despite high number of haplotypes (27 haplotypes, Moraes-Barros *et al.*, data not published).

Although it is expected that the species suffered high differentiation during its range expansion, mammals capable of crossing long distances typically exhibit low levels of differentiation (Wilson & Strobeck 1999; Polziehn *et al.* 2000), and animals inhabiting large open areas often display high genetic diversity with small structuring, as shown in studies with large mammals from Africa and Australia (Simonsen *et al.* 1998; Lorenzen *et al.* 2008; Neaves *et al.* 2009). This was the case for lesser anteater populations in central Brazil – Cerrado, Caatinga and Pantanal biomes – that showed to be less differentiated than the others.

When considering all samples (subset "All Individuals"), only one main split was found, separating some individuals of the Amazon biome from all the others. Further testing different configurations of sample assembly led to detection of a subtle north to south genetic differentiation (as illustrated in Figure3). This hierarchical testing of genetic structure, where successive subsets of samples are tested according to the outcomes of previous Bayesian clustering runs, has proven to be useful for widespread species with less severe genetic boundaries (Balkenhol *et al.* 2014). Furthermore, the use of both Bayesian

87

clustering softwares was important to differentiate between clear genetic structuring (such as Amazonian individuals) and patterns of isolation by distance (such as Atlantic Forest populations). Geneland takes explicitly into account the geographical information, while Structure does not. Geneland reported sharped genetic discontinuities, and almost no migrant individuals, as it was partially driven by the large geographical distances across samples coupled with the major genetic break. Structure, on the other hand, reported a variety of spread migrants, facilitating the observation of gradual genetic changes across sampled individuals. Therefore, Structure worked as a control for the role of the geographical information when assigning individuals to clusters by genetic markers.

Nevertheless, comparing results from *a priori* and *a posteriori* approaches, *a posteriori* resulted in a much more comprehensible outcome regarding the definition of genetic differentiation. In this approach no small populations were considered, individuals were inspected directly as mixtures of clusters or assembled in genetic cohesive groups, and a major, clear pattern could emerge.

Genetic structure or isolation by distance in the Atlantic Forest/East Subset

The extant distribution of genetic diversity of the lesser anteater in the Atlantic Forest most likely indicates a stepping-stone model of population structure. The highest significant value of population pairwise F_{ST} was found between populations BA1 and ES1 for *a priori* analysis. In fact, high F_{ST} values were found between most populations from the northern part against the southern part (Figure 2). Many studies regarding species distributed along the Brazilian Atlantic Forest have reported a division between northern and southern populations, most of them sharing the location in which this discontinuity occurs, the Doce river, which have been considered since then a biotic convergence zone (Pellegrino et al. 2005; Grazziotin et al. 2006; Cabanne et al. 2008; Lara-Ruiz et al. 2008; Thomé et al. 2010; Silva et al. 2012). Even though the a priori analyses showed a genetic differentiation between closely located populations, in localities that are coincidently above and below Doce River, this outcome could be due to limited and gaped sampling in the area, given the contrasting evidences from *a posteriori* analyses. In *a posteriori* analyses, especially in Structure outcome, the scenario seemed to be better explained by isolation by distance (e.g., the method by Pritchard et al 2010, L(K), indicated K=3 for East subset, contrasting with Evanno's, Delta K, K=2, and this further subdivision followed the gradient of the samples' distribution in the area, corroborating an isolation by distance pattern). Furthermore, Partial Mantel tests did not corroborate a barrier to gene flow in this subset when correcting for geographical distance. Finally, it has been proposed that the central portion of the Atlantic Forest supported a stable climatic area of evergreen forest during the Last Glacial Maximum (LGM, 21000 years BP) in the Quaternary, functioning as a refuge that maintained constant populations of many taxa (Carnaval et al. 2008; 2009, reviewed and tested in Martins 2011), while other parts of the forest, such as the south-southeast were transformed into semi-arid formations. Yet, given the species high adaptability to different environments, it is possible that populations never ceased to occur across the entire biome, even during the LGM. In fact, populations and/or individuals from Central Brazil did not show any sign of genetic divergence from Atlantic Forest populations in a posteriori analyses of Lower and East transects, corroborating the hypothesis of temporal continuity of the species within coastal Brazil. Therefore, despite the clear population genetic differentiation along the East subset, especially regarding samples from Atlantic Forest biome, this differentiation might be solely the result of isolation by distance, and not due to

barriers to gene flow. This result further confirm the better performance of individual-based rather than population-based approach for the species.

Non-distinctiveness of Central Brazil populations

As aforementioned, no genetic divergence was found between Atlantic Forest and Central Brazil (biomes Pantanal and Cerrado) populations in a posteriori analyses of Lower and East transects. Apparently, open land areas such as these biomes do not represent barriers to gene flow for the species, especially because lesser anteaters are not forest dependent, and are adapted to a variety of different landscapes. Furthermore, one important factor that limits the species occurrence and density is the food supply, ants and termites, in the environment (Montgomery 1985; Redford 1985). Termites and ants are abundant in Cerrado, and several nests are accessible at one time, allowing anteaters to maximize food consumption (Rodrigues et al. 2001). Studies with other non-forest dependent mammals also did not find genetic distinctiveness in this region (e.g. the crab-eating fox, Tchaicka et al. 2007), reflecting these biomes may have constituted a net of connections between forested biomes in the past, given its mosaic of vegetation units fluctuating between open grasslands and dry forests, constituting complementary habitats and food resources for nonforest dependent taxa (Oliveira-Filho & Ratter 1995). This could probably explain the miscellaneous nature of Central Brazil populations with Atlantic Forest and a part of Amazonian individuals.

The Amazonian genetic complexity

Amazonian individuals were clearly differentiated from the rest. F_{ST} values (Figure 2, Table S4) in *a priori* analyses showed a differentiation of Amazonian populations (AM1, PER1,

PICE, MT2) from the other populations. *A posteriori* analysis confirmed this divergence (All Individuals, also corroborated by Major Upper and Lower Transect, Figure 3).

A marked evidence of isolation by distance was observed inside Major Upper Transect (Figure 5 and Table 3), which was also the longest geographical distance tested. However, Cluster 2 inside Major Upper transect composed mainly by Amazonian individuals did not show a significant correlation between genetic and geographical distance, and partial Mantel tests corroborate the existence of a barrier between clusters of Major Upper Transect when correcting for distance (Table 3). This pattern can only be partially explained by major river barriers. Amazon river, for instance, is known to be an effective barrier to gene flow for a number of taxa (e.g. Floyd et al. 2004 for bird species and Peres et al. 1996 for primates), but we sampled few individuals of lesser anteater from the same cluster in both river sides to properly discuss this hypothesis. Yet, the Amazon basin is a very complex system, and it has been claimed to explain morphological variation, divergence and speciation of many taxa, since Wallace's (1852) first hypothesis of riverine barriers shaping primates distribution patterns. In fact, Amazonian rivers are dynamic systems with complex histories of lateral channel migration and shifting river courses that could cause taxa segregation, as well as merely being secondary contact places for taxa that diverged elsewhere (Silva & Patton 1998). Lesser anteaters seem to have been influenced by river barriers, although it does not fully explain its genetic structure pattern.

The genetic composition inside Amazon clusters is complex, as revealed by *a posteriori* analysis, which indicated at least two populations composing the genetic ancestry of the region (Figure 3). Values of allelic richness and H_E also points towards a more diverse

91

genetic composition of the lesser anteater inside the biome as opposed to the rest of the sampling area (Figure 5).

Amazon forest holds one of the highest species richness in the world (Voss & Emmons 1986; Costa & Leite 2000). Due to its intricate development, many factors are held responsible for the complexity and diverse composition of the biota. The Andean uplift and its effect on regional climate is thought to have changed the Amazonian landscape drastically by reconfiguring drainage patterns and creating a vast influx of fragments into the basin. The successive fragmentation and wetlands formation diversified many taxa, amongst them Xenarthrans (Hoorn *et al.* 2010). The Amazon biome also presents the highest variation in morphology. Wetzel (1975) described the greatest variation in color of pelage and a clinal decrease in size of specimens along a boundary marked by the Amazon river and its western tributaries, where non-vested yellow and melanistic (black, brown) specimens occur, as well as partially and fully vested lesser anteaters. This is opposed to the pattern found in the rest of the distribution, where a predominance of vested individuals takes place.

Furthermore, Ohana (2011) developed another morphological study with lesser anteaters comprising a wider sampling and using geometric morphometrics to analyze pelage color, skull and jaw characters. This work described a significant difference of pelage coloration characters between specimens from the north of the Amazonas river and surrounding regions (including neighboring countries), and specimens from south of Amazonas river and eastern Madeira, and proposed a new taxonomic rearrangement spliting the first group into another species, *Tamandua nigra*, and the second group into the already known *Tamandua tetradactyla*. The mixed pelage patterns inside the Amazon biome described by

92

these studies gives further insight to the complex genetic composition of the species in this region, depicted by at least two different clusters.

Although all these evidences point towards a differentiation of Amazonian individuals, caution must be taken to refute the hypothesis of isolation by distance. The fact that some individuals in Cluster 2 of Major Upper Transect were not clustered together when tested in "All Individuals", and that Structure showed much more mixture than Geneland in the "All Individuals" transect (despite results were roughly similar, Figure 3) shows a significant role of geographic information shaping the results, and so the hypothesis of isolation by distance may not be completely ruled out.

Despite our sampling in the Amazon is not sufficient to fully disentangle barriers to gene flow, and a more detailed evolutionary history of the lesser anteater within Amazon forest, morphological evidences from Wetzel (1975) and especially from Ohana (2011), coupled with our description of genetic composition and structure, points to a scenario of historical diversification of the species inside the Amazon biome, and later dispersal to the other environments.

Conclusions

This study is the first investigation of the continental wide genetic structure in this vagile mammal of South America, the lesser anteater *Tamandua tetradactyla*. We were able to describe the patterns of gene flow, and partially reconstruct the biogeographic history of the species throughout an extensive sampling area. A marked differentiation of Amazonian individuals from the rest of individuals emerged. Amazon individuals had higher diversity than the rest sampled areas, and a complex genetic composition, depicted by two genetic

clusters: one with exclusive Amazonian origin, and another that is similar with the other biomes. Although it is not possible to fully discard isolation by distance, there seems to be a barrier that drives genetic differentiation from Amazonian individuals from the rest.

Atlantic Forest samples displayed a likely isolation by distance pattern, reflected in a northern-southern division, and non-differentiation from Cerrado and Pantanal individuals.

These results add knowledge to the species biogeography, and points out to the need of further studies for this species, with an extensive local sampling (especially in Amazon Forest) in order to understand the complex dynamics that led to the extant diversity in this region.

Acknowledgements

This work was done under the SISBIO/IBAMA authorization for scientific activities number 24001-5. We would like to thank José Abílio Ohana and Flavia Miranda for sample donations and valuable discussions. We'd also like to thank Dr. Fabrício R. Santos and Dr. Eduardo Eizirik for sample donations. C.L.Clozato was supported by CAPES scholarship, and J. S. Morgante was granted with a research award from FAPESP (08/52207-0).

References

- Arita HT, Robinson JG, Redford KH (1990) Rarity in Neotropical forest mammals and its ecological correlates. *Conservation Biology*, **4**, 181–192.
- Balkenhol N, Holbrook JD, Onorato D, Zager P, White C, Waits LP (2014) A multimethod approach for analyzing hierarchical genetic structures: a case study with cougars *Puma concolor*. *Ecography*, **37**, 001-012.
- Beheregaray L B (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, **17**, 3754–3774

- Brieger FG (1969) Contribuição à fitogeografia do Brasil com referência especial as orquídeas. *Anais do XX Congresso Nacional de Botânica*, 41–44.
- Cabanne GS, D'Horta F, Sari E, Santos F, Miyaki CY (2008) Nuclear and mitochondrial phylogeography of the Brazilian Atlantic forest endemic *Xiphorhynchus fuscus* (Aves: Dendrocolaptidae): biogeography and systematics implications. *Molecular Phylogenetics and Evolution*, **49**, 760–773.
- Carnaval AC, Hickerson MJ, Haddad, Célio F B, Miguel T. Rodrigues CM (2009) Stability predicts genetic diversity in the Brazilian Atlantic Forest Hotspot. *Science*, **323**, 785–789.
- Carnaval AC, Moritz C (2008) Historical climate modeling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography*, **35**, 1187–1201.
- Ceballos G, and Ehrlich PR (2009) Discoveries of new mammal species and their implications for conservation and ecosystem services. *Proceedings of the National Academy of Sciences*, **106**, 3841-3846.
- Clozato CL, Moraes-Barros N, Santos FR, Morgante JS (2013) Historical and non-invasive samples: a study case of genotyping errors in newly isolated microsatellites for the lesser anteater (*Tamandua tetradactyla* L., Pilosa). *Molecular ecology resources*, in press.
- Costa LP (2003) The historical bridge between the Amazon and the Atlantic Forest of Brazil: a study of molecular phylogeography with small mammals. *Journal of Biogeography*, **30**, 71–86.
- Costa LP, Leite YLR, da Fonseca G a. B, da Fonseca MT (2000) Biogeography of South American Forest Mammals: Endemism and Diversity in the Atlantic Forest1. *Biotropica*, **32**, 872–881.
- Cushman S, Shirk AJ, Landguth EL (2012) Landscape genetics and limiting factors. *Conservation Genetics*, 14, 263–274.
- Cushman S, Wasserman T, Landguth E, Shirk A (2013) Re-Evaluating Causal Modeling with Mantel Tests in Landscape Genetics. *Diversity*, **5**, 51–72.
- Eisenberg JF (1989) Mammals of the neotropics vol I: Panama, Colombia, Venezuela, Guyana, Suriname and French Guiana. University of Chicago Press, Chicago.
- 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, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
- Floyd EH, Sewhal JN (2004) The Amazon River as a dispersal barrier to passerine birds: effects of river width, habitat and taxonomy. *Journal of Biogeography*, **31**(11), 1809-1818.
- Gardner AL (2008) Magnaorder Xenarthra. In: Mammals of South America, Vol 1: Marsupials, Xenarthrans, Shrews, and Bats (ed. Gardner AL), pp 127–176, The University of Chicago Press, Chicago 2008.
- Goslee SC, Urban DL (2007) The *ecodist* package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**(7), 1-19.
- Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485:486.
- Grazziotin F, Monzel M, Echeverrigaray S, Bonatto S (2006) Phylogeography of the Bothrops jararaca complex (Serpentes: Viperidae): past fragmentation and island colonization in the Brazilian Atlantic forest. *Molecular Ecology*, **15**, 3969–3982.
- Guillot G, Mortier F, Estoup A (2005) Geneland: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.
- Haffer J (1969) Speciation in Amazonian forest birds. Science, 165, 131–137.
- Hartl DL, Clark AG (2006) Testing Hardy-Weinberg Equilibrium. In: *Principles of Population Genetics* (Hartl DL, Clark, AG), pp. 70–82. Sinauer Associates, Massachusetts.
- Hayssen V (2011) Tamandua tetradactyla (Pilosa: Myrmecophagidae). Mammalian Species, 43, 64–74.
- Hoorn C, Wesselingh FP, ter Steege H *et al.* (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, **330**, 927–31.
- Jenkins CN, Pimm SL, and Joppa LN (2013) Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences*, **110**, E2602-E2610.
- Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, **49**, 561–576.
- Kuchta S, Tan A (2005) Isolation by distance and post-glacial range expansion in the rough-skinned newt, *Taricha granulosa*. *Molecular Ecology*, **14**, 225–244.

- Lara-Ruiz P, Chiarello AG, Santos FR (2008) Extreme population divergence and conservation implications for the rare endangered Brazilian Atlantic forest sloth, Bradypus torquatus (Pilosa: Bradypodidae). *Biological Conservation*, **141**, 1332– 1342.
- Lorenzen ED, Arctander P, Siegismund HR (2008) High variation and very low differentiation in wide ranging plains zebra (*Equus quagga*): insights from mtDNA and microsatellites. *Molecular ecology*, **17**, 2812–2824.
- Malécot G (1948) Les Mathematiques de L'Herédité. Masson, Paris.
- Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in ecology & evolution*, **28**, 614–21.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.
- Mares M, Barquez RM, Braun JK, Ojeda RA (1996) Observations on the mammals of Tucuman Province, Argentina. Systematics, distribution, and ecology of the Didelphimorphia, Xenarthra, Chiroptera, Primates, Carnivora, Perissodactyla, Artiodactyla, and Lagomorpha. Annals of Carnegie Museum, 65, 89–152.
- Martins FM, Templeton AR, Pavan ACO, Kohlbach BC, Morgante JS (2009) Phylogeography of the common vampire bat (*Desmodus rotundus*): marked population structure, Neotropical Pleistocene vicariance and incongruence between nuclear and mtDNA markers. *BMC evolutionary biology*, **9**, 294–307.
- Meirmans PG (2012) The trouble with isolation by distance. *Molecular ecology*, **21**, 2839–2846.
- Mello Martins F (2011) Historical biogeography of the Brazilian Atlantic forest and the Carnaval-Moritz model of Pleistocene refugia: what do phylogeographical studies tell us? *Biological Journal of the Linnean Society*, **104**, 499–509.
- Miquel C, Bellemain E, Poillot C *et al.* (2006) Quality indexes to assess the reliability of genotypes in studies using noninvasive sampling and multiple-tube approach. *Molecular Ecology Notes*, **6**, 985–988.
- Montgomery GG (1985) *The Evolution and Ecology of Armadillos, Sloths and Vermilinguas.* Smithsonian Institution Press, Washington DC.

- Montgomery GG, Lubin YD (1977) Prey influences on movements of neotropical anteaters. In: *Proceedings of the 1975 predator symposium*. (ed. R.L. Phillips), pp. 103–131, University of Montana Printing Department, Missoula.
- Moraes-Barros N, Silva J, Miyaki CY, Morgante JS (2006) Comparative phylogeography of the Brazilian Atlantic forest endemic sloth (*Bradypus torquatus*) and the widespread three-toed sloth (*Bradypus variegatus*) (Bradypodidae, Xenarthra). *Genetica*, 126: 189–198.
- Murphy WJ, Eizirik E, Johnson WE *et al.* (2001) Molecular phylogenetics and the origins of placental mammals. *Nature*, **409**, 614–618.
- Neaves LE, Zenger KR, Prince RI, Eldridge MD, Cooper DW (2009) Landscape discontinuities influence gene flow and genetic structure in a large, vagile Australian mammal, *Macropus fuliginosus*. *Molecular Ecology*, **18**(16), 3363-3378.
- Novak RM (1983) *Walkers Mammals of the world*. The Johns Hopkins University Press, Baltimore and London.
- Ohana JAB (2011) Variação morfológica do tamanduá-mirim, Tamandua tetradactyla (*Linnaeus, 1758*)(*Pilosa, Vermilingua*). Master's Thesis. Universidade Federal do Pará/Museu Paraense Emílio Goeldi.136 pp.
- Oliveira-Filho AT, Ratter JA (2002) Vegetation physiognomies and woddy flora of the Cerrado Biome. In: *The Cerrados of Brazil*. (eds. Oliveira PS, Marquis RJ), pp 91–120, Columbia University Press, NewYork.
- Peres CA, Patton JL, Nazareth F, Silva M (1996) Riverine Barriers and Gene Flow in Amazonian Saddle-Back Tamarins. *Folia Primatologica*, **67**, 113-124.
- Parera A (2002) Los mamíferos de la Argentina y la región austral de Sudamérica. Editorial El Ateneo, Buenos Aires.
- Pellegrino K, Rodrigues M, Waite A, Morando M, Yassuda Y, Sites J Jr. (2005) Phylogeography and species limits in the *Gymnodactylus darwinii* complex (Gekkonidae, Squamata): genetic structure coincides with river systems in the Brazilian Atlantic forest. *Biological Journal of the Linnean Society*, **85**, 13–26.
- Pilot M, Jedrzejewski W, Branicki W *et al.* (2006) Ecological factors influence population genetic structure of European grey wolves. *Molecular ecology*, **15**, 4533–4553.
- Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–859.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

- R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <u>http://www.R-project.org</u>.
- Redford KH (1985) Food habits of armadillos (Xenarthra: Dasypodidae). In: *The evolution and ecology of armadillos, sloths and vermilinguas* (ed. Montgomery GG), pp 429–437, Smithsonian Institution Press, Washington, DC.
- Redford KH, Eisenberg JF (1992) Mammals of the Neotropics. Vol. 2: The Southern Cone: Chile, Argentina, Uruguay, Paraguay. The University of Chicago Press, Chicago and London.
- Redford KH, Fonseca GAB (1986) The role of gallery forests in the zoogeography of the cerrado's non-volant mammalian fauna. *Biotropica*, **18**(2), 126-135.
- Rice WR (1989) Analyzing table of statistical tests. Evolution, 43, 223–225.
- Rodrigues FHG, Marinho-filho J, Santos HG (2001) Home ranges of translocated lesser anteaters *Tamandua tetradactyla* in the Cerrado of Brazil. *Oryx*, **35**, 166–169.
- Rodrigues FGH, Medri IM, De Miranda GHB.Camilo-Alves C, Mourão G (2008) Anteater behavior and ecology. In: *The biology of the Xenarthra* (eds. Vizcaíno SF, Loughry, WJ), pp. 257–268, University Press of Florida, Gainesville.
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
- Silva MN, Patton JL (1998) Molecular phylogeography and the evolution and conservation of Amazonian mammals. *Molecular ecology*, **7**, 475–486.
- Silva MS, Moraes-Barros N, Ribas C, Ferrrand N, Morgante JS (2012) Divide to conquer : a complex pattern of biodiversity depicted by vertebrate components in the Brazilian Atlantic Forest. *Biological Journal of the Linnean Society*, **107**, 39–55.
- Simonsen BT, Siegismund HR, Arctander P (1998) Population structure of African buffalo inferred from mtDNA sequences and microsatellite loci: high variation but low differentiation. *Molecular ecology*, **7**, 225–237.
- Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, 35, 627–632.
- Tchaicka L, Eizirik E, De Oliveira TG, Cândido JF, Freitas TRO (2007) Phylogeography and population history of the crab-eating fox (*Cerdocyon thous*). *Molecular ecology*, **16**, 819–838.

- Thomé MT, Zamudio K, Giovanelli J, Haddad C, Baldissera F Jr, Alexandrino J (2010) Phylogeography of endemic toads and post-Pliocene persistence of the Brazilian Atlantic forest. *Molecular Phylogenetics and Evolution*, **55**, 1018–1031.
- Turchetto-Zolet A C, Pinheiro F, Salgueiro F, Palma-Silva C (2013) Phylogeographical patterns shed light on evolutionary process in South America. *Molecular ecology*, **22**, 1193–213.
- Voss RS, Emmons LH (1996) Mammalian diversity in Neotropical lowland rainforests: a preliminary assessment. *Bulletin of the American Museum of Natural History*, **230**, 1–115.
- Vuilleumier F, Monasterio M (1986) *High altitude tropical biogeography*. Oxford University Press, New York.

Wallace AR (1852) On the monkeys of the Amazon. *Proceedings of the Zoological Society of London*, **20**, 107-110.

- Webb SD (2006) The Great American Biotic Interchange: patterns and processes. *Annals of the Missouri Botanical Garden*, **93**, 245-257.
- Wetzel RM (1975) The species of Tamandua Gray (Edentata, Myrmecophagidae). Proceedings of the Biological Society of Washington, **88**, 95–112.
- Wetzel RM (1985) The identification and distribution of recent Xenarthra (=Edentata). In: *The evolution and ecology of armadillos, sloths and vermilinguas* (ed. Montgomery GG), pp 5–21. Smithsonian Institution Press, Washington, DC.
- Wilson G a, Strobeck C (1999) Genetic variation within and relatedness among wood and plains bison populations. *Genome*, **42**, 483–496.
- Worley K, Strobeck C, Arthur S *et al.* (2004) Population genetic structure of North American thinhorn sheep (Ovis dalli). *Molecular ecology*, **13**, 2545–2556.

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

Capítulo 3

Diversidade de sequências e sinais de seleção no gene DRB de MHC Classe II do tamanduá mirim (*Tamandua tetradactyla*) ao longo de diferentes biomas.

Sequence diversity and signs of selection in MHC Class II DRB gene of the lesser anteater (*Tamandua tetradactyla*) across different biomes

Apresentação Capítulo 3

Este capítulo apresenta pela primeira vez na literatura dados do Complexo Principal de Histocompatibilidade (*Major Histocompatibility Complex*, MHC) para uma espécie do grupo Xenarthra. Esta região genômica é conhecida por apresentar sinais clássicos de seleção (positiva e/ou balanceadora) e pode estar correlacionada com a adaptação de indivíduos a pressão de patógenos no ambiente, entre outros fatores.

Utilizando sequenciamento de nova geração, uma metodologia moderna capaz de capturar maior número de variantes genéticas, foi investigada a distribuição da diversidade do éxon 2 do gene *DRB* em indivíduos provenientes dos grandes biomas da América do Sul, e o panorama encontrado foi comparado com os resultados de microssatélites descritos no capítulo anterior. Assim, foi possível comparar os padrões gerais de diversidade e estruturação genética de populações de tamanduá mirim revelados por regiões genômicas neutras e sob-seleção.

Sequence diversity and signs of selection in MHC Class II DRB gene of the lesser anteater (*Tamandua tetradactyla*) across different biomes

CAMILA L. CLOZATO^{1,5} NADIA MORAES-BARROS^{1,2}, JOÃO S. MORGANTE¹, CAMILA J. MAZZONI⁴ and SIMONE SOMMER^{3,5}

Keywords Major histocompatibility complex, DRB, next generation sequencing, selection, drift, *Tamandua tetradactyla*, Brazil.

(Artigo em preparação)

Abstract

The genes of the major histocompatibility complex (MHC) code for proteins involved in antigen recognition and activation of the adaptive immune response, and are thought to be regulated by natural selection, especially due to pathogen-driven selective pressure. In this study we aimed to investigate the distribution of DRB exon 2 gene diversity using Next Generation Sequencing across five Brazilian biomes and compare the MHC pattern with that of neutral markers (microsatellites). We found a noticeable high level of diversity in DRB (60 amino acid alleles in 65 individuals) and clear signatures of historical positive selection acting on this gene. Higher allelic richness and proportion of private alleles was found in rainforest biomes, especially Amazon Forest, a megadiverse biome, possibly harboring greater pathogen richness as well. Nevertheless, neutral markers showed a similar pattern to DRB, demonstrating the strength of demography in shaping MHC diversity and structure.

Introduction

The major histocompatibility complex (MHC) is one of the most important immunogenic systems for infectious disease resistance in vertebrates (Hedrick & Kim 2000). Since these genes are highly variable and are thought to play an essential role in the adaptive immune response of vertebrates, they can be useful for investigating the role of natural selection on genetic diversity in wild populations (Bernatchez & Landry 2003).

It has been proposed that pathogen-mediated selection (PMS) is one of the main driving forces maintaining diversity at MHC loci (Doherty & Zinkernagel 1975; Apanius et al. 1997; Jeffery & Bangham 2000; Bernatchez & Landry 2003). Some hypotheses to explain PMS have been suggested: heterozygote advantage (Doherty & Zinkernagel 1975), rareallele advantage (Slade & McCallum 1992) and fluctuating selection (Hill 1991). All three mechanisms, or a combination of the three, could be the driver of MHC diversity (Hughes & Nei 1988; Takahata & Nei 1990; Apanius et al. 1997). Since MHC is known to respond to PMS, the factors that drive pathogen diversity in different environments could also represent important causal predictors. As pathogens depend on hosts to survive, host diversity is one consequential factor driving pathogen richness (Dunn et al. 2010). Notwithstanding, annual precipitation and, in a smaller proportion, temperature, are important factors explaining diversity of many groups of pathogens, such as bacteria, viruses, fungi, protozoa and helminths (Guernier et al. 2004; Froeschke et al. 2010). Therefore, the composition (different kinds) and diversity (species richness) of pathogens are likely to be heterogeneous along distinct environments and higher in wetter areas harboring high host species richness.

105

South America is a large territory that includes a vast array of climates, closely associated with vegetation formation. It comprises a tropical region near the equatorial zone, as well as a subtropical region with temperate climates (Fittkau *et al.* 1969; Sylvestre 2009). The continent shows complex geomorphological patterns (large river plains, *e.g.* the Amazon basin, and extensive mountain chains, *e.g.*, the Andean Cordillera) (Clapperton 1993). South America harbors the greatest biodiversity on Earth, containing five of the world's biodiversity 'hotspots' (Myers *et al.* 2000). Moreover, South America shows a complex biogeography, composed of several biomes/ecoregions differentiated mostly by vegetation and climate conditions (Morrone 2004; 2006).

The Brazilian flora composition can be divided into six major biomes. The tropical rainforests (Amazon Forest and Atlantic Forest), the Brazilian Savanna (Cerrado), the wetlands (Pantanal), the semi-arid Northeast vegetation (Caatinga), and the southern grasslands (Campos Sulinos) (adapted from Veloso *et al.* 1991 and reviewed in Joly *et al.* 1999). Different levels of species richness and endemism are observed between these formations. The Amazon Forest is a megadiverse biome (Hoorn *et al.* 2010) and as well as the Atlantic Forest considered as a hotspot of biodiversity with high levels of endemism (Myers *et al.* 2010). Thus, they are expected to harbor more pathogens than drier or less diverse biomes. This vast continent favors a differential distribution and composition of pathogens along its range.

The lesser anteater, *Tamandua tetradactyla* (Linnaeus 1758), is a medium-sized mammal of the family Myrmecophagidae, order Pilosa. It is part of one of the most ancient lineages of placental mammals, the Xenarthra magnaorder (Murphy *et al.* 2001), a group that evolved and diversified in South America (Webb 2006). The species has a wide geographic

distribution: it occurs in South America east of the Andes, from Venezuela and Trinidad until the north of Argentina, and south of Brazil and Uruguay, in elevations up to 2000 meters (Novak *et al.* 1983; Wetzel 1985; Gardner 2008). Although it has a preference for forested areas, it is largely found also in open grassland savanna-like areas such as Cerrado, wetlands such as Pantanal, in mountain tropical regions (Eisenberg 1989) and transitional forests (Mares *et al.* 1996). In fact, the species inhabits all major biomes in South America. The lesser anteater is, thus, a suitable model to study the MHC diversity along different habitats (with different predicted pathogen diversity) in South America.

In this study we characterized MHC Class II DRB exon 2 diversity in *Tamandua tetradactyla* and examined the distribution of this diversity across Brazilian biomes. Our overall aim was to investigate if there are different composition of alleles specific to certain geographic regions which could be indicative of local adaptation to differential pools of pathogens in the landscape, or if alleles are randomly distributed in space, meaning they are all equally adapted. In the first scenario, we would expect higher levels of genetic diversity and presence of private alleles in rainforest biomes harboring higher pathogen pressure. Additionally, we compared this pattern of MHC spatial diversity with that of neutral markers, such as microsatellites (after Chapter 2), to examine how strongly related to demographic processes are the MHC loci in the lesser anteater.

Materials and Methods

Sampling design

Genetic samples of *Tamandua tetradactyla* were collected across five Brazilian biomes: Atlantic Forest (coastal tropical rainforest, AF, n=29), Amazon Forest (tropical rainforest, AM, n=12), Caatinga (northeast arid grassland vegetation, CA, n=3), Cerrado (central open grassland vegetation with gallery forests, CE, n=16), Pantanal (open grassland seasonal wetlands, PT, n=11) between 2006 and 2012 through fieldwork, recovery from road kills and collaborations with researchers and/or institutions. Additionally, two samples from Peru and one from French Guiana were obtained through donation, and assigned to the Amazonian biome. A total of 71 samples were used for this study. Details on the origin of samples and a map of sampling sites are provided in Table S1 and Fig. 1. The biome assignment is based on the geographic coordinates and biomes' definition from the Brazilian Institute of Geography and Statistics (IBGE, <u>http://www.ibge.gov.br/home/</u>) database.

Plucked hair was kept in a dry recipient, blood and tissue samples were kept in 70% ethanol and kept in 4°C until processed for DNA extraction. Genomic DNA extraction was performed with Proteinase K digestion enhanced with DDT 1M, followed by salt precipitation (Sambrook *et al.* 1989) and stored at -20°C before use. All field work and sample management was performed under the SISBIO/IBAMA authorization for scientific activities number 24001-5/53695225.


Figure 1 – Map of sampling sites of all individuals used in this study across five biomes: Atlantic Forest (AF), Amazon Forest (AM), Caatinga (CA), Cerrado (CE) and Pantanal (PT). Shaded area represents the distribution of *T. tetradactyla*, lines inside Brazil represent biome's limits according to IBGE (<u>http://www.ibge.gov.br/home/</u>), circles show sampling localities. Pie charts indicate the proportion of five most frequent alleles (TateDRB*40, 04, 39b, 01a and 25), private alleles and other rare alleles (see legend) in each biome.

Preparatory work for the next generation sequencing approach: primer design and standardization of amplification efficiency by Single Strand Conformation Polymorphism (SSCP) and Sanger sequencing

The SSCP technique was used (i) to screen for the best suitable primer pair that catches the most diversity at the target locus, MHC Class II DRB exon 2, and minimize the occurrence of artefacts, such as null alleles and (ii) to account for possible differences in the amplification efficiency across loci in the species (Sommer et al. 2013). For this purpose, 10 random samples with high quality DNA were used. Tested primers were binding to conserved sites of DRB intron 1 or exon 2 (forward), and to DRB exon 2 or intron 3 (reverse). Amplification reactions (PCRs) were conducted in a 20 µl final volume, with 100 ng of template DNA, 0.375 µM of each primer, 5x HotStar HiFidelity PCR Buffer (including MgSO₄ and dNTPs mix), 5x Q-Solution and 0.5 U of HotStar HiFidelity DNA Polymerase (Qiagen). Thermocycling program consisted of an initial denaturation of 5' at 94°C, 35 cycles of 1' at 94°C, 1' at 53-55°C and 1' at 72°C, with a final extension of 10' at 72°C. The chosen primer pair with best JF1eV (5'vield was GAGTGTCATTTYGAGAACGGGACSGAG-3') YML10 (5'and TCGCCGCTGCACTGTGAACGTCTC-3') (Sommer et al. 2013) both binding in the exon 2, amplifying 242 bp. Amplicons derived from successful amplifications were genotyped by SSCP on polyacrylamide gels (Sommer & Tichy 1999). SSCP analysis was performed twice per individual on separate gels using independent PCRs to confirm the banding pattern of all detected alleles. Details about the SSCP preparation, reagents and electrophoretic conditions are described in Appendix S1, Supporting Information. Single strand bands were excised from gel matrix and diluted in distilled water for reamplification

in a final volume of 20 μ l, 0.375 μ M of each primer, 1.75 μ M dNTP mix, 2.5 μ l and 0.5 U of Taq polymerase (MP Biomedicals), using the same program as above. PCR products were purified and sequenced in both directions using BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) on an ABI PRISM 310 (Applied Biosystems).

Sequence electropherograms were visually inspected using Chromas Lite 1.02 (Technelysium Pty Ltd, Queensland, Australia). Alignment and sequence translation were performed with ClustalW algorithm implemented in MEGA 6.0 (Tamura *et al.* 2013). To check for the histocompatibility nature of the sequences, homology was verified using GenBank database (https://www.ncbi.nlm.nih.gov/genbank/) with the online tool BlastN (http://www.ncbi.nlm.nih.gov/BLAST). To define a sequence as a putative MHC DRB exon 2 allele the criteria used was its occurrence in at least two independent reactions from the same individuals or detection in at least two distinct individuals. Putative alleles were named according to the nomenclature rules defined in Klein *et al.* (1990) after confirmation by subsequent pyrosequencing.

Next generation sequencing (NGS) approach

The library preparation for pyrosequencing on a 454 GS Junior Titanium platform (Roche) was performed using fusion primers composed of four parts: (i) adaptor lib A sequence (forward: 5'-CGTATCGCCTCCCTCGCGCCA-3' or reverse: 5'-CTATGCGCCTTGCCAGCCCGC-3'), (ii) internal library key (TCAG), (iii) 10 base pair long barcodes called multiplex identifiers (MIDs) to identify each individual and (iv) the sequence of the specific chosen primer pair (forward: JF1eV or reverse: YML10 see above). A compilation of all fusion primers is available in Table S2, Supporting

Information. In total, forward fusion primer was 62 bp long, and reverse was 59 bp. For individual barcoding, eleven forward and ten reverse fusion primers were used. The amount of different MIDs combinations allowed pooling a maximum of 100 tagged samples in each run. In order to safely assign putative alleles, every individual was separately amplified and sequenced twice using different barcodes (amplicon replicates), following Sommer et al. (2013) recommendations to deal with artefacts and allelic dropout. Thus, to genotype all individuals, two independent pyrosequencing runs were necessary. PCR was done in 25 μ L reaction volumes containing 100 ng of template DNA, 0.4 µM of each fusion primer, 0.2 mM dNTPs, 2.5 µL FastStart buffer and 1.25 U FastStart HiFi Polymerase (Roche Diagnostics GmbH, Germany). Thermocycling program consisted of a denaturation step for 2' at 94°C, followed by 35 cycles for 30" at 94°C; 30" at 55°C, 1' at 72°C and a final extension for 7'at 72°C. PCR products were purified through gel band extraction using peqGOLD Gel Extraction Kit (PeqLab, Germany) and quantified by the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen Corporation). Subsequently, all amplicons were diluted to 200,000 molecules/µl and pooled. Emulsion PCR, beads recovery and DNA library enrichment was performed according to the manufacturer's instructions. The enriched library pool was then sequenced on a PicoTiter plate in a 454 GS Junior Titanium (Roche Diagnostics GmbH).

Quality check of NGS reads and putative allele assignment

Genome Sequencer FLX System Software was used for initial image and signal processing using the standard amplicon pipeline option. Quality check and filtering steps were performed following the approach and recommendations of Sommer *et al.* (2013). Briefly, all reads much shorter than expected were excluded (~310 bp, including fusion primers and target DNA fragment). Reads with incomplete or incorrect MID sequences were also removed. Reads with incomplete or incorrect primer regions or less than 95% bases with Phred quality score Q>20 were removed.

Alignments within individual amplicons were performed using the software Geneious Pro v.5.6.5 (Drummond *et al.* 2012). Reads with changes in the reading frame and indels other than 3 bp long (or multiples of 3 bp, corresponding to codons) were excluded due to the biological incompatibility of these features with functional MHC genes. All singletons reads were considered artefacts. All remaining reads were clustered based on identity (called clusters), and these consensus sequences were considered hereafter as variants. The subsequent pipeline used in this study to discriminate "artefacts" from "putative alleles" is illustrated and described in detail in Sommer *et al.* (2013). Variants within each amplicon were organized based on their frequency, which was considered a criterion for grouping likely putative alleles and probable artefacts since putative alleles were assumed to be more frequent than artefacts. At this stage the major artefacts which still need to be sorted out are "chimera" (artificial and low frequency combination of two parental common variants). A third category included "unclassified variants" defined as being present in both individual amplicon replicates and more frequent than putative artefacts but less frequent than putative alleles. After this first step within each amplicon, a second step was performed crosschecking for correspondence of these variants in the amplicon independent replicate. Variants were classified as "putative allele" if they were present in both individual amplicon replicates and more frequent than any artefact. Once "unclassified variants" were considered consistent among amplicons, they were categorized as "putative alleles with low amplification efficiency".

Characterization of MHC class II DRB exon 2 in lesser anteater across biomes

Sequences of putative MHC alleles were edited, aligned and translated using MEGA 6.0 (Tamura et al. 2013). This software was also applied to count the number of variable positions and the mean number of differences between alleles. Nucleotide and amino acid genetic distances were calculated using Kimura-2-parameters model of substitution and Poisson-corrected distance, respectively. For the genetic distance parameter, calculations were performed for the entire dataset (all sequences), as well as for each biome separately. The relative rates of non-synonymous (dN) and synonymous (dS) base pair substitutions were calculated according to Nei & Gojobori (1986), using Jukes-Cantor correction for multiple hits. These calculations were performed for all sites, for putative antigen biding sites (ABS) separately, assuming correspondence to human ABS of HLA-DR1 molecule (after Brown et al. 1993), and for all sites excluding ABS. The dN/dS ratios after 1000 bootstrap replicates were compared with an implemented two-tailed Z-test (Nei & Kumar 2000) to test for positive selection. The number of probable loci as product of gene duplications was estimated as the half of the maximum number of alleles for one individual. The phylogenetic relationships between MHC class II DRB exon 2 alleles were investigated by a Bayesian phylogenetic tree obtained by Mr. Bayes vs. 3.2.0 (Ronquist et al. 2012). A mixed model for mixed rates for amino acid with Gamma distribution was used. Values of posterior probability were obtained with 300,000 generations and 1000 burn-in, sampled every 1000 chain to check MCMC convergence. Each terminal (allele) was labeled by its occurrence in each biome.

Allelic richness was calculated based on allele frequencies. Since there are several loci, and no assumption could be made on their inheritance pattern, the assignment of an allele to a particular locus could not be done. Therefore, alleles were classified based on frequency ranges (five intervals, most to least common), and these were considered as "artificial loci" for each individual. The R package *standArich* (Alberto 2006) was used to calculate allelic richness based on a rarefaction method. Nucleotide diversity was calculated for the entire sequence and for ABS site only with the R package *pegas* (Paradis 2014), considering all MHC alleles and alleles that occur in each biome separately. To envision the sharing of MHC alleles between biomes and individuals, Circus online tool (http://circos.ca/), a method of circular visualization of tables, was used (Krzywinski *et al.* 2009). Similarly, the evolutionary relationship between MHC alleles, biomes and individuals was observed through a network taking the required number of mutations between alleles into account by using a Median-Joining algorithm (Bandelt *et al.* 1999) implemented in the software Network vs. 4.612 (http://www.fluxus-engineering.com/sharenet.htm).

We used the *ecodist* package (Goslee & Urban 2007) in R to test the correlation of a matrix of raw individual genetic and geographical distance (Euclidean distances between each pair of individuals measured in log-kilometers) applying Mantel tests with 1000 replications (Mantel 1967). We also performed Partial Mantel tests to correct for geographic distance considering different biomes as predictor variables for genetic distance.

Microsatellite data analyses

All samples used in this study were also genotyped at eight microsatellite loci (H5, E12, G3, F1R, B2, C10, A9, A8). Isolation and characterization of loci are described in Chapter 1. The sequences of microsatellite fragments are deposited in GenBank (accession numbers KF746177-KF746185), and genotyping data of loci for samples used in this study were

extracted from Chapter 2. Number of alleles, number of private alleles (averaged across loci and biomes) and observed and expected heterozygosity (H_0/H_E) were calculated with GenAlEx package 6.41 (Peakall & Smouse 2006). Allelic richness and Mantel tests were also performed for microsatellites as described for DRB data.

Results

MHC Class II DRB exon 2 diversity and selection pattern in Tamandua tetradactyla

A total of 65 out of 71 samples presented consistent 454 pyrosequencing data for further analyses steps. Six samples did not yield enough coverage in one or both replicate amplicons, and were subsequently excluded. After applying the workflow described in the methodology section to classify variants, final data filtering yielded 6,402 reads classified as "putative artefacts", 1,194 as "unclassified variants" and 91,374 as "putative alleles". For the 65 individuals, the number of reads per amplicon after final data filtering ranged from 110 to 5,344, and the number of reads per individual varied between 232 and 10,426.

We validated 70 MHC Class II DRB exon 2 alleles in the nucleotide level and 60 alleles in the amino acid level (Table 1, Fig. 2). Average coverage per allele was 1,305 reads ranging from 4 to 18,413 (Table 1). Alignment of the nucleotide sequences against Genebank database using BlastN confirmed the MHC nature of all alleles. *T. tetradactyla* MHC Class II DRB allele sequences will be deposited in Genbank for future reference. One allele (TateDRB*48) presented a 3 bp insertion at position 13-15 of the nucleotide alignment, and 13 alleles (TateDRB*06, 17, 20, 29, 43, 44a, 44b, 44c, 44d, 45, 46, 47, 49) presented a 3 bp deletion at position 169-171. Both indels represent one single codon (inframe alteration), so the reading frame was unaltered.

Table 1 – MHC class II DRB exon 2 alleles observed in *Tamandua tetradactyla* across five biomes. *N* is the number of individuals that contain a specific allele. Relative frequency of a specific allele in the overall data set. MHC alleles that are similar at the amino acid level but different at the nucleotide level are indicated with "a", "b", "c" and "d". Coverage refers to overall number of reads for each allele. The total number of validated reads was 91,374.

| DRB Allele | Ν | Frequency | Coverage | DRB Allele | Ν | Frequency | Coverage | |
|-------------|----|-----------|----------|-------------|----|-----------|----------|--|
| TateDRB*01a | 21 | 0,323 | 7289 | TateDRB*34 | 1 | 0,015 | 11 | |
| TateDRB*01b | 4 | 0,061 | 9408 | TateDRB*35 | 1 | 0,015 | 6 | |
| TateDRB*02 | 3 | 0,046 | 132 | TateDRB*36 | 3 | 0,046 | 115 | |
| TateDRB*03 | 14 | 0,215 | 6686 | TateDRB*37a | 2 | 0,030 | 143 | |
| TateDRB*04 | 32 | 0,492 | 18413 | TateDRB*37b | 5 | 0,076 | 1707 | |
| TateDRB*05 | 9 | 0,138 | 1938 | TateDRB*37c | 2 | 0,030 | 43 | |
| TateDRB*06 | 1 | 0,015 | 281 | TateDRB*38 | 11 | 0,169 | 140 | |
| TateDRB*07 | 2 | 0,030 | 58 | TateDRB*39a | 4 | 0,061 | 29 | |
| TateDRB*08 | 2 | 0,030 | 63 | TateDRB*39b | 26 | 0,400 | 10568 | |
| TateDRB*09 | 5 | 0,076 | 3172 | TateDRB*39c | 1 | 0,015 | 240 | |
| TateDRB*10 | 1 | 0,015 | 16 | TateDRB*39d | 4 | 0,061 | 147 | |
| TateDRB*11 | 2 | 0,030 | 768 | TateDRB*40 | 37 | 0,569 | 971 | |
| TateDRB*12 | 1 | 0,015 | 95 | TateDRB*41 | 1 | 0,015 | 144 | |
| TateDRB*13 | 1 | 0,015 | 464 | TateDRB*42 | 4 | 0,061 | 555 | |
| TateDRB*14 | 3 | 0,046 | 894 | TateDRB*43 | 7 | 0,107 | 1208 | |
| TateDRB*15 | 2 | 0,030 | 160 | TateDRB*44a | 11 | 0,169 | 4314 | |
| TateDRB*16 | 2 | 0,030 | 15 | TateDRB*44b | 5 | 0,076 | 636 | |
| TateDRB*17 | 7 | 0,107 | 74 | TateDRB*44c | 15 | 0,230 | 4095 | |
| TateDRB*18a | 1 | 0,015 | 7 | TateDRB*44d | 2 | 0,030 | 1193 | |
| TateDRB*18b | 1 | 0,015 | 44 | TateDRB*45 | 3 | 0,046 | 309 | |
| TateDRB*19 | 1 | 0,015 | 4 | TateDRB*46 | 1 | 0,015 | 273 | |
| TateDRB*20 | 1 | 0,015 | 23 | TateDRB*47 | 2 | 0,030 | 39 | |
| TateDRB*21 | 1 | 0,015 | 29 | TateDRB*48 | 4 | 0,061 | 225 | |
| TateDRB*22 | 1 | 0,015 | 24 | TateDRB*49 | 8 | 0,123 | 74 | |
| TateDRB*23 | 1 | 0,015 | 8 | TateDRB*50 | 1 | 0,015 | 4 | |
| TateDRB*24 | 2 | 0,030 | 522 | TateDRB*51 | 2 | 0,030 | 35 | |
| TateDRB*25 | 19 | 0,292 | 8168 | TateDRB*52 | 1 | 0,015 | 5 | |
| TateDRB*26 | 1 | 0,015 | 6 | TateDRB*53 | 4 | 0,061 | 77 | |
| TateDRB*27 | 1 | 0,015 | 66 | TateDRB*54 | 1 | 0,015 | 4 | |
| TateDRB*28 | 10 | 0,153 | 2341 | TateDRB*55 | 1 | 0,015 | 4 | |
| TateDRB*29 | 11 | 0,169 | 1338 | TateDRB*56 | 2 | 0,030 | 7 | |
| TateDRB*30 | 7 | 0,107 | 161 | TateDRB*57 | 1 | 0,015 | 4 | |
| TateDRB*31 | 2 | 0,030 | 450 | TateDRB*58 | 1 | 0,015 | 6 | |
| TateDRB*32 | 1 | 0,015 | 33 | TateDRB*59 | 1 | 0,015 | 4 | |
| TateDRB*33 | 2 | 0,030 | 885 | TateDRB*60 | 1 | 0,015 | 4 | |

| | | | * | * | | | | | * * | | | | | | * | | | | | • | • | * * | | | | * | * * | | * | | * | | * * | | • • |
|-------------|-----|---------|------------|-------------|------------|-----|---------|--------|----------|------------|-------|--------|-------|---|------------|-------|----------|---|------------|-----|-----|------|----------|------------|--------------|-------|------------|-------|-----|-------|-----------|-----|-------|-----|-------|
| TateDRB*01 | RΝ | /Н | F - L | . Q I | RQI | Y P | N Q | ΕE | ΗV | н | FD | S D | V G | Е | FV | AV | ΤI | L | ER | ΡD | ΑE | ΥY | NS | R K | DL | LΕ | DS | R A | ΑA | V D | TF | CR | ΗN | ΥR | VG |
| TateDRB*02 | | | | | | | | | . L | | | | | | | | | | G. | | | | | | | | | | | | | | | | |
| TateDRB*03 | | | | | | | | | | | | | | | | | | | | | | | | . R | | | | | | | | | | | |
| TateDRB*04 | | | | D. | Υ. | | R | | Υ. | R | | | | | | | | | G. | . s | . K | . W | τ | Q . | | | . W | 7 | | | . Y. | | | | |
| TateDRB*05 | | | | D. | Υ. | | R | | Υ. | R | | | | | | | | | G. | . S | . K | . W | ſ | Q. | | | . W | 1 | | | | | | | |
| TateDRB*06 | | | | D . | Υ. | | R | | Υ. | R | | | | | ΥA | | | | GΟ | . т | | . W | t | Q. | ΝI | | QR | | E | | К- | | | | . F |
| TateDRB*07 | | | | D. | Υ. | | R | | Υ. | R | | | | к | ΥA | | | | G. | | | . W | t | ō. | | | R | | | | . Y | | Υ. | | |
| TateDRB*08 | G. | R | Y | н. | Υ. | | R | | ΤA | R | | | | | ΥA | | | | G. | | | . W | 7 | ō. | ΝI | | οк | | Е | | . v | | | | |
| TateDRB*09 | | R | Y | н. | Υ. | | R | | ΤA | R | | | | | ΥA | | | | G. | | | . W | T | õ. | ΝI | | οĸ | | Е | | . v | | | | |
| TateDRB*10 | | R | Y | н. | Υ. | | R | | ТА | R | | | | | ΥA | | | | G. | . т | | . W | <i>.</i> | õ. | ΝI | | οĸ | | Е | | . v | | | | |
| TateDRB*11 | | R | Y | н. | Υ. | | R | | ТА | R | | | | | ΥA | | | | G. | . т | | . W | r | õ. | SI | | οĸ | | Е | | . v | | | | |
| TateDRB*12 | | R | Y | H. | Y. | | R | | T A | R | | | | | ΥA | | A | | G. | . т | | w | 7 | ο. | NI | | оĸ | | Е | | . v | | | | |
| TateDRB*13 | | R | Y | н. | Υ. | | R | | ΤA | R | | | | | ΥA | | | | G. | | | . W | т. т | ō. | ΝI | | οĸ | | Е | | . v | | | | |
| TateDRB*14 | | R | Y | Ε. | н. | | W | | Y A | R | | | | | | | | | G. | | | . W | T | õ. | | | | | | | | | | | A A |
| TateDRB*15 | | R | Y | E. | H. | | W | | Y A | R | | | | | | | | | G. | | | W | 7 | ò. | | | | | | | | G | | H. | AA |
| TateDRB*16 | | R | Y | Ε. | н. | | w | | Y A | R | | | | | | | | | G. | | | . W | | õ. | | | | | | | | | | | ΑT |
| TateDRB*17 | | R | Y | E. | н | 7 | R | | . N | 4. | | | | | Υ. | ν. | | | G. | . 1 | | . F | | | . 1 | | . R | | | | к - | | | | |
| TateDRB*18 | | R | Y | E | н | 7 | R | | . N | 4 | | | | | | | | | G. | I | | F | | | . 1 | | | | | | | | | | |
| TateDRB*19 | | R | v | E | н | , | R | | · N | 1 | | | | | | | | | G | Ť | | HF | | | Ť | | | | | | | | | | |
| TateDRB*20 | | R | v | Ē. | v | | R | • • | v | | | | | | v | | | | G | s | | F | • • | 0 | w | | R | | T | • • | к. | | • • | G | |
| TateDRB*21 | | R | v | ~ . | v | | R | ĸ | v | | | N | | | • • | | | | G . | | | w | , | õ | NI | P | O K | | E | • • | | | • • | | |
| TateDRB*22 | | R | v | | v | | R | ĸ | v | | | N | | - | | | | | G | | | w | , | õ. | SI | • • | 0 K | | F | | | | | • • | |
| TateDRB*23 | | R | Y . | | Ŷ | | R | ĸ | Ŷ | | | N | | | | | | | G | | ĸ | · w | , | õ | · · | | Y K | | τõ | | AY | | | | |
| TateDRB*24 | | R | н. | E | н | , | w | | Y A | R | | | | 1 | | | | 1 | G | | | w | , | õ | 1.1 | | | | ~ | | | | | | |
| TateDRB*25 | | R | - R | s. | v | | s | • • | 04 | R | | | | | • • | · · | | | 6 0 | | | w | , | õ. | · · | D | KF | | : 0 | | VI | ĸ | • • | • • | Ť |
| TateDRB*26 | • • | R | P | | v | | ŝ | • • | 0 4 | R | | | | | • • | ÷, | | | G O | | | w | , | <u>0</u> . | Ť | . D | KE | | i õ | | VI. | ĸ | , c | • • | F |
| TateDRB*27 | | R | r | | v | | s | | 0 4 | R | | | | - | • • | · . | • • | | G O | | | w | , | õ. | Ť | . D | KE | 0 | i õ | | VI. | ĸ | | | F |
| TateDRB*28 | | R | R | с Б., | v | | s | | 0 4 | R | | | | - | • • | · . | • • | | a o | | | w | , | õ. | Ť | · D | KE | ~ ~ | ž | | VI. | ĸ | • • | • • | Ē |
| TateDRB*20 | | R | 1 | с Б. | н. | | R | | F | 0 | | | | 1 | v | v | | 1 | G | Ť | | . F | • • | õ. | w | | R | | Ŷ | | к. | | | G | v |
| TateDRB*30 | | R | | F. | н | | R | • • | • | ~ | | G | | | • • | | | | G . | ÷÷ | | w | , | õ. | | | V K | | ÷ 0 | | AV | | • • | | |
| TateDRB*31 | • • | R | | F. | н | | R | • • | R | | | | | | • • | | | | с. с | ÷ | | w | , | ŏ. | | | v K | | | | ΔV | | • • | • • | |
| TateDRB*32 | | R | | F. | н | | P | | n . | | | | | G | • • | | | | G . | ÷ | | w | , | õ. | | | v K | | | | ΔV | | | | |
| TateDRB*33 | | R | | Е. Е | н | | P | | • • | | | | м. | • | • • | | | | G. | ÷÷ | | w | , | õ. | | | VK | | | | ΔV | | • • | • • | |
| TateDRB*34 | | P | | Е. Е | ш. | | P | | | <u>ر</u> ا | | | | 1 | | | | | G . | ÷÷ | | . w | , | ŏ. | | | VK | | | | A V | | | | · · · |
| TateDRB*35 | | P | | Б. Б | . u | | P | | | | | | P | 1 | | | | | с. с | ÷ ‡ | | . w | , | ŏ. | | | VK | | | | AV | | | | |
| TateDRB*36 | | R | | F. | н. | | R | • • | • • | | | LK | K . | | • • | | | | с. с | ÷ | | w | , | ŏ. | | • • | VK | | | • • | A V | | D | • • | |
| TateDRB*37 | • • | R | | F. | н. | | P | • • | • • | | | | | | • • | • • | | | G . | ÷ † | | w | , | õ. | | • • | VK | | 10 | | A V | | | • • | |
| TateDRB*38 | • • | R | | Ъ. П | v | н | P | • • | т | v | | | | | • • | | | P | G. | | | . w | , | Q. | 1.1 | | K | | 20 | | A V | | v | • • | |
| TateDRB*30 | 6 | IC. | | D . | | | R | | • • | | | | | | | | | 1 | u . | | | | • • | ς. | | | | | ~ ~ | | AI. | | • • | | |
| TateDRB*40 | G . | | | | | | ÷ | | | | | | | 1 | | | | | | | | | | | | | • • | | | | | | | | · · · |
| TateDPP*41 | 0. | | | | | | | • • | • • | | | | | - | • • | | | | | | | | • • | | | | • • | | - | • • | | | • • | • • | |
| TateDRB*//2 | Υ. | P | | н | v | | P | | т 4 | R | | | | - | v A | | | | G | | | w | , | 0 | | | P | | - | | v | | v | | |
| TateDRB*/3 | | R | г т | н. | v | | P | | ТА | R | | | | - | v A | | | | G 0 | | | w | , | õ. | NI | | O R | | F | v | K I | | • • | • • | F |
| TateDRB*44 | | P | г т | н. | v. | | P | | TA | P P | | | | | v A | | | | a a | ÷ | | . w | , | ŏ. | NI | | O P | | F | | K . | | | | · - |
| TateDRB*45 | | P | г т | ш. | v. | | P | | TA | | | | | 1 | v A | | | | 6 0 | ÷ | | . w | , | ŏ. | NI | | Q R | | F | G | v v | | | | |
| TateDRD 45 | | P | L T | . u | v | | P | • • | ТР | | | | | - | v A | | | | 6 0 | · + | | | , | Q. | NI | | | | Ē | . 0 | v v | | • • | • • | |
| TateDRD 40 | • • | R. R. | L | п. | | | D D | • • | T | R. | • • • | | | - | I A V A | | | | 00 | · + | | - 10 | , | Q. | NI | | QR | | E | | к v | | • • | | . r |
| TateDPD*49 | • • | P | с гт. | , <u></u> . | · · · | | ĸ | | 1 7 | I K | | | | - | 1 1 | | | | υų | | | - ** | • • | Q . | 18 1 | | QR | | E | | K | | | • • | |
| TateDRD +8 | 6 | D | L I V T | п | v | | P | | т. | | | | | 1 | · · | | 1.1 | 1 | G O | · . | 1.1 | | , | · · | NI | 1.1 | O P | | Ē | | v | | | | |
| TateDRD*49 | σ. | P | L | п. | . I . | | D | 1.1 | 1.7 | I K | | | 1.1 | 1 | 1 1 | ••••; | , · · | 1 | c Q | 11 | 1.1 | | , | Q. | 18 1 | 1.1 | VV | | | | AV | | | | |
| TateDRB*50 | | P | v | Е. | v. | | P | v | · · | | · · · | NT . | | | • • | 1.1 | . | | с. | | · · | | , | Q. | | | VV | | 10 | | AI. | | • • | | |
| TateDRB*31 | • • | ĸ | 1 | P | . 1 . V | | P | ĸ., | v . | P | | IN . | A | | • • | • • | | | с. С | | . K | . W | , | Q. | | | 1 K | , . (| J V | | AI. | • • | • • | • • | . r |
| TateDRB*32 | • • | P | T | . ע. די | - 1 . P | | л р | • • | т. т. | R. | | | л. | v | v A | • • | • • | | с. | . 5 | . K | W | , | Q. | 14 | | . W | | - | | | • | • • | • • | |
| TateDRB*33 | | K. P | ь v | п. | . ע. יע | | л. р | v · | т А v | r K | | | 1.1 | ĸ | 1 A | ••••• | | 1 | с. | 1.1 | 1.1 | . W | | ¥. | . 1 N T | | Q K O V | | E | | - v | | P | • • | . r |
| TateDKB*54 | | K. | 1 V | | . I. | | K. | к V | 1. V | - | • • • | NT | | 1 | • • | | | 1 | с. | | • • | . w | | Q. | N I | | ųκ | • • | E | | | | к. | · · | · · |
| TateDKB*55 | • • | ĸ | 1 V | | . Y . | | ĸ | к. | т. Е | | • • • | IN . | | 1 | | • • | | | ч. с | | . К | w | • • | Q. | . w | • • | o r | 11 | | | | • • | • • | • • | |
| TateDRB*30 | | R. D | 1 V | Е. U | . L . V | | R. D | | г. т. | ų. | | | | 1 | IA VA | | | | с. | · · | | . L | , | | N I | | QK | . 1 | - S | | . 1. V | | • • | • • | . v |
| TateDKB*3/ | σ. | K. D | 1 v | н. | . 1. | | R. D | | 1 A | IK. | | | | 1 | i A v A | | • • | D | с. | . 1 | | . W | , | Q. | IN I NI I | | QK | | E | | - V. | • • | • • | • • | • • |
| TateDKB*38 | 1.5 | K. | 1 V | н. | . I. | | K. | | 1 A | IK. | | | | 1 | IA VA | ••••• | | r | с. | | | . w | · · · | Q. | IN I NI I | | QK | • • | E | | - V. | • • | • • | • • | • • |
| TateDKB*39 | . 4 | 1 K | 1 | н. | . I. | | K. | • • | 1 A | IK. | | | | 1 | r A | ••••• | | 1 | с. | ÷ 2 | | . w | | Q. | IN I | | V K | , · · | E | | - V. | | • • | · · | · · |
| TateDKB*60 | • • | ĸ | | Е. | н. | | ĸ | • • | • • | | • • • | | | - | • • | • • | | | σ. | . 1 | | . w | • • | Q. | | | . w | | - | • • | . r | | • • | • • | • • |

Figure 2 – Alignment of *Tamandua tetradactyla* MHC DRB exon 2 alleles (amino acids). Dots represent identity to the first sequence, and asterisks represent putative ABS sites inferred after Brown *et al.* 1993. The first inferred ABS position in the alignment corresponds to position number 26 of the β -chain residues of the DRB gene in the human sequence (Brown *et al.* 1993).

Sixteen out of the 70 alleles were frequent and detected in more than 10% of the individuals (22.9%), whereas more than half (41, 58,5%) of alleles were rare and found in only one or two individuals. Rare alleles were distributed in all sampling regions (Fig. 1). All alleles,

also the rare ones were confirmed and validated since they occurred in both independent amplicon replicates (Table 1, Fig. 4a).

Number of DRB alleles across all individuals varied between two and 13, indicating that at least seven loci were amplified. The general MHC diversity features are summarized in Table 2. A total of 75 nucleotide sites were invariable, and 114 were variable. Among these, 35 sites were singletons and 79 sites were parsimony informative. Overall allele sequence diversity was 1.000 ± 0.003 , and nucleotide diversity was 0.132 ± 0.004 . When considering nucleotide sequences, all inferred ABS sites were variable in at least one position of the codon. In the translated amino acid sequences, all but two inferred ABS positions were variable (Figure 2).

The ratio *dN/dS* was 2.94, *i.e.*, a higher value of *dN* than expected in neutral regions was observed, which is compatible with a scenario where this genomic region is under positive selection. This ratio was clearly driven by ABS positions, which held the most part of non-synonymous substitutions. The two-tailed Z-test corroborated the hypothesis of positive selection acting on ABS sites (p-value=0.035) (Table 3). The estimates of genetic distances between alleles were more pronounced in among amino acid sequences than nucleotide sequences, which points out to the functionality of these different alleles.

Table 2 – Genetic diversity in the overall *Tamandua tetradactyla* data set and in each biome for MHC class II DRB exon2 and for eight microsatellite loci (N= sample size; H_O/H_E = observed and expected heterozygosity; mean number of alleles and private alleles are averaged across all loci).

| | | | | MHC | Microsatellites | | | | | | |
|--------------------|----|----------------------------------|---------------------------------|---------------------------|-----------------------|---|-----------|-------------------|-------------------|--|--|
| Dataset | N | DRB nucleoti de alleles | DRB amino acid alleles | Nucleotide alleles per | Amino ac | id sequences | Mean | Mean number of | Mean number of | | |
| | | (195 bp) | (65 aa) | individual | Variable positions | Mean number of differences (+s.e.) | H_0/H_E | alleles | alleles | | |
| All samples | 65 | 70 | 60 | 1-13 | 53/65 (82%) | $14.602 (\pm 1.977)$ | 0.38/0.49 | 2.83(±0.69) | $0.27(\pm 0.13)$ | | |
| Atlantic Forest | 29 | 43 | 36 | 1-11 | 48/65 (74%) | 14.311 (±1.941) | 0.40/0.52 | 6.00(±1.55) | 0.75(±0.25) | | |
| Amazon Forest | 9 | 32 | 28 | 3-10 | 46/65 (71%) | 14.860 (±2.116) | 0.50/0.56 | 5.13(±1.30) | 0.75(±0.36) | | |
| Cerrado | 16 | 32 | 29 | 2-9 | 45/65 (69%) | 15.481 (±2.138) | 0.36/0.48 | 5.13(±1.30) | $0.50(\pm 0.26)$ | | |
| Caatinga | 3 | 15 | 14 | 2-5 | 33/65 (51%) | 13.124 (±1.876) | 0.33/0.47 | $2.875(\pm 0.47)$ | $0.125(\pm 0.25)$ | | |
| Pantanal | 8 | 21 | 15 | 4-13 | 38/65 (58%) | 15.048 (±2.135) | 0.29/0.40 | 3.50 (±0.92) | $0.00(\pm 0.00)$ | | |

Table 3 – The estimated rates (\pm s.e.) of dN and dS substitutions for all sites (65 codons), ABS (21 codons) and non-ABS (44 codons), and their ratio (dN/dS) for MHC class II DRB exon 2 sequences in *Tamandua tetradactyla*. The probability that dN and dS are different (null hypothesis of neutrality is dN=dS) was calculated using a two-tailed codon based Z-test, with 1000 bootstrap replications to obtain the *p*-value. Significance at 0.05 is indicated by an asterisk.

| Positions | dN | dS | dN/dS | p-value |
|-----------|--------------------|--------------------|-------|---------|
| All sites | 0.306(±0.081) | $0.104(\pm 0.044)$ | 2.94 | 0.142 |
| ABS | $0.279(\pm 0.070)$ | 0.138(±0.058) | 2.02 | 0.035* |
| Non-ABS | 0.099(±0.021) | $0.143(\pm 0.041)$ | 0.69 | 0.948 |

Comparison of individual MHC genotypes obtained by the SSCP and NGS approach

A comparison of DRB putative alleles retrieved from SSCP and NGS techniques for the same ten individuals revealed that extra alleles could be detected using NGS in all but two of the samples (Student's t-test for paired data=3.772, p-value=0.004, Table S3 Supporting Information, Fig. 3). The SSCP technique was able to detect 25 MHC Class II DRB exon 2 alleles in the nucleotide level and 20 alleles in the amino acid level among the ten samples analyzed for this approach. In SSCP, the number of different alleles found in one individual varied between three and eight (Table S3, Supporting Information), which would indicate the presence of at least four DRB loci, an underestimation of this parameter if compared to NGS.



Figure 3 – Comparison of number of MHC Class II DRB exon 2 putative alleles retrieved for ten individuals through SSCP (white bars) technique and Next Generation Sequencing (NGS, black bars). Significantly more MHC alleles per individual were detected by the NGS approach (Student's T test for paired data = 3.772, p-value=0.004).

MHC and diversity pattern across biomes

Some alleles were more frequent and common than others (Table 1). Alleles present in more than 10% of individuals (TateDRB*01, 40, 04, 39b, 01a, 25, 44c, 03, 29, 38, 44a, 28 – Figure 4A) were distributed in three or more biomes. Some other alleles, less frequent (<10% of individuals) but yet not rare (present in two or less individuals), were likewise distributed in different landscapes (TateDRB*48, 17, 37, 42, 09, 53, 43, 49).

Therefore, alleles with high or intermediate frequency showed to be distributed in several different landscapes. When considering samples partitioned by biomes, Atlantic Forest presented a set of sequences with most variable positions (74%), and Caatinga with the

least (51%). The highest mean number of differences between sequences was found in Cerrado (Table 2).

DRB alleles with frequencies high as 0.5 occurred in all five biomes indicating their wide geographic range (Fig. 4a, 4b, Fig S1), whereas alleles with low frequencies (0.1-0.2) occurred in three or less biomes (Figure 4b).



Figure 4 – (A) Frequency of MHC class II DRB nucleotide alleles in 65 individuals. Sixteen alleles were detected in more than 10% of the individuals, whereas more than half of alleles are found in only one or two individuals, but confirmed and validated since they occur in both independent replicates. (B) Boxplots of MHC class II allele frequencies divided in intervals of 0.1, and the number of biomes (1 to 5) in which they occur.

A total of 33 (50.8%) DRB amino acid alleles were private to single biomes (Fig. 1). Amazon Forest presented the highest number of private alleles: 11 out of 33 alleles (33%). Atlantic Forest had 12 private alleles (30%), Cerrado had 5 (15%), Caatinga had 3 (9%) and Pantanal had 2 (6%).

Phylogenetic tree of DRB alleles resulted in roughly nine major clades, with variable branch lengths reflecting the differences between alleles (Fig. 5). Among the nine clades displayed in the phylogeny, seven contain at least one geographically disseminated allele of high or intermediate frequency. Considering that each clade could possibly represent one locus or closely related loci, these seven clades would be composed of one widespread allele (or more) and a few biome-specific alleles. The DRB phylogenetic tree showed that the basal cluster of alleles comprised mainly those with high frequency (TateDRB*01, 03, 39, 40), which are probably early alleles and could represent ancient loci. Other clades are predominantly clusters of alleles specific to biomes AF and AM. There is also constant grouping of alleles between these biomes along the phylogenetic tree, and vast allele sharing between them (Fig. S1, Fig. S2).

Figure 5 Bayesian _ phylogenetic tree of MHC Class II DRB alleles with values of posterior probabilities for nodes. Terminals are labeled with biomes in which alleles occur: Atlantic Forest (triangle), Amazon Forest (square), Cerrado (circle), Caatinga (diamond), Pantanal (star).



Comparison of DRB and microsatellite diversity and correlation with geographic distance

Allelic richness corrected for three individuals (all biomes) ranged between 2.6 (CA) and 3.7 (AM); for eight individuals (excluding CA) ranged between 4.2 (PT) and 6.8 (AM); and for nine individuals (excluding CA and PT) ranged between 5.8 (CE) and 7.2 (AM). In all cases, the highest allelic richness was found in Amazonian biome (Fig. 6A). A similar pattern emerged for microsatellite data. For a minimum of three samples, allelic richness ranged between 2.3 (PT) and 3.0 (AM); for eight samples it ranged from 3.6 (PT) and 4.9 (AM); and for nine samples it ranged from 4.2 (AF) and 5.1 (AM) (Fig. 6B). In all cases, a higher value of allelic richness was found for DRB data (Student's t-test for paired data=3.933, p-value = 0.011).

In DRB, nucleotide diversity ranged between 0.117 (CA) and 0.142 (CE) for the whole sequence. Nucleotide diversity in ABS ranged between 0.011 (CA) and 0.013 (CE). Overall, nucleotide diversity inside ABS was always higher than the entire sequence (Fig. 6D). Since nucleotide diversity measures genetic variation in sequences, it is clear that ABS are more diverse either considering the whole sample set or inside biomes if compared to whole sequences. In fact, non-synonymous nucleotide substitutions were more frequently observed in ABS than non-ABS for all sequences and in each biome separately (Table 4).



Figure 6 – Allelic richness values (\pm standard deviations) across biomes for MHC class II DRB exon 2 (A) and for microsatellite loci (B). Plots are shown after standardization for a minimum of three samples (all biomes), eight samples (excluding CA) and nine samples (excluding CA and PT), from left to right. (C) Mantel tests showing the relationship of individual genetic and geographic distance (Log 10) for DRB alleles and microsatellite loci, and their respective correlation values. (D) Nucleotide diversity (\pm standard deviations) across DRB exon 2 alleles for the overall data set and for each biome separately. Values for the whole sequence (black dots) and for ABS only (white dots) are displayed.

Table 4 – Average nucleotide and amino acid genetic distances among *Tamanduatetradactyla* MHC class II DRB exon 2 alleles.

| Dataset | K2P nucleot | ide distance | | Poisson-corrected amino acid distance | | | | | | |
|-----------|---------------|---------------|---------------|---------------------------------------|---------------|---------------|--|--|--|--|
| | All sites | ABS | Non-ABS | All sites | ABS | Non-ABS | | | | |
| All | 0.339(±0.219) | 0.302(±0.071) | 0.121(±0.021) | 0.306(±0.059) | 0.594(±0.260) | 0.217(±0.048) | | | | |
| Sequences | | | | | | | | | | |
| Atlantic | 0.105(±0.017) | 0.242(±0.045) | 0.109(±0.017) | 0.270(±0.043) | 0.468(±0.023) | 0.193(±0.036) | | | | |
| Forest | | | | | | | | | | |
| Amazon | 0.147(±0.018) | 0.248(±0.046) | 0.107(±0.017) | 0.274(±0.045) | 0.481(±0.132) | 0.194(±0.040) | | | | |
| Forest | | | | | | | | | | |
| Cerrado | 0.162(±0.018) | 0.275(±0.049) | 0.116(±0.019) | 0.291(±0.047) | 0.506(±0.133) | 0.208(±0.043) | | | | |
| Caatinga | 0.130(±0.018) | 0.233(±0.048) | 0.089(±0.018) | 0.241(±0.043) | 0.446(±0.120) | 0.164(±0.038) | | | | |
| Pantanal | 0.153(±0.019) | 0.254(±0.049) | 0.113(±0.019) | 0.285(±0.048) | 0.472(±0.139) | 0.213(±0.047) | | | | |

In microsatellites, the highest mean number of microsatellite alleles was observed in the Atlantic Forest, while the highest number of private alleles was observed both in Atlantic Forest and Amazon Forest. Also, mean H_O/H_E was higher in Amazon and Atlantic Forest than in other biomes (Table 2).

Positive and significant correlation between genetic and geographical distances was found for both datasets (DRB and microsatellites), although the degree of this correlation was stronger for microsatellites than for DRB (DRB: r=0.081, p-value=0.002; microsatellites: r=0.164, p-value=0.001) (Fig. 6C). When using a Partial Mantel test to correct for the effects of geographic distance (*i.e.*, testing the response of genetic distance to different biomes, representing different environments), a similar positive correlation is observed for both MHC and microsatellites (MHC: r=0.354 p-value=0.001; microsatellites: r=0.351, p-value=0.001).

Discussion

In this study we examined the MHC Class II DRB exon 2 sequence of the lesser anteater for the first time. In fact, this was the first characterization and description of diversity of a MHC gene for a member of the magna-order Xenarthra, one of the basal lineages of placental mammals, along with Afrotheria (Murphy *et al.* 2001). Across a wide sampling region, we aimed to investigate if there are different composition of alleles specific to certain geographic regions which could be indicative of local adaptation, or if alleles are randomly distributed in space.

MHC Class II DRB diversity

We were able to describe 60 DRB exon 2 alleles detected in 65 individuals of *Tamandua tetradactyla* on the amino acid level, and 70 on the nucleotide level (Table 1). Since our sampling strategy covered a broad territorial extension and many different landscapes (referred as "biomes" in this study), we believe to have reported the approximate extent of DRB exon 2 variation across the species distribution.

Considering the limited amount of samples successfully processed, this is a considerable high level of diversity, even for a locus already expected to be highly variable. This could be the due of two complementary facts. First, the individuals sampled were spread in a vast geographical area, comprising distinct environments. The sampling sites of *Tamandua tetradactyla* were chosen intending to represent the actual geographic distribution of the species along South America (mainly Brazil, which comprises the majority of the

continent's territory) (Fig. 1). Therefore, the samples were originated from a variety of distinct environments, and the high diversity of DRB exon 2 may mirror the species wide distribution. Secondly, the NGS technique used in this study provided a wider capture of MHC DRB alleles. NGS technique has proven to be more sensitive to identify different sequence variants than SSCP approach (Fig. 3), especially for alleles with low amplification efficiency (Sommer *et al.* 2013). Other studies have also reported this trend (*e.g.* Promerová *et al.* 2012), and this may have contributed to the high amount of alleles found for *Tamandua tetradactyla* as well.

Among mammals, different levels of MHC Class II genes diversity are shown. In nonvolant widely distributed mammals, as much as 36 alleles were described for more than one hundred samples of the European dog *Canis lupus* (Seddon & Ellegren 2002), and 52 alleles described for 25 samples of the Asiatic lion *Panthera leo* (Sachdev *et al.* 2005), both using traditional methods (respectively, SSCP and cloning followed by sequencing). On the other hand, 58 DRB alleles were described based on 36 samples of *Delomys sublineatus*, a small mammal with a smaller distribution (compared to lesser anteater), using the NGS approach (Sommer *et al.* 2013). Thus, the technique used to describe the diversity of MHC genes in a species sample set is relevant, along with sample size and the biology of the species.

Nevertheless, the number of DRB alleles found in *T. tetradactyla* may be related to copy number variation in the species. In an extreme example, one individual (TTPAN7) displayed 13 alleles in the nucleotide level, which reveals the high number of loci in the species. Therefore, the diversity described in this study is also the reflection of variation in

several DRB loci, probably generated by gene duplication, not distinguishable from each other.

Evidence for historical selection

A higher number of non-synonymous (*dN*) versus synonymous (*dS*) nucleotide substitutions was found for the entire sequence of DRB exon 2 gene and in putative ABS, but not in non-ABS, leading to elevated *dN/dS* ratios (Table 3). This scenario, along with the known biological significance of MHC Class II genes in peptide presentation for the immune system, is compatible with the hypothesis of positive selection acting in this region. Also, non-synonymous nucleotide substitutions were more frequently observed in the ABS than non-ABS for all sequences and in each biome separately, as well as higher nucleotide diversity (Table 4, Fig. 6D), an indicative of selection acting more strongly on ABS. These findings corroborates literature on functional importance of these sites (Doherty & Zinkernagel 1975; Hughes & Nei 1988) and supports evidence for the correct analogy of ABS positions inferred from human after Brown *et al.* (1993) (Fig.2).

Distribution and diversity of DRB alleles among biomes

Our results show that common and frequent DRB alleles are geographically widespread, occurring in all biomes, while other less frequent alleles are restricted to one biome. In fact, all biomes showed some level of private alleles in combination with other rare alleles (Fig. 1). These common alleles are most likely ancient, and supposedly very important to the immune response of the species, with probably similar pathogen-recognition capabilities, and possibly recognizing a vast array of common antigens. Therefore, they are spread to

different populations and are likely kept in high frequency in all of them by positive selection.

At the same time, several rare alleles were found to be scattered among biomes, and some of them were private to one biome. Nevertheless, the alleles were highly divergent, as evidenced by the phylogenetic tree, pairwise genetic distance both in nucleotides and amino acids, and especially in ABS, which showed higher nucleotide diversity and genetic distance than the whole sequence. These parameters are indicative of dissimilar functional properties of these alleles, and the fact that private alleles were found within biomes is considered as evidence that specific alleles may be important in environments which harbor specific pathogens. Hence, we found different allelic compositions throughout biomes, comprising common, rare and private alleles. Moreover, rainforest biomes displayed the highest diversity, depicted by higher number of private alleles and higher allelic richness. Between the biomes sampled in this study, the rainforests in fact display higher general diversity among all taxa. As already mentioned, the Amazon and the Atlantic Forest are regions with high levels of biodiversity and endemism. Thus, they are also expected to harbor more pathogens than drier or less diverse biomes. This elevated exposure to pathogens is expected to be mirrored in the MHC genetic diversity of the species, as it was observed. On the other hand, our data also controlled for demographic patterns (neutral markers), and the configuration of genetic diversity seemed to be similar to MHC. The effect of demographic processes such as drift and migration are not negligible on shaping the MHC diversity throughout the species ranges. Studies in humans both on MHC (Prugnolle et al. 2005) and on several genomic SNPs with signs for positive selection (Fumagalli et al. 2011), correlating with pathogen pressure, show a strong signature of demography despite the observed influence of pathogen mediated selection. In fact, Fumagalli *et al.* (2011) argued that the diversity of the local pathogenic environment could represent a predominant driver of local adaptation, and although background demography usually makes the strongest contribution in explaining the genetic variance among populations, specific alleles could be correlated with a certain pathogenic environment. Studies with non-model organisms also showed that demography is important on shaping MHC diversity, even in the presence of clear signatures of selection in this gene family (Alcaide 2010; Miller *et al.* 2010), with isolation by distance being a common pattern emerged both from microsatellites and MHC. In fact, demography probably does not overcome balancing selection when populations are fragmented and small (Strand *et al.* 2012). Nevertheless, differences in spatial distribution of alleles (and genetic differentiation measured by F_{ST} , when possible) indicate possible local adaptation to different environments (Ekblom *et al.* 2007; Miller *et al.* 2010).

Although our study does not have data about pathogens that interact with our *T*. *tetradactyla*'s samples, it is possible that the higher MHC diversity found in the rainforest biomes could be the product of a more diverse pathogenic environment (and hence, a stronger selective pressure), coupled with the effects of demography as well, since these biomes display a greater diversification in general than the other sampled landscapes. Also, the sample size used in this study was relatively small to effectively investigate differential allele selection in distinct environments. Bigger samples sizes with finer spatial distribution coupled with information about pathogen interaction could help to clarify to what extent (and which) private alleles actually represent local adaptations in the species.

Conclusions

Our results indicate that MHC variation across lesser anteater populations in different biomes show clear signs of natural selection, as well as probable local adaptation driven by different compositions of pathogens in distinct environments. Rainforest biomes (Amazon and Atlantic Forest) show higher overall neutral and adaptive genetic diversity, as expected from known patterns of species diversity in South America biogeography. The higher proportion of private alleles and allelic richness in these biomes point out to a different composition of potentially important genetic variation that should be taken into account for future conservation plans.

Acknowledgements

We would like to thank Fabrício R. Santos, Flávia Miranda and José Abílio Ohana for sample donations. We would also like to thank Anke Schmidt, Saskia Wutke and Susan Mbedi for wet laboratory assistance. C.L.Clozato was supported by CAPES (PhD and PSDE) scholarship, and J. S. Morgante was granted with a research award from FAPESP (08/52207-0).

References

- Alberto F (2006) StandArich v1.0: an R package to estimate population allelic richness using standardized sample size. Available from: http://www.ccmar.ualg.pt/maree/software.php?soft=sarich.
- Alcaide M (2010) On the relative roles of selection and genetic drift in shaping MHC variation. *Molecular Ecology*, **19**, 3842–4.
- Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology*, **17**, 179–224.

- Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, **16**, 363–77.
- Brown JH, Jardetzky TS, Gorga JC et al. (1993) 3-Dimensional structure of the human class-II histocompatibility antigen HLA-DR1. *Nature*, **364**, 33–39.
- Clapperton CM (1993) *Quaternary Geology and Geomorphology of South America*. Elsevier, New York.
- Clozato CL, Moraes-Barros N, Morgante JS (2014) Main biogeographic patterns in a widedistributed mammal from South America: genetic structure assessment of the lesser anteater (*Tamandua tetradactyla*, Pilosa), *in prep*.
- Clozato CL, Moraes-Barros N, Santos FR, Morgante JS (2014) Historical and non-invasive samples: a study case of genotyping errors in newly isolated microsatellites for the lesser anteater (*Tamandua tetradactyla* L., Pilosa). *Molecular Ecology Resources*, **14**, 531-540.
- Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at H-2 gene complex. *Nature*, **256**, 50–52.
- Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2012) Geneious v5.6.5 created by Biomatters.
- Dunn RR, Davies TJ, Harris NC, Gavin MC (2010) Global drivers of human pathogen richness and prevalence. *Proceedings of the Royal Society of London B*, **277** (1694): 2587-2595.
- Ekblom R, Saether SA, Jacobsson P *et al.* (2007) Spatial pattern of MHC class II variation in the great snipe (Gallinago media). *Molecular Ecology*, **16**, 1439–51.
- Fittkau EJ, Illies J, Klinge H, Schwabe GH, Sioli H (1969) *Biogeography and Ecology in South-America*. Junk, The Hague.
- Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admettla A, Pattini L, Nielsen R (2011) Signatures of environ- mental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet*, **7**, e1002355.
- Gardner AL (2008) Magnaorder Xenarthra. In: *Mammals of South America, Vol 1: Marsupials, Xenarthrans, Shrews, and Bats* (ed. Gardner AL), pp 127–176, The University of Chicago Press, Chicago 2008.
- Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**(7),1-19.
- Guernier V, Hochberg ME, Guegan JF (2004) Ecology drives the worldwide distribution of human diseases. *PLoS Biology*, **2**, 740–746.

- Hedrick PW, Kim KJ (2000) Genetics of complex polymorphisms: Parasites and maintenance of the major histocompatibility complex variation. In: *Evolutionary Genetics: from Molecules to Morphology* (eds Singh RS, Krimbas CB), pp. 204–234. Cambridge University Press, Cambridge.
- Hillebrand H (2004) On the generality of the latitudinal diversity gradient. Am Nat, 163, 192–211.
- Hoorn C, Wesselingh FP, ter Steege H, Bermudez MA, Mora A, Sevink J, et al. (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, **330**, 927–931.
- Horton R, Gibson R, Coggil P, Miretti M, Alceck RJ, *et al.* (2008) Variation analysis and gene annotation of eight MHC haplotypes: The MHC Haplotype Project. *Immunogenetics*, **60**, 1-18.
- Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class-I loci reveals overdominant selection. *Nature*, **335**, 167–170.
- Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics*, **32**, 415–435.
- Jeffery KJM, Bangham CRM (2000) Do infectious diseases drive MHC diversity? *Microbes Infect.* 2, 1335–1341.
- Joly CA, Aidar MPM, Klink CA, McGrath DG, Moreira AG, Moutinho P, Nepstad DC, Oliveira AA, Pott A, Rodal MJN, Sampaio EVSB (1999). Evolution of the Brazilian phytogeography classification systems: implications for biodiversity conservation. *Ciência e Cultura* **51**(5/6), 331-348.
- Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllesten UB, Heise ER, Jones PP, Parham P, Wakeland EK, Watkins D (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics*, **31**, 217–219.
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.
- Mares M, Barquez RM, Braun JK, Ojeda RA (1996) Observations on the mammals of Tucuman Province, Argentina: Systematics, distribution, and ecology of the Didelphimorphia, Xenarthra, Chiroptera, Primates, Carnivora, Perissodactyla, Artiodactyla, and Lagomorpha. Annals of Carnegie Museum, 65, 89–152.
- Miller HC, Allendorf F, Daugherty CH (2010) Genetic diversity and differentiation at MHC genes in island populations of tuatara (Sphenodon spp.). *Molecular ecology*, **19**, 3894–908.
- Mittelbach GG, Schemske DW, Cornell HV et al. (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315–331.

- Morrone JJ (2004) Panbiogeografia, componentes bióticos y zonas de transición. *Revista Brasileira de Entomologia*, **48**, 149–162.
- Morrone JJ (2006) Biogeographic areas and transition zones of Latin America and the Caribbean islands based on panbiogeographic and cladistic analyses of the entomofauna. *Annual Review of Entomology*, **51**, 467–494.
- Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ (2001) Molecular phylogenetics and the origins of placental mammals. *Nature*, **409**, 614–618.
- Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*, **3**, 418–426.
- Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Novak RM (1983) *Walkers Mammals of the world*. The Johns Hopkins University Press, Baltimore and London.
- Nunn CL, Altizer SM, Sechrest W, Cunningham AA (2005) Latitudinal gradients of parasite species richness in primates. *Diversity and Distributions*, **11**, 249–256.
- Paradis E, Schliep K, Potts A, Winter D (2012) Population and Evolutionary genetics analysis system. Available from: http://ape-package.ird.fr/pegas/.
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.
- Promerová M, Babik W, Bryja J *et al.* (2012) Evaluation of two approaches to genotyping major histocompatibility complex class I in a passerine-CE-SSCP and 454 pyrosequencing. *Molecular Ecology Resources*, **12**, 285–92.
- Prugnolle F, Manica A, Charpentier M, Guégan JF, Guernier V, Balloux F (2005) Pathogen-driven selection and worldwide HLA class I diversity. *Curr Biol*, **15**, 1022–1027.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck (2012) MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**(3), 539-542.
- Sachdev M, Sankaranarayanan R, Reddanna P, Thangaraj K, Singh L (2005) Major histocompatibility complex class I polymorphism in Asiatic lions. *Tissue Antigens*, 66, 9–18.

- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
- Seddon JM, Ellegren H (2002) MHC class II genes in European wolves: a comparison with dogs. *Immunogenetics*, **54**, 490–500.
- Slade RW, McCallum HI (1992) Overdominant vs. frequency-dependent selection at MHC loci. *Genetics*, **132**, 861–862.
- Sommer S, Courtiol A, Mazzoni CJ (2013) MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC Genomics*, **14**, 542-559.
- Sommer S, Tichy H (1999) Major histocompatibility complex (MHC) class II polymorphism and paternity in the monogamous *Hypogeomys antimena*, the endangered, largest endemic Malagasy rodent. *Molecular Ecology*, **8**, 1259-1272.
- Sylvestre F (2009) Moisture pattern during the Last Glacial Maximum in South America. In: Past Climate Variability in South America and Surrounding Regions, Developments in Paleo environmental Research (eds Vimeux F, Sylvestre F, Khodri M). Springer, Netherlands.
- Takahata N, Nei M (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of Major Histocompatibility Complex loci. *Genetics*, **124**, 967–978.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**, 2725-2729.
- Veloso HP, Rangel-Filho ALR, Lima JCA (1991) *Classificação da vegetação brasileira adaptada a um sistema universal*. MEFP/IBGE/DRNEA, Rio de Janeiro. 123p.
- Webb SD (2006) The Great American Biotic Interchange: patterns and processes. *Annals of the Missouri Botanical Garden*, **93**, 245-257.
- Wetzel RM (1985) The identification and distribution of recent Xenarthra (=Edentata). In: *The evolution and ecology of armadillos, sloths and vermilinguas* (ed. Montgomery GG), pp 5–21. Smithsonian Institution Press, Washington, DC.
- Yuhki N, O'Brien SJ (1990) DNA variation of the mammalian major histocompatibility complex reflects genomic diversity and population history. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 836–40.

Conclusão Geral

Conclusão Geral

O objetivo desta tese de doutorado foi descrever a diversidade genética e estruturação populacional da espécie *Tamandua tetradactyla* ao longo de sua distribuição, principalmente no Brasil, utilizando para isso marcadores genéticos neutros e não-neutros.

A começar pela grande variedade de tipos de amostras usadas (pele de museu, sangue, músculo, carcaça de atropelamento, pelo arrancado, etc), foi necessário um estudo formal para validar o uso de determinados materiais para etapas posteriores no estudo da variação genética da espécie. Portanto, no capítulo 1 avaliamos o potencial do uso de amostras de pele de museu (amostras históricas) e de pelo arrancado com o bulbo (amostras nãoinvasivas) como material biológico válido para procedimentos de biologia molecular, comparando a performance da genotipagem de microssatélites nestas amostras e em amostras frescas, com DNA íntegro. Foi possível reportar uma qualidade de resultados superior de amostras de pelo arrancado em relação às peles de museu. Através de um índice de qualidade assinalado a cada amostra e lóco, foi possível detectar e quantificar erros de genotipagem, selecionar amostras e lócos com resultados de genotipagem confiáveis para análises posteriores, e excluir resultados duvidosos.

Uma vez excluídos erros de genotipagem e amostras com qualidade duvidosa, o estudo descritivo da diversidade genética e estruturação populacional de *Tamandua tetradactyla* através de marcadores neutros pôde ser realizado.

Através de uma amostragem extensa que perpassou os principais biomas brasileiros (Floresta Amazônica, Mata Atlântica, Pantanal, Cerrado e Caatinga), foi possível

correlacionar a diferenciação genética de populações/indivíduos de tamanduá-mirim com as diferentes paisagens em que a espécie ocorre.

Foram utilizadas duas abordagens distintas de análise: agrupamento de indivíduos em populações (*a priori*) e uma análise de atribuição baseada puramente em indivíduos (*a posteriori*). As análises resultaram em padrões similares, destacando principalmente uma diferenciação genética entre indivíduos provenientes da Floresta Amazônica em relação aos demais. Valores significativos de F_{ST} foram encontrados entre as populações da Amazônia e populações adjacentes na análise *a priori*, e houve uma clara distinção dos indivíduos amazônicos nos gráficos de atribuição da análise *a posteriori*. Além disso, foi possível observar uma composição genética complexa neste bioma, cujos indivíduos foram atribuídos a mais de um grupo: um *cluster* exclusivo amazônico, que demonstra maior diversidade genética em termos de riqueza alélica e heterorigozigose; e outro grupo que é similar geneticamente às populações fora da Amazônia.

A análise *a priori* também destacou uma diferenciação entre populações da Mata Atlântica do norte e do sul. Embora essa diferenciação tenha sido refletida em valores altos de F_{ST} entre estas populações, uma análise de correlação entre a distância genética e provável presença de barreiras ao fluxo gênico, corrigida pela distância geográfica entre as populações, revelou que a diferenciação encontrada é, provavelmente, produto de isolamento por distância. O padrão de atribuição genética dos indivíduos em grupos distintos da análise *a posteriori* confirmou este resultado.

Em geral, foi encontrada uma fraca estruturação genética ao longo das populações de tamanduá mirim amostradas, sobressaindo a diferenciação entre os indivíduos da Amazônia

e os demais, cuja diversidade genética também se mostrou mais elevada. Ao longo das populações da Mata Atlântica, um padrão de isolamento por distância foi encontrado. Populações dos biomas Pantanal e Cerrado não demonstraram diferenciação genética significativa em relação às demais.

O capítulo 3 foi dedicado à investigação da distribuição da diversidade genética de um marcador não-neutro, o gene DRB do MHC Classe II, região genômica conhecida por apresentar sinais clássicos de seleção positiva e/ou balanceadora. Através do uso de Sequenciamento de Nova Geração foi possível explorar a diversidade deste gene ao longo de populações de tamanduá-mirim distribuídas nos principais biomas brasileiros, e comparar esta diversidade com aquela encontrada nos marcadores microssatélites do capítulo 2.

Foi encontrada uma alta diversidade do gene DRB, que exibiu uma variedade de alelos no nível de nucleotídeos e proteico, um provável reflexo de múltiplas cópias do gene na espécie combinada com a grande multiplicidade de ambientes que a espécie ocupa. Sinais claros de seleção positiva histórica agindo neste gene foram observados, como valores elevados de substituições não-sinônimas e maior diversidade nos códons de ligação de antígeno.

Uma maior riqueza alélica e maior número de alelos privados (exclusivos) foram observados nos biomas florestados, Mata Atlântica e Floresta Amazônica, especialmente nesta última. Possivelmente, estes ambientes também apresentam maior riqueza de espécies de patógenos (assim como riqueza de espécies em geral), o que representa uma elevada pressão do ambiente.
Além disso, os marcadores microssatélites (neutros) mostraram um padrão de diversidade genética similar ao gene DRB (não-neutro), demonstrando uma importante influência da demografia na diversidade e estrutura genética do Complexo Principal de Histocompatibilidade.

Anexos

Supporting Information

Table S1 List of samples used in this study. Sample ID, sample type (historical, noninvasive and fresh tissue datasets), institution of origin (Museu de Biologia Professor Mello Leitão – MBML, Universidade Federal de Santa Catarina – UFSC, Museu de História Natural do Capão da Imbuia – MNHCI, Museu Nacional – MN, Laboratório de Biologia Evolutiva e Conservação de Vertebrados – LABEC and Laboratório de Biodiversidade e Evolução Molecular – LBEM), Brazilian federal state and locality where specimen was collected, age (years, absolute numbers or fractions), and QI value obtained in this study. ¹Population of historical samples (n=29), ² Population of noninvasive samples (n=8), ³ Population of fresh tissue samples (n=13).

| Sample ID | Sample type | Instutution | State | Locality | Age | QI |
|-----------|-------------|-------------|-------|------------------------|-----|------|
| MBML1955 | Historical | MBML | ES | Santa Teresa | 15 | 0.62 |
| MBML2451 | Historical | MBML | ES | Santa Teresa | 10 | 0.61 |
| MBML2353 | Historical | MBML | ES | Viana | 10 | 0.78 |
| MBML2797 | Historical | MBML | ES | Vila Velha | 11 | 0.66 |
| MBML2799 | Historical | MBML | ES | Aracruz | 5 | 0.52 |
| MBMLsf1 | Historical | MBML | ES | Santa Teresa | N/A | 0.52 |
| MBMLsf2 | Historical | MBML | ES | Santa Teresa | N/A | 0.57 |
| MBMLsf3 | Historical | MBML | ES | Santa Teresa | N/A | 0.72 |
| MBMLsf4 | Historical | MBML | ES | Santa Teresa | N/A | 0.60 |
| UFSC771 | Historical | UFSC | SC | Ilha de Santa Catarina | 16 | 0.44 |
| UFSC844 | Historical | UFSC | SC | Ilha de Santa Catarina | 11 | 0.47 |
| UFSC3042 | Historical | UFSC | SC | Ilha de Santa Catarina | 4 | 0.76 |
| UFSC3790 | Historical | UFSC | SC | Xavantina | 12 | 0.52 |
| UFSC2833 | Historical | UFSC | SC | Ilha de Santa Catarina | 16 | 0.52 |
| UFCS838 | Historical | UFSC | MS | Campo Grande | 11 | 0.72 |
| UFSC3948 | Historical | UFSC | SC | Joinville | N/A | 0.81 |

| UFSC658 | Historical | UFSC | SC | Praia Grande | 16 | 0.89 |
|------------------------|------------|-------|----|----------------------|-----|------|
| MHNCI4036 ¹ | Historical | MHNCI | PR | Mauá da Serra | 6 | 0.38 |
| MHNCI5789 ¹ | Historical | MHNCI | PR | Morretes | 8 | 0.38 |
| MHNCI5989 ¹ | Historical | MHNCI | SC | Caxambu | 6 | 0.92 |
| MHNCI6048 ¹ | Historical | MHNCI | PR | Antonina | 7 | 0.70 |
| MHNCI5709 ¹ | Historical | MHNCI | PR | Icaraíma | 6 | 0.81 |
| MHNCI5780 ¹ | Historical | MHNCI | PR | Paranaguá | 6 | 0.72 |
| MHNCI5788 ¹ | Historical | MHNCI | PR | Paranaguá | 8 | 0.89 |
| MHNCI5082 ¹ | Historical | MHNCI | PR | Diamante do Norte | 7 | 0.71 |
| MHNCI5708 ¹ | Historical | MHNCI | PR | Diamante do Norte | 13 | 0.74 |
| MHNCI3967 ¹ | Historical | MHNCI | PR | Pinhão | 12 | 0.71 |
| MHNCI3971 ¹ | Historical | MHNCI | PR | Guarapuava | 2 | 0.62 |
| MHNCI3986 ¹ | Historical | MHNCI | PR | Blumenau | 12 | 0.87 |
| MHNCI3929 ¹ | Historical | MHNCI | PR | Arapoti | 12 | 0.66 |
| MHNCI3930 ¹ | Historical | MHNCI | PR | Reserva de Iguaçu | 10 | 0.77 |
| MHNCI3938 ¹ | Historical | MHNCI | SC | Água Doce | 18 | 0.66 |
| MHNCI2501 ¹ | Historical | MHNCI | PR | São José dos Pinhais | 26 | 0.72 |
| MHNCI1695 ¹ | Historical | MHNCI | PR | Piraquara | 52 | 0.92 |
| MHNCI370 ¹ | Historical | MHNCI | PR | Curitiba | 52 | 0.66 |
| MNHCI371 ¹ | Historical | MHNCI | PR | Curitiba | 64 | 0.55 |
| MHNCI165 ¹ | Historical | MHNCI | PR | Curitiba | 4 | 0.44 |
| MHNCI6206 ¹ | Historical | MHNCI | PR | Diamante do Norte | 3 | 0.92 |
| MHNCI6207 ¹ | Historical | MHNCI | PR | Morretes | 4 | 0.81 |
| MHNCI6208 ¹ | Historical | MHNCI | PR | Paranaguá | 4 | 0.55 |
| MHNCI6213 ¹ | Historical | MHNCI | PR | Matinhos | 4 | 0.74 |
| MHNCI6192 ¹ | Historical | MHNCI | PR | Diamante do Norte | 4 | 0.62 |
| MHNCI6196 ¹ | Historical | MHNCI | PR | Diamante do Norte | 4 | 0.77 |
| MHNCI6199 ¹ | Historical | MHNCI | PR | Morretes | 4 | 0.77 |
| MHNCI6200 ¹ | Historical | MHNCI | PR | Alexandra-Paranaguá | 24 | 0.70 |
| MHNCI1696 ¹ | Historical | MHNCI | PR | Matinhos | N/A | 0.77 |
| MHNCI2685 | Historical | MHNCI | PR | Paranaguá | 6 | 0.40 |
| MHNCI5795 | Historical | MHNCI | PR | Antonina | 6 | 0.33 |
| MHNCI5768 | Historical | MHNCI | PR | Paranaguá | 8 | 0.40 |
| MHNCI5241 | Historical | MHNCI | PR | Foz do Iguaçu | 19 | 0.66 |
| MHNCI2500 | Historical | MHNCI | PR | Contenda | 21 | 0.40 |
| MHNCI2503 | Historical | MHNCI | PR | Arapoti | 46 | 0.66 |
| MHNCI407 | Historical | MHNCI | PR | Congonhinhas | 26 | 0.53 |
| MHNCI1694 | Historical | MHNCI | PR | São Mateus do Sul | 55 | 0.53 |

| - | MNHCI249 | Historical | MHNCI | PR | Rio do Meio | 69 | 0.46 |
|---|----------|------------|-------|----|----------------|-----|------|
| | MHNCI20 | Historical | MHNCI | PR | Palmas | 69 | 0.46 |
| | MHNCI21 | Historical | MHNCI | PR | Curitiba | 72 | 0.26 |
| | MHNCI22 | Historical | MHNCI | PR | Curitiba | 65 | 0.66 |
| | MHNCI134 | Historical | MHNCI | PR | Castro | 75 | 0.53 |
| | MN2328 | Historical | MN | TO | Palma | N/A | 0.72 |
| | MN2329 | Historical | MN | TO | Palma | N/A | 0.47 |
| | MN2335 | Historical | MN | ТО | Palma | 80 | 0.70 |
| | MN2341 | Historical | MN | AM | Manacapuru | 80 | 0.61 |
| | MN2342 | Historical | MN | AM | Manacapuru | 75 | 0.72 |
| | MN3846 | Historical | MN | GO | Anápolis | 76 | 0.66 |
| | MN5050 | Historical | MN | PA | Curralinho | 76 | 0.74 |
| | MN5052 | Historical | MN | PA | Curralinho | 76 | 0.55 |
| | MN5053 | Historical | MN | PA | Curralinho | 75 | 0.76 |
| | MN5056 | Historical | MN | MS | Maracaju | 76 | 0.55 |
| | MN5057 | Historical | MN | PA | Curralinho | 76 | 0.71 |
| | MN5058 | Historical | MN | PA | Curralinho | 76 | 0.66 |
| | MN5060 | Historical | MN | PA | Curralinho | 75 | 0.62 |
| | MN5061 | Historical | MN | MS | Maracaju | 76 | 0.89 |
| | MN5063 | Historical | MN | PA | Curralinho | 76 | 0.79 |
| | MN5065 | Historical | MN | PA | Curralinho | 76 | 0.62 |
| | MN5067 | Historical | MN | PA | Curralinho | 76 | 0.75 |
| | MN5070 | Historical | MN | PA | Curralinho | 76 | 0.83 |
| | MN5071 | Historical | MN | PA | Curralinho | 76 | 0.85 |
| | MN5072 | Historical | MN | PA | Curralinho | 74 | 0.70 |
| | MN5509 | Historical | MN | PA | Santa Cruz | 74 | 0.89 |
| | MN5638 | Historical | MN | PA | Belterra | 74 | 0.55 |
| | MN5726 | Historical | MN | PA | Belterra | 74 | 0.61 |
| | MN5743 | Historical | MN | PA | Belterra | 71 | 0.85 |
| | MN5883 | Historical | MN | ES | Santa Teresa | N/A | 0.72 |
| | MN5967 | Historical | MN | AM | Juruá | 69 | 0.66 |
| | MN7620 | Historical | MN | MG | Araguari | 61 | 0.59 |
| | MN11607 | Historical | MN | SP | Formosa | N/A | 0.44 |
| | MN23954 | Historical | MN | BA | Ilhéus | 65 | 0.55 |
| | MN23956 | Historical | MN | MG | Viçosa | 32 | 0.76 |
| | MN24833 | Historical | MN | RJ | Silva Jardim | 32 | 0.43 |
| | MN24854 | Historical | MN | GO | Mineiros | 25 | 0.57 |
| | MN26851 | Historical | MN | RJ | Rio das Flores | 14 | 0.66 |

| MN50820 | Historical | MN | SP | Iguapé | 105 | 0.55 |
|---------------------------------|---------------------------------|-------|----|-----------------------|-----|------|
| MN2340 | Historical | MN | MG | Paraopeba | N/A | 0.53 |
| MN4538 | Historical | MN | GO | Anápolis | 76 | 0.66 |
| MN5062 | Historical | MN | PA | Curralinho | 76 | 0.80 |
| MN5066 | Historical | MN | PA | Curralinho | 72 | 0.66 |
| MN5507 | Historical | MN | ES | Santa Teresa | 74 | 0.59 |
| MN5510 | Historical | MN | PA | Belterra | 74 | 0.73 |
| MN23952 | Historical | MN | PA | Belterra | 34 | 0.53 |
| MN19956 | Historical | MN | PA | Belterra | N/A | 0.53 |
| TTPAN6 ² | Noninvasive | LABEC | MS | Corumbá | 2 | 0.61 |
| TTPAN7 ² | Noninvasive | LABEC | MS | Aquidauana | 2 | 0.70 |
| TTPAN8 ² | Noninvasive | LABEC | MS | Paranaíba | 1 | 0.81 |
| TTPAN9 ² | Noninvasive | LABEC | MS | Paranaíba | 1 | 0.78 |
| TTPAN10 ² | Noninvasive | LABEC | MS | Paranaíba | 1 | 0.50 |
| TTPAN11 ² | Noninvasive | LABEC | MS | Paranaíba | 1 | 0.66 |
| $TTPAN12^2$ | Noninvasive | LABEC | MS | Corumbá | 0 | 0.92 |
| TTPAN13 ² | Noninvasive | LABEC | MS | Corumbá | 0 | 0.89 |
| TTPR01 | Noninvasive | LABEC | PR | Curitiba | 6 | 0.66 |
| TTBA02 | Noninvasive | LABEC | BA | Santo Amaro | 0.5 | 0.81 |
| TTBA03 | Noninvasive | LABEC | BA | Camacari | 0.5 | 0.89 |
| TTBA04 | Noninvasive | LABEC | BA | Salvador | 0.5 | 0.83 |
| TTBA05 | Noninvasive | LABEC | BA | Camacari | 0.5 | 0.79 |
| TTBA06 | Noninvasive | LABEC | BA | Camacari | 0.5 | 0.81 |
| T18CE | Noninvasive | LABEC | CE | Maranguape | 0.8 | 0.75 |
| T19CE | Noninvasive | LABEC | CE | Maranguape | 0.8 | 0.78 |
| T20CE | Noninvasive | LABEC | CE | Maranguape | 0.8 | 0.81 |
| T21PI | Noninvasive | LABEC | PI | Cocal de Telha | 0.8 | 0.70 |
| TTMT1 | Noninvasive | LABEC | MT | Pontes e Lacerda | 0.1 | 0.93 |
| M0203 ³ | Fresh Tissue | LBEM | MG | Brasilândia | 5 | 1.0 |
| M0658 ³ | Fresh Tissue | LBEM | MG | Mariana | 7 | 1.0 |
| M0659 ³ | Fresh Tissue | LBEM | MG | Belo Horizonte | N/A | 1.0 |
| M0714 ³ | Fresh Tissue | LBEM | MG | Belo Horizonte | N/A | 1.0 |
| M0723 ³ | Fresh Tissue | LBEM | MG | Grão Mogol | 6 | 0.93 |
| M0720 ³ | Fresh Tissue | LBEM | MG | Grão Mogol | 6 | 1.0 |
| M0716 ³ | Fresh Tissue | LBEM | MG | Sto Antônio do Amparo | 9 | 1.0 |
| M0719 ³ | Fresh Tissue | LBEM | MG | Viçosa | 7 | 0.93 |
| M0661 ³ Fresh Tissue | | LBEM | MG | Abre Campo | 8 | 1.0 |
| M0666 ³ | M0666 ³ Fresh Tissue | | MG | Acaiaca | 8 | 1.0 |

| M0657 ³ | Fresh Tissue | LBEM | MG | Viçosa | 7 | 0.98 |
|---------------------|--------------|-------|----|-----------------------|-----|------|
| TTCA01 ³ | Fresh Tissue | LBEM | MG | Barão de Cocais | 3 | 1.0 |
| TTNA01 ³ | Fresh Tissue | LBEM | MG | Belo Horizonte | 3 | 1.0 |
| M0688 | Fresh Tissue | LBEM | SP | Bauru | N/A | 1.0 |
| M0689 | Fresh Tissue | LBEM | SP | Lorena | N/A | 0.98 |
| M0691 | Fresh Tissue | LBEM | SP | Bauru | 5 | 1.0 |
| M0692 | Fresh Tissue | LBEM | SP | Bauru | 6 | 1.0 |
| M0694 | Fresh Tissue | LBEM | SP | Bauru | 5 | 1.0 |
| M0695 | Fresh Tissue | LBEM | SP | Bauru | N/A | 1.0 |
| TMSJBV | Fresh Tissue | LABEC | SP | São João da Boa Vista | 4 | 1.0 |
| TM003 | Fresh Tissue | LABEC | SP | Sorocaba | 7 | 0.93 |
| TMIguape | Fresh Tissue | LABEC | SP | Iguape | 8 | 1.0 |
| PauloA | Fresh Tissue | LABEC | SP | Caraguatatuba | 15 | 1.0 |

Figure S1 Relationship between the age (years) of noninvasive samples and their QI values, evaluated through Pearson's correlation coefficient r (r= -0.7193593, p-value = 0.0007655), for the dataset without the outlier TTPR01 (Table S1, Supporting Information), which was an 6 years old sample.

0.9 * * * ∗ * * ≹ * * * * * * * * 0.6 * Т 0.5 0.0 0.5 1.0 1.5 2.0

Noninvasive w/o Outlier

Age (Years)

Figure S2 Plot of QI values of samples against age (years) depicted by the natural history collections of origin (symbols in legend). No linear relationship is found, and samples from the same institution and around the same age can vary in a wide range of QI values.

٦



Age (Years)

Table S1 - List of samples used in this study. Sample ID, biome of origin (assingment of individuals to biomes according to Figure 1, Cerrado – CE, Pantanal – PT, Atlantic Forest – AF, Caatinga – CA, Amazon Forest – AM), country, Brazilian federal state, locality and institution of origin where the voucher specimen, tissue or DNA is stored (Laboratório de Biologia Evolutiva e Conservação de Vertebrados – LABEC, Laboratório de Biodiversidade e Evolução Molecular – LBEM, Museu de Biologia Professor Mello Leitão – MBML, Universidade Federal de Santa Catarina – UFSC, Museu de História Natural do Capão da Imbuia – MNHCI, Museu Nacional – MN, Museu de Zoologia da Universidade federal de São Paulo – MZUSP, Universidade Federal de Rondônia – UNIR, Universidade Federal do Pará – UFPA, Laboratório de Biologia Genômica e Molecular – LBGM). Asterisks indicate historical samples.

| Sample ID | Biome | Country | Federal State | Locality/Municipality | Origin |
|-----------|-------|---------|---------------|-----------------------|--------|
| M0203 | CE | Brazil | MG | Brasilândia | LBEM |
| M0658 | CE | Brazil | MG | Mariana | LBEM |
| M0723 | CE | Brazil | MG | Grão Mogol | LBEM |
| M0720 | CE | Brazil | MG | Grão Mogol | LBEM |
| M0667 | CE | Brazil | GO | São Simão | LBEM |
| TTTA01 | CE | Brazil | MT | Nova Xavantina | LBEM |
| TTTA02 | CE | Brazil | MT | Nova Xavantina | LBEM |
| TTJE01 | CE | Brazil | GO | Catalão | LBEM |
| TTTA03 | CE | Brazil | MT | Barra do Garças | LBEM |
| TTTA04 | CE | Brazil | MT | Nova Xavantina | LBEM |
| TTTA05 | CE | Brazil | MT | Vale dos Sonhos | LBEM |
| TTTA06 | CE | Brazil | MT | Nova Xavantina | LBEM |
| TTPAN8 | CE | Brazil | MS | Paranaíba | LABEC |
| TTPAN9 | CE | Brazil | MS | Paranaíba | LABEC |
| TTPAN10 | CE | Brazil | MS | Paranaíba | LABEC |
| TTPAN11 | CE | Brazil | MS | Paranaíba | LABEC |
| TMUnimat | CE | Brazil | MT | Nova Xavantina | LABEC |
| Btte008 | CE | Brazil | MS | Anaurilândia | LBGM |

| Btte006 | CE | Brazil | MS | Bataguassu | LBGM |
|----------|----|--------|----|------------------------|-------|
| Btte002 | CE | Brazil | MA | Carolina | LBGM |
| TTBA01 | CE | Brazil | BA | Luís Eduardo Magalhães | LABEC |
| TM1823 | CE | Brazil | PI | Itaueira | LABEC |
| MN2328* | CE | Brazil | ТО | Palma | MN |
| MN2335* | CE | Brazil | ТО | Palma | MN |
| MN3846* | CE | Brazil | GO | Anápolis | MN |
| MN5061* | CE | Brazil | MS | Maracaju | MN |
| MN7620* | CE | Brazil | MG | Araguari | MN |
| MN24854* | CE | Brazil | GO | Mineiros | MN |
| UFCS838* | CE | Brazil | MS | Campo Grande | UFSC |
| TTPAN1 | РТ | Brazil | MS | Aquidauana | LABEC |
| TTPAN2 | РТ | Brazil | MS | Aquidauana | LABEC |
| TTPAN3 | РТ | Brazil | MS | Aquidauana | LABEC |
| TTPAN4 | РТ | Brazil | MS | Aquidauana | LABEC |
| TTPAN5 | РТ | Brazil | MS | Aquidauana | LABEC |
| TTPAN6 | РТ | Brazil | MS | Corumbá | LABEC |
| TTPAN7 | РТ | Brazil | MS | Aquidauana | LABEC |
| T2JUPAN | РТ | Brazil | MS | Aquidauana | LABEC |
| TTPAN12 | РТ | Brazil | MS | Corumbá | LABEC |
| TTPAN13 | РТ | Brazil | MS | Corumbá | LABEC |
| M0659 | AF | Brazil | MG | Belo Horizonte | LBEM |
| M0714 | AF | Brazil | MG | Belo Horizonte | LBEM |
| M0716 | AF | Brazil | MG | Sto Antônio do Amparo | LBEM |
| M0719 | AF | Brazil | MG | Viçosa | LBEM |
| M0661 | AF | Brazil | MG | Abre Campo | LBEM |
| M0666 | AF | Brazil | MG | Acaiaca | LBEM |
| M0657 | AF | Brazil | MG | Viçosa | LBEM |
| M0703 | AF | Brazil | BA | Una | LBEM |
| M0689 | AF | Brazil | SP | Lorena | LBEM |
| M0691 | AF | Brazil | SP | Bauru | LBEM |
| M0694 | AF | Brazil | SP | Bauru | LBEM |
| M0695 | AF | Brazil | SP | Bauru | LBEM |
| M0973 | AF | Brazil | PR | Telemaco Borba | LBEM |
| M0974 | AF | Brazil | PR | Telemaco Borba | LBEM |
| M0975 | AF | Brazil | PR | Telemaco Borba | LBEM |
| TTGI01 | AF | Brazil | MG | Resplendor | LBEM |
| TTCA01 | AF | Brazil | MG | Barão de Cocais | LBEM |
| TTNA01 | AF | Brazil | MG | Belo Horizonte | LBEM |
| MBML2453 | AF | Brazil | ES | Santa Teresa | MBML |
| TMSJBV | AF | Brazil | SP | São João da Boa Vista | LABEC |
| TM003 | AF | Brazil | SP | Sorocaba | LABEC |
| TRS01 | AF | Brazil | RS | Vacaria | LABEC |

| TT01 | AF | Brazil | PR | Antonina | LABEC |
|------------|----|--------|----|------------------------|-------|
| TMIguape | AF | Brazil | SP | Iguape | LABEC |
| PauloA | AF | Brazil | SP | Caraguatatuba | LABEC |
| TTBA03 | AF | Brazil | BA | Camacari | LABEC |
| TTBA04 | AF | Brazil | BA | Salvador | LABEC |
| TTBA05 | AF | Brazil | BA | Camacari | LABEC |
| TTBA06 | AF | Brazil | BA | Camacari | LABEC |
| T1AL | AF | Brazil | AL | Maceió | LABEC |
| T2AL | AF | Brazil | AL | Maceió | LABEC |
| T3AL | AF | Brazil | AL | Rio Largo | LABEC |
| T4AL | AF | Brazil | AL | Maceió | LABEC |
| T5PE | AF | Brazil | PE | Recife | LABEC |
| T6PE | AF | Brazil | PE | Recife | LABEC |
| T7PE | AF | Brazil | PE | Goiana | LABEC |
| T8PE | AF | Brazil | PE | Igarassu | LABEC |
| T9PE | AF | Brazil | PE | Recife | LABEC |
| T10PE | AF | Brazil | PE | Igarassu | LABEC |
| T11PB | AF | Brazil | PB | João Pessoa | LABEC |
| T12PB | AF | Brazil | PB | João Pessoa | LABEC |
| AF116PB | AF | Brazil | PB | João Pessoa | LABEC |
| Btte016 | AF | Brazil | SC | Itapoá | LBGM |
| Btte010 | AF | Brazil | RS | Vacaria | LBGM |
| Btte015 | AF | Brazil | PR | Antonina | LBGM |
| Btte023 | AF | Brazil | SC | Blumenau | LBGM |
| MBML2353* | AF | Brazil | ES | Viana | MBML |
| MBML2797* | AF | Brazil | ES | Vila Velha | MBML |
| MBMLSF2* | AF | Brazil | ES | Santa Teresa | MBML |
| MBMLSF3* | AF | Brazil | ES | Santa Teresa | MBML |
| UFSC3042* | AF | Brazil | SC | Ilha de Santa Catarina | UFSC |
| UFSC3948* | AF | Brazil | SC | Joinville | UFSC |
| UFSC658* | AF | Brazil | SC | Praia Grande | UFSC |
| MHNCI5989* | AF | Brazil | SC | Caxambu do Sul | MHNCI |
| MHNCI6048* | AF | Brazil | PR | Antonina | MHNCI |
| MHNCI5709* | AF | Brazil | PR | Icaraíma | MHNCI |
| MHNCI5780* | AF | Brazil | PR | Paranaguá | MHNCI |
| MHNCI5788* | AF | Brazil | PR | Paranaguá | MHNCI |
| MHNCI5082* | AF | Brazil | PR | Diamante do Norte | MHNCI |
| MHNCI5708* | AF | Brazil | PR | Diamante do Norte | MHNCI |
| MHNCI3967* | AF | Brazil | PR | Pinhão | MHNCI |
| MHNCI3971* | AF | Brazil | PR | Guarapuava | MHNCI |
| MHNCI3986* | AF | Brazil | SC | Blumenau | MHNCI |
| MHNCI3929* | AF | Brazil | PR | Arapoti | MHNCI |
| MHNCI3930* | AF | Brazil | PR | Reserva de Iguaçu | MHNCI |

| MHNCI3938* | AF | Brazil | SC | Água Doce | MHNCI |
|------------|----|---------------|----|----------------------|-------|
| MHNCI2501* | AF | Brazil | PR | São José dos Pinhais | MHNCI |
| MHNCI1695* | AF | Brazil | PR | Piraquara | MHNCI |
| MHNCI370* | AF | Brazil | PR | Curitiba | MHNCI |
| MHNCI6206* | AF | Brazil | PR | Diamante do Norte | MHNCI |
| MHNCI6207* | AF | Brazil | PR | Morretes | MHNCI |
| MHNCI6208* | AF | Brazil | PR | Paranaguá | MHNCI |
| MHNCI6213* | AF | Brazil | PR | Matinhos | MHNCI |
| MHNCI6192* | AF | Brazil | PR | Diamante do Norte | MHNCI |
| MHNCI6196* | AF | Brazil | PR | Diamante do Norte | MHNCI |
| MHNCI6199* | AF | Brazil | PR | Morretes | MHNCI |
| MHNCI6200* | AF | Brazil | PR | Alexandra-Paranaguá | MHNCI |
| MHNCI1696* | AF | Brazil | PR | Matinhos | MHNCI |
| MN5883* | AF | Brazil | ES | Santa Teresa | MHNCI |
| MN23954* | AF | Brazil | BA | Ilhéus | MHNCI |
| MN2326* | AF | Brazil | RJ | Teresópolis | MHNCI |
| MN24834* | AF | Brazil | RJ | Campos | MHNCI |
| TTBA02 | CA | Brazil | BA | Santo Amaro | LABEC |
| T13CE | CA | Brazil | CE | Fortaleza | LABEC |
| T14CE | CA | Brazil | CE | Fortaleza | LABEC |
| T15CE | CA | Brazil | CE | Fortaleza | LABEC |
| T16CE | CA | Brazil | CE | Ibiapaba | LABEC |
| T17CE | CA | Brazil | CE | Sobral | LABEC |
| T18CE | CA | Brazil | CE | Maranguape | LABEC |
| T19CE | CA | Brazil | CE | Maranguape | LABEC |
| T20CE | CA | Brazil | CE | Maranguape | LABEC |
| T21PI | CA | Brazil | PI | Cocal de Telha | LABEC |
| TM35 | CA | Brazil | PI | Juazeiro do Piauí | LABEC |
| TTAMZ01 | AM | Brazil | PA | Marabá | LABEC |
| TMPar | AM | Brazil | AM | Parintins | LABEC |
| TMAM | AM | Brazil | PA | Oriximiná | LABEC |
| 13H47 | AM | French Guyana | - | Kourou | LABEC |
| TOC015 | AM | Brazil | MA | Imperatriz | LABEC |
| TOC065 | AM | Brazil | MA | Imperatriz | LABEC |
| TOC001 | AM | Brazil | MA | Imperatriz | LABEC |
| UFROM302 | AM | Brazil | RO | Porto Velho | UNIR |
| UFROM303 | AM | Brazil | RO | Porto Velho | UNIR |
| TTMT1 | AM | Brazil | MT | Pontes e Lacerda | LABEC |
| TTMT2 | AM | Brazil | MT | Pontes e Lacerda | LABEC |
| TTMT3 | AM | Brazil | MT | Pontes e Lacerda | LABEC |
| TMPeru1 | AM | Peru | - | - | LABEC |
| TMPeru2 | AM | Peru | - | - | LABEC |
| TMPeru3 | AM | Peru | - | - | LABEC |

| TMN142 | AM | Brazil | AM | Manaus | UFPA |
|-------------|----|--------|----|---------------|-------|
| New002 | AM | Brazil | PA | Belém | UFPA |
| TMN141 | AM | Brazil | AM | Manaus | UFPA |
| TMN93 | AM | Brazil | AM | Manaus | UFPA |
| J103 | AM | Brazil | PA | Juruti | UFPA |
| New001 | AM | Brazil | PA | Barbacena | UFPA |
| JE23 | AM | Brazil | PA | Itaituba | UFPA |
| TTRO01 | AM | Brazil | RO | Costa Marques | LABEC |
| MN2341* | AM | Brazil | AM | Manacapuru | MN |
| MN2342* | AM | Brazil | AM | Manacapuru | MN |
| MN5050* | AM | Brazil | PA | Curralinho | MN |
| MN5052* | AM | Brazil | PA | Curralinho | MN |
| MN5053* | AM | Brazil | PA | Curralinho | MN |
| MN5057* | AM | Brazil | PA | Curralinho | MN |
| MN5058* | AM | Brazil | PA | Curralinho | MN |
| MN5060* | AM | Brazil | PA | Curralinho | MN |
| MN5063* | AM | Brazil | PA | Curralinho | MN |
| MN5065* | AM | Brazil | PA | Curralinho | MN |
| MN5067* | AM | Brazil | PA | Curralinho | MN |
| MN5070* | AM | Brazil | PA | Curralinho | MN |
| MN5071* | AM | Brazil | PA | Curralinho | MN |
| MN5072* | AM | Brazil | PA | Curralinho | MN |
| MN5726* | AM | Brazil | PA | Belterra | MN |
| MN5743* | AM | Brazil | PA | Belterra | MN |
| MN5967* | AM | Brazil | AM | Juruá | MN |
| MZUSP13680* | AM | Brazil | RR | Catrimani | MZUSP |
| MZUSP13681* | AM | Brazil | RR | Catrimani | MZUSP |
| | | | | | |

Table S2 – Parameters of genetic diversity (N – number of individuals sampled in the population that were amplified successfully for each locus; A – Allelic richness estimated based in a minimum of three individuals; H_0/H_E – observed and expected heterozygosity and F_{IS} – population inbreeding coefficient) within 20 populations of *a priori* analysis for microsatellite loci. Both fresh and historical samples were considered. Individuals belonging to each population are listed in Appendix S1. Asterisks indicate significant departure from Hardy-Weinberg equilibrium at 0.05 confidence level after correction for multiple tests.

| Рор | | | PA1 | | PER1 | | | | | AM1 | | | | AM2 | | | |
|-------|----|------------|------------|-----------------|------|------|-----------|-----------------|---|------|------------|-----------------|---|------|------------|-----------------|--|
| Locus | Ν | Α | H_0/H_E | F _{IS} | Ν | Α | H_0/H_E | F _{IS} | Ν | Α | H_0/H_E | F _{IS} | Ν | Α | H_0/H_E | F _{IS} | |
| H5 | 14 | 1.07 | 0.07/0.07 | 0.00 | 3 | 1.00 | - | - | 5 | 1.00 | - | - | 6 | 1.16 | 0.16/0.16 | 0.00 | |
| E12 | 6 | 1.59 | 0.17/0.59* | 0.73 | 3 | 1.60 | 0.66/0.60 | -0.14 | 4 | 1.85 | 0.85/0.85 | 0.00 | 6 | 1.80 | 0.66/0.80 | 0.18 | |
| G3 | 11 | 1.90 | 0.36/0.90* | 0.60 | 3 | 1.80 | 0.66/0.80 | 0.20 | 5 | 1.88 | 0.40/0.88* | 0.57 | 5 | 1.93 | 0.80/0.93 | 0.15 | |
| F1R | 13 | 1.34 | 0.23/0.34 | 0.33 | 3 | 1.00 | - | - | 5 | 1.00 | - | - | 6 | 1.00 | - | - | |
| B2 | 13 | 1.87 | 0.76/0.87 | 0.12 | 3 | 1.73 | 0.66/0.73 | 0.11 | 5 | 1.88 | 0.80/0.88 | 0.11 | 4 | 1.96 | 1.00/0.96 | -0.04 | |
| C10 | 13 | 1.47 | 0.38/0.47 | 0.18 | 3 | 1.60 | 0.66/0.60 | -0.14 | 5 | 1.53 | 0.53/0.53 | 0.00 | 6 | 1.53 | 0.16/0.53 | 0.70 | |
| A9 | 14 | 1.27 | 0.21/0.26 | 0.21 | 3 | 1.73 | 0.66/0.73 | 0.11 | 4 | 1.75 | 0.25/0.50 | 0.70 | 6 | 1.78 | 0.33/0.78* | 0.60 | |
| A8 | 12 | 1.50 | 0.33/0.50 | 0.35 | 3 | 1.86 | 0.66/0.86 | 0.27 | 4 | 1.82 | 0.75/0.82 | 0.10 | 5 | 1.82 | 0.60/0.82 | 0.29 | |
| Рор | N | 1T1 | | | | | MT2 | | | PI | CE1 | | | N | IA1 | | |
| Locus | Ν | Α | H_0/H_E | F _{IS} | Ν | Α | H_0/H_E | F _{IS} | Ν | Α | H_0/H_E | F _{IS} | Ν | А | H_0/H_E | F _{IS} | |
| H5 | 7 | 1.00 | - | - | 3 | 1.00 | - | - | 3 | 1.00 | - | - | 3 | 1.00 | - | - | |
| E12 | 7 | 1.79 | 0.28/0.79* | 0.65 | 3 | 1.73 | 0.33/0.73 | 0.60 | 3 | 1.53 | 0.53/0.53 | 0.00 | 3 | 1.60 | 0.33/0.60 | 0.50 | |
| G3 | 7 | 1.82 | 0.85/0.82 | 0.67 | 3 | 1.93 | 0.67/0.93 | 0.33 | 3 | 1.60 | 0.66/0.60 | -0.14 | 3 | 1.80 | 0.66/0.80 | 0.20 | |
| F1R | 7 | 1.58 | 0.28/0.58 | 0.52 | 3 | 1.00 | - | - | 3 | 1.00 | - | - | 3 | 1.00 | - | - | |
| B2 | 6 | 1.87 | 0.50/0.87* | 0.45 | 3 | 1.80 | 1.00/0.80 | -0.33 | 3 | 1.86 | 0.33/0.86 | 0.66 | 3 | 1.83 | 0.67/0.93 | 0.33 | |
| C10 | 7 | 1.26 | 0.26/0.26 | 0.00 | 3 | 1.53 | 0.77/0.73 | -0.33 | 3 | 1.73 | 0.33/0.73 | 0.60 | 3 | 1.73 | 1.00/0.73 | -0.50 | |
| A9 | 7 | 1.47 | 0.28/0.47 | 0.41 | 3 | 1.73 | 0.33/0.73 | 0.60 | 3 | 1.80 | 0.80/0.80 | 0.00 | 3 | 1.00 | - | - | |

| A8 | 6 | 1.43 | 0.16/0.43 | 0.64 | 3 | 1.73 | 0.66/0.73 | 0.11 | 2 | 2.00 | 1.00/1.00 | 0.00 | 3 | 1.86 | 0.67/0.86 | 0.27 |
|-------|----|------|------------|-----------------|---|------|-----------|-----------------|---|------|------------|-----------------|---|------|-----------|-----------------|
| Рор | | | CE1 | | | | PB1 | | | | PE1 | | | | AL1 | |
| Locus | Ν | A | H_0/H_E | F _{IS} | Ν | А | H_0/H_E | F _{IS} | Ν | Α | H_0/H_E | F _{IS} | Ν | А | H_0/H_E | F _{IS} |
| Н5 | 6 | 1.00 | - | - | 3 | 1.00 | - | - | 6 | 1.00 | - | - | 4 | 1.03 | 0.16/0.16 | 0.00 |
| E12 | 6 | 1.75 | 0.66/0.75 | 0.13 | 3 | 1.73 | 0.67/0.73 | 0.11 | 6 | 1.53 | 0.17/0.53 | 0.70 | 4 | 1.78 | 0.50/0.78 | 0.40 |
| G3 | 6 | 1.69 | 0.50/0.69 | 0.30 | 3 | 2.00 | 1.00/1.00 | 0.00 | 6 | 1.74 | 0.67/0.74 | 0.11 | 4 | 1.82 | 0.25/0.82 | 0.72 |
| F1R | 6 | 1.16 | 0.17/0.17 | 0.00 | 3 | 1.33 | 0.33/0.33 | 0.00 | 6 | 1.00 | - | - | 4 | 1.25 | 0.25/0.25 | 0.00 |
| B2 | 6 | 1.93 | 1.00/0.93 | -0.08 | 3 | 1.86 | 1.00/0.87 | -0.20 | 6 | 1.71 | 0.50/0.71 | 0.31 | 4 | 1.89 | 0.75/0.89 | 0.18 |
| C10 | 6 | 1.48 | 0.66/0.48 | -0.42 | 3 | 1.53 | 0.53/0.53 | 0.00 | 6 | 1.71 | 0.50/0.71 | 0.31 | 4 | 1.82 | 0.25/0.82 | 0.72 |
| A9 | 6 | 1.63 | 0.33/0.63 | 0.50 | 3 | 1.53 | 0.53/0.53 | 0.00 | 6 | 1.16 | 0.16/0.16 | 0.00 | 4 | 1.53 | 0.25/0.53 | 0.57 |
| A8 | 6 | 1.31 | 0.17/0.31 | 0.50 | 3 | 1.73 | 0.33/0.73 | 0.60 | 6 | 1.43 | 0.50/0.43 | -0.15 | 4 | 1.82 | 0.50/0.82 | 0.42 |
| Рор | | | BA1 | | | | ES1 | | | | MG1 | | | | SP1 | |
| Locus | Ν | А | H_0/H_E | F _{IS} | Ν | Α | H_0/H_E | F _{IS} | Ν | Α | H_0/H_E | F _{IS} | Ν | Α | H_0/H_E | F _{IS} |
| Н5 | 3 | 1.03 | 0.16/0.16 | 0.00 | 5 | 1.00 | - | - | 7 | 1.14 | 0.14/0.14 | 0.00 | 3 | 1.00 | - | - |
| E12 | 3 | 1.80 | 0.80/0.80 | 0.00 | 4 | 1.71 | 0.50/0.71 | 0.33 | 7 | 1.69 | 0.71/0.69 | -0.03 | 3 | 1.80 | 0.67/0.80 | 0.20 |
| G3 | 3 | 1.00 | - | - | 5 | 1.73 | 0.60/0.73 | 0.20 | 7 | 1.84 | 0.85/0.84 | -0.01 | 3 | 1.60 | 0.33/0.60 | 0.50 |
| F1R | 3 | 1.00 | - | - | 2 | 1.50 | 0.50/0.50 | 0.00 | 7 | 1.36 | 0.42/0.36 | -0.20 | 3 | 1.33 | 0.33/0.33 | 0.00 |
| B2 | 3 | 1.80 | 0.67/0.80 | 0.20 | 2 | 1.66 | 0.66/0.66 | 0.00 | 6 | 1.68 | 0.67/0.68 | 0.02 | 3 | 1.80 | 0.67/0.80 | 0.20 |
| C10 | 3 | 1.86 | 0.67/0.86 | 0.27 | 5 | 1.00 | - | - | 7 | 1.00 | - | - | 3 | 1.33 | 0.33/0.33 | 0.00 |
| A9 | 3 | 1.00 | - | - | 3 | 1.53 | 0.53/0.53 | 0.00 | 7 | 1.26 | 0.28/0.26 | -0.09 | 3 | 1.00 | - | - |
| A8 | 3 | 1.53 | 0.67/0.53 | -0.33 | 3 | 2.00 | 1.00/1.00 | 0.00 | 7 | 1.65 | 0.29/0.65 | 0.58 | 3 | 1.60 | 0.33/0.60 | 0.50 |
| Рор | | | PRSC1 | | | | SCMS1 | | | | MS1 | | | | MS2 | |
| Locus | Ν | А | H_0/H_E | F _{IS} | Ν | Α | H_0/H_E | F _{IS} | Ν | А | H_0/H_E | F _{IS} | Ν | Α | H_0/H_E | F _{IS} |
| Н5 | 15 | 1.06 | 0.06/0.06 | 0.00 | 6 | 1.00 | - | - | 8 | 1.00 | - | - | 6 | 1.00 | - | - |
| E12 | 15 | 1.70 | 0.20/0.70* | 0.72 | 6 | 1.83 | 0.66/0.83 | 0.21 | 8 | 1.80 | 0.50/0.81* | 0.39 | 5 | 1.73 | 0.20/0.73 | 0.75 |
| G3 | 13 | 1.86 | 0.38/0.86* | 0.56 | 5 | 1.77 | 0.40/0.77 | 0.51 | 8 | 1.80 | 0.25/0.80* | 0.70 | 3 | 1.73 | 0.33/0.73 | 0.60 |
| F1R | 13 | 1.34 | 0.23/0.34 | 0.33 | 6 | 1.16 | 0.16/0.16 | -0.00 | 8 | 1.12 | 0.12/0.12 | 0.00 | 5 | 1.46 | 0.20/0.46 | 0.60 |
| B2 | 12 | 1.81 | 0.75/0.81 | 0.08 | 6 | 1.93 | 1.00/0.93 | -0.07 | 7 | 1.90 | 0.57/0.90* | 0.38 | 4 | 1.85 | 0.75/0.85 | 0.14 |

| C10 | 15 | 1.51 | 0.27/0.51 | 0.48 | 6 | 1.16 | 0.16/0.16 | 0.00 | 8 | 1.00 | - | - | 6 | 1.43 | 0.16/0.43 | 0.64 |
|-----|----|------|-----------|-------|---|------|-----------|------|---|------|-----------|------|---|------|-----------|------|
| A9 | 13 | 1.45 | 0.30/0.45 | 0.33 | 6 | 1.30 | 0.30/0.30 | 0.00 | 8 | 1.43 | 0.25/0.43 | 0.44 | 6 | 1.65 | 0.50/0.65 | 0.25 |
| A8 | 11 | 1.45 | 0.54/0.45 | -0.22 | 4 | 1.25 | 0.25/0.25 | 0.00 | 7 | 1.44 | 0.28/0.43 | 0.36 | 4 | 1.60 | 0.25/0.60 | 0.62 |

Table S3 – Parameters of genetic diversity (N – number of individuals sampled in the population that were amplified successfully for each locus; A – Allelic richness estimated based in a minimum of three individuals; H_0/H_E – observed and expected heterozygosity and F_{IS} – population inbreeding coefficient) within 16 populations of *a priori* analysis for microsatellite loci (populations left with less than three individuals after removing historical samples were excluded). Only fresh samples were considered in this analysis. Individuals belonging to each population are listed in Appendix S1. Asterisks indicate significant departure from Hardy-Weinberg equilibrium at 0.05 confidence level after correction for multiple tests.

| Рор | BA1 | | | | MC | G1 | L | | | PER1 | | | | AM2 | | | |
|-------|-----|------|------------|-----------------|----|-----------|-----------|-----------------|-----|------|------------|-----------------|----|------------|------------|-----------------|--|
| Locus | Ν | А | H_O/H_E | F _{IS} | Ν | А | H_O/H_E | F _{IS} | Ν | А | H_O/H_E | F _{IS} | Ν | А | H_O/H_E | F _{IS} | |
| H5 | 3 | 1.00 | - | - | 7 | 1.20 | 0.14/0.14 | 0.00 | 3 | 1.43 | 0.16/0.16 | 0.00 | 4 | 1.00 | - | - | |
| E12 | 3 | 1.80 | 0.80/0.80 | 1.00 | 7 | 1.81 | 0.71/0.69 | 0.40 | 3 | 1.70 | 0.66/0.60 | -0.03 | 4 | 1.80 | 0.75/0.82 | 0.66 | |
| G3 | 3 | 1.80 | 0.85/0.85 | 0.00 | 7 | 1.81 | 0.85/0.84 | 0.70 | 3 | 1.85 | 0.66/0.80 | -0.01 | 3 | 1.82 | 1.00/0.86 | 0.67 | |
| F1R | 3 | 1.00 | - | - | 7 | 1.13 | 0.42/0.36 | 0.30 | 3 | 1.36 | 0.66/0.60 | -0.20 | 4 | 1.59 | 0.45/0.49 | 0.53 | |
| B2 | 3 | 1.83 | 0.66/0.80 | 0.50 | 6 | 1.90 | 0.66/0.68 | 0.39 | 3 | 1.68 | 0.66/0.73 | 0.02 | 3 | 1.88 | 1.00/1.00 | 0.46 | |
| C10 | 3 | 1.53 | 0.66/0.86 | 1.00 | 7 | 1.00 | - | - | 3 | 1.03 | 0.66/0.66 | 0.00 | 4 | 1.27 | 0.25/0.67 | 1.00 | |
| A9 | 3 | 2.00 | 0.29/0.29 | 0.00 | 7 | 1.43 | 0.28/0.26 | 0.40 | 3 | 1.27 | 0.66/0.73 | -0.09 | 4 | 1.47 | 0.25/0.82* | 0.42 | |
| A8 | 3 | 2.00 | 0.67/0.53 | 0.35 | 7 | 1.44 | 0.28/0.65 | 0.37 | 3 | 1.66 | 0.66/0.86 | 0.58 | 3 | 1.44 | 1.00/0.73 | 0.64 | |
| Рор | МЛ | 71 | | | МЛ | [2 | | | PIC | CE1 | | | MA | A 1 | | | |
| Locus | Ν | А | H_O/H_E | F _{IS} | Ν | А | H_O/H_E | F _{IS} | Ν | А | H_O/H_E | F _{IS} | Ν | А | H_O/H_E | F _{IS} | |
| H5 | 7 | 1.10 | 0.28/0.28 | 0.00 | 3 | 1.00 | - | - | 3 | 1.00 | - | - | 3 | 1.20 | 0.33/0.33 | 0.00 | |
| E12 | 7 | 1.73 | 0.29/0.79* | 0.75 | 3 | 1.73 | 0.33/0.73 | 0.60 | 3 | 1.80 | 0.53/0.63 | 1.00 | 3 | 1.79 | 0.33/0.60 | 0.40 | |
| G3 | 7 | 1.73 | 0.29/0.82* | 0.60 | 3 | 1.93 | 0.66/0.93 | 0.33 | 3 | 1.20 | 0.66/0.66 | 0.00 | 3 | 1.82 | 0.66/0.80 | 0.72 | |
| F1R | 7 | 1.47 | 0.29/0.58* | 0.60 | 3 | 1.00 | - | - | 3 | 1.00 | - | - | 3 | 1.25 | 0.29/0.29 | 0.00 | |
| B2 | 6 | 1.86 | 0.50/0.87* | 0.14 | 3 | 1.80 | 1.00/0.80 | -0.33 | 3 | 1.80 | 0.33/0.86* | 0.20 | 3 | 1.89 | 0.66/0.93 | 0.18 | |

| C10 | 7 | 1.44 | 0.29/0.26 | 0.64 | 3 | 1.53 | 0.66/0.53 | -0.33 | 3 | 1.86 | 0.33/0.73 | 0.27 | 3 | 1.82 | 1.00/0.73 | 0.73 |
|-------|-----|------|-----------|-----------------|----|------|-----------|-----------------|----|------|------------|-----------------|----|------|------------|-----------------|
| A9 | 7 | 1.65 | 0.28/0.47 | 0.25 | 3 | 1.73 | 0.33/0.73 | 0.60 | 3 | 1.00 | - | - | 3 | 1.54 | 0.29/0.33 | 0.57 |
| A8 | 6 | 1.60 | 0.17/0.44 | 0.63 | 3 | 1.73 | 0.66/0.73 | 0.11 | 2 | 1.53 | 1.00/1.00 | 0.33 | 3 | 1.82 | 0.66/0.86 | 0.43 |
| Pop | CE | 1 | | | PB | 1 | | | PE | 1 | | | AL | 1 | | |
| Locus | Ν | А | H_O/H_E | F _{IS} | Ν | А | H_O/H_E | F _{IS} | Ν | А | H_O/H_E | F _{IS} | Ν | А | H_O/H_E | F _{IS} |
| H5 | 6 | 1.00 | - | - | 3 | 1.00 | - | - | 6 | 1.00 | - | - | 4 | 1.10 | 0.16/0.16 | 0.00 |
| E12 | 6 | 1.53 | 0.66/0.75 | 0.71 | 3 | 1.73 | 0.66/0.73 | 0.11 | 6 | 1.76 | 0.16/0.53 | 0.13 | 4 | 1.60 | 0.50/0.78 | 0.50 |
| G3 | 6 | 1.74 | 0.50/0.69 | 0.11 | 3 | 2.00 | 1.00/1.00 | 0.00 | 6 | 1.70 | 0.66/0.74 | 0.30 | 4 | 1.80 | 0.25/0.82* | 0.20 |
| F1R | 6 | 1.10 | 0.16/0.16 | 0.00 | 3 | 1.33 | 0.33/0.33 | 0.00 | 6 | 1.17 | 0.33/0.33 | 0.00 | 4 | 1.10 | 0.25/0.25 | 0.00 |
| B2 | 5 | 1.71 | 1.00/0.93 | 0.32 | 3 | 1.87 | 1.00/0.86 | -0.20 | 6 | 1.93 | 0.50/0.71 | -0.08 | 4 | 1.93 | 0.75/0.89 | 0.33 |
| C10 | 6 | 1.71 | 0.66/0.48 | 0.32 | 3 | 1.53 | 0.40/0.53 | 1.00 | 6 | 1.49 | 0.70/0.50 | -0.43 | 4 | 1.73 | 0.25/0.82 | -0.50 |
| A9 | 6 | 1.17 | 0.63/0.63 | 0.00 | 3 | 1.53 | 0.40/0.53 | 1.00 | 6 | 1.64 | 0.16/0.16 | 0.50 | 4 | 1.10 | 0.25/0.53 | 0.30 |
| A8 | 6 | 1.44 | 0.16/0.32 | -0.15 | 3 | 1.73 | 0.33/0.73 | 0.60 | 6 | 1.32 | 0.50/0.43 | 0.50 | 4 | 1.87 | 0.50/0.82* | 0.27 |
| Pop | SPI | 1 | | | PR | SC1 | | | MS | 51 | | | MS | 52 | | |
| Locus | N | А | H_O/H_E | F _{IS} | Ν | А | H_O/H_E | F _{IS} | Ν | А | H_O/H_E | F _{IS} | Ν | А | H_O/H_E | F _{IS} |
| H5 | 3 | 1.00 | - | - | 3 | 1.00 | - | - | 8 | 1.00 | - | - | 6 | 1.10 | 0.20/0.20 | 0.00 |
| E12 | 3 | 1.82 | 0.66/0.80 | 0.10 | 3 | 1.53 | 0.80/0.80 | 0.00 | 8 | 1.60 | 0.50/0.80* | -0.14 | 5 | 1.80 | 0.20/0.73* | 0.20 |
| G3 | 3 | 1.87 | 0.33/0.60 | -0.20 | 3 | 1.60 | 0.80/0.80 | 0.00 | 8 | 1.80 | 0.25/0.80 | 0.20 | 3 | 1.60 | 0.33/0.73 | 0.50 |
| F1R | 3 | 1.10 | 0.33/0.33 | 0.00 | 3 | 1.00 | - | - | 8 | 1.10 | 0.12/0.12 | 0.00 | 5 | 1.33 | 0.20/0.46 | 0.20 |
| B2 | 3 | 2.00 | 0.66/0.80 | 0.20 | 2 | 1.86 | 0.50/0.83 | 0.66 | 7 | 1.73 | 0.57/0.90* | 0.11 | 4 | 1.80 | 0.75/0.85 | 0.20 |
| C10 | 3 | 1.68 | 0.33/0.33 | 0.66 | 3 | 1.73 | 0.50/0.53 | 0.60 | 8 | 1.60 | 0.50/0.54 | 0.14 | 6 | 1.33 | 0.16/0.43 | 0.20 |
| A9 | 3 | 1.82 | 0.43/0.45 | 0.72 | 1 | 1.80 | 1.00/1.00 | 0.00 | 8 | 1.73 | 0.25/0.43 | 0.11 | 6 | 1.10 | 0.50/0.65 | 0.10 |
| A8 | 3 | 1.73 | 0.33/0.60 | -0.50 | 1 | 2.00 | 1.00/1.00 | 0.00 | 7 | 1.87 | 0.28/0.43 | 0.27 | 4 | 1.60 | 0.25/0.60 | 0.50 |

Table S4 – Values of F_{ST} between pairs of populations in *a priori* analysis. Asterisks refer to significance at a confidence level, with p-values lower than 0.05 (*), 0.01 (**) and 0.001 (***), after correction for multiple tests.

| Рор | MT2 | MT1 | MS1 | MS2 | SCMS 1 | SP1 | PRSC1 | MG1 | ES1 | BA1 | AL1 | PE1 | PB1 | CE1 | PICE1 | MA1 | PA1 | AM2 | AM1 | PER1 |
|-------|-----------|----------|-----------|---------|----------|--------|-----------|-----------|----------|----------|--------|----------|--------|---------|----------|--------|----------|--------|--------|--------|
| MT2 | 0.0000 | | | | | | | | | | | | | | | | | | | |
| MT1 | 0.2370** | 0.0000 | | | | | | | | | | | | | | | | | | |
| MS1 | 0.3454** | 0.0591 | 0.0000 | | | | | | | | | | | | | | | | | |
| MS2 | 0.2438 | 0.0041 | 0.0041 | 0.0000 | | | | | | | | | | | | | | | | |
| SCMS1 | 0.3275* | 0.0446 | 0.0446 | 0.0472 | 0.0000 | | | | | | | | | | | | | | | |
| SP1 | 0.3129 | 0.1409 | 0.1409 | 0.0677 | 0.0384 | 0.0000 | | | | | | | | | | | | | | |
| PRSC1 | 0.2243** | 0.0212* | 0.0707 | 0.0697 | 0.0611 | 0.0321 | 0.0000 | | | | | | | | | | | | | |
| MG1 | 0.3505* | 0.0234 | 0.0348 | 0.0174 | 0.0710* | 0.0516 | 0.0682 | 0.0000 | | | | | | | | | | | | |
| ES1 | 0.3714* | 0.0118 | 0.1137* | 0.0073 | 0.0916 | 0.0266 | 0.0628 | 0.0523 | 0.0000 | | | | | | | | | | | |
| BA1 | 0.3213 | 0.1794 | 03114** | 0.2741* | 0.2756* | 0.3475 | 0.2451** | 0.3256** | 0.4491* | 0.0000 | | | | | | | | | | |
| AL1 | 0.1088 | 0.1020 | 0.1428* | 0.0338 | 0.1512 | 0.1086 | 0.0805 | 0.1026 | 0.1172 | 0.1913 | 0.0000 | | | | | | | | | |
| PE1 | 0.2780* | 0.1721** | 0.1964** | 0.0776 | 0.1899** | 0.1984 | 0.2131*** | 0.2269*** | 0.2697** | 0.2340* | 0.0944 | 0.0000 | | | | | | | | |
| PB1 | 0.1401 | 0.0125 | 0.0540 | 0.0439 | 0.0163 | 0.0449 | 0.0325 | 0.0489 | 0.0895 | 0.1692 | 0.0228 | 0.0917 | 0.0000 | | | | | | | |
| CE1 | 0.1833* | 0.0204 | 0.0850 | 0.0202 | 0.0746 | 0.0652 | 0.0736 | 0.0393 | 0.1065* | 0.2567** | 0.0312 | 0.1056* | 0.0418 | 0.0000 | | | | | | |
| PICE1 | 0.1384 | 0.2605* | 0.2465* | 0.1776 | 0.3266* | 0.3641 | 0.2344** | 0.3387** | 0.3128* | 0.3606 | 0.1184 | 0.2751** | 0.2030 | 0.2406* | 0.0000 | | | | | |
| MA1 | 0.2137 | 0.0452 | 0.1186 | 0.0543 | 0.1177 | 0.0780 | 0.0967 | 0.0854* | 0.2681* | 0.2227 | 0.0327 | 0.0532 | 0.0705 | 0.0135 | 0.2739 | 0.0000 | | | | |
| PA1 | 0.3011*** | 0.0118 | 0.0326 | 0.0194 | 0.0386 | 0.0293 | 0.0157 | 0.0135 | 0.0646 | 0.2361** | 0.0172 | 0.1085* | 0.0427 | 0.0568 | 0.2619** | 0.0031 | 0.0000 | | | |
| AM2 | 0.1530 | 0.0499 | 0.0645 | 0.0217 | 0.0767 | 0.0886 | 0.0265 | 0.0668 | 0.0413 | 0.2309* | 0.0265 | 0.1896** | 0.0383 | 0.0577 | 0.1006 | 0.0828 | 0.0064 | 0.0000 | | |
| AM1 | 0.0835 | 0.1075 | 0.2277*** | 0.1397 | 0.1961 | 0.2098 | 0.1005 | 0.2221*** | 0.2154 | 0.3438* | 0.0488 | 0.2757** | 0.0898 | 0.1156 | 0.1113 | 0.2042 | 0.1778** | 0.0405 | 0.0000 | |
| PER1 | 0.1052 | 0.1964* | 0.2890** | 0.2111* | 0.2751* | 0.2145 | 0.1497** | 0.2448** | 0.2365* | 0.3230 | 0.0452 | 0.2881* | 0.1344 | 0.1573* | 0.1547 | 0.2247 | 0.1621* | 0.0791 | 0.0379 | 0.0000 |

Appendix S1 – Supporting Information

List of *Tamandua tetradactyla* individuals belonging to each *a priori* population, corresponding to populations in Table S2. Asterisks indicate populations excluded from analysis with fresh samples only (Table S3).

| A priori populations | Individuals |
|----------------------|-------------|
| Pará 1 (PA1)* | MN5050 |
| (n=14) | MN5052 |
| | MN5053 |
| | MN5057 |
| | MN5058 |
| | MN5060 |
| | MN5061 |
| | MN5063 |
| | MN5065 |
| | MN5067 |
| | MN5070 |
| | MN5071 |
| | MN5072 |
| | New002 |
| Peru 1 (PER1) | TMPeru |
| (n=3) | TMPeru2 |
| | TMPeru3 |
| Amazônia 1 (AM1)* | MN2341 |
| (n=5) | MN2342 |
| | TMN142 |
| | TMN141 |
| | TMN93 |
| Amazônia 2 (AM2) | MN5726 |
| (n=6) | MN5743 |
| | TMPar |
| | TMAM |
| | J103 |
| | JE23 |
| Mato Grosso 1 (MT1) | TTTA01 |
| (n=7) | TTTA02 |
| | TTTA03 |
| | TTTA04 |
| | TTTA05 |

| | TTTA06 |
|----------------------------|----------|
| | TMUnimat |
| Mato Grosso 2 (MT2) | TTMT1 |
| (n=3) | TTMT2 |
| | TTMT3 |
| Piauí and Ceará 1 (PICE 1) | T16CE |
| (n=3) | T21PI |
| | TM35 |
| Maranhão 1 (MA1) | TOC015 |
| (n=3) | TOC065 |
| | TOC001 |
| Ceará 1 (CE1) | T13CE |
| (n=6) | T14CE |
| | T15CE |
| | T18CE |
| | T19CE |
| | T20CE |
| Paraíba 1 (PB1) | T11PB |
| (n=3) | T12PB |
| | AF116PB |
| Pernambuco 1 (PE1) | T5PE |
| (n=6) | T6PE |
| | T7PE |
| | T8PE |
| | T9PE |
| | T10PE |
| Alagoas 1 (AL1) | T1AL |
| (n=4) | T2AL |
| | T3AL |
| | T4AL |
| Bahia 1 (BA1) | TTBA03 |
| (n=3) | TTBA05 |
| | TTBA06 |
| Espírito Santo 1 (ES1)* | MN5883 |
| (n=5) | MBML2353 |
| | MBML2797 |
| | MBMLSF2 |
| | MBMLSF3 |
| Minas Gerais 1 (MG1) | M0659 |
| (n=7) | M0714 |
| | TTNA01 |
| | TTCA01 |

| | M0719 |
|--------------------------------|-----------|
| | M0661 |
| | M0666 |
| São Paulo 1 (SP1) | M0691 |
| (n=3) | M0694 |
| | M0695 |
| Paraná and Santa Catarina 1 | |
| (PRSC1) | Btte016 |
| (n=15) | Btte015 |
| | Btte023 |
| | MHNCI2501 |
| | MHNCI1695 |
| | MHNCI370 |
| | MHNCI5780 |
| | MHNCI5788 |
| | UFSC3948 |
| | MHNCI1696 |
| | MHNCI3986 |
| | MHNCI6208 |
| | MHNCI6213 |
| | MHNCI6199 |
| | MHNCI6200 |
| Santa Catarina and Mato Grosso | |
| do Sul 1 (SCMS1)* | MHNCI6192 |
| (n=6) | MHNCI6196 |
| | MHNCI5709 |
| | MHNCI5082 |
| | MHNCI5708 |
| | MHNCI6206 |
| Mato Grosso do Sul 1 (MS1) | TTPAN1 |
| (n=8) | TTPAN2 |
| | TTPAN3 |
| | TTPAN4 |
| | TTPAN5 |
| | TTPAN6 |
| | TTPAN7 |
| | T2JUPAN |
| Mato Grosso do Sul 2 (MS2) | TTPAN8 |
| (n=6) | TTPAN9 |
| | TTPAN10 |
| | TTPAN11 |
| | TTPAN12 |
| | TTPAN13 |
| | |

Material Suplementar Capítulo 3

Appendix S1 – Supporting Information

Details about SSCP gel preparation

Individual amplicons were screened by SSCP analysis (Orita *et al.* 1989), a method that can detect variants separated by only a single base difference (Sunnucks *et al.* 2000). For denaturation, 2–4 μ l of PCR products were mixed with 6 μ l of loading dye (30% glycerol, 0.25% bromophenol blue, 0.25% xylene cyanol), heated at 95°C for 5' and placed on ice for 5'. This mix was loaded on 15% non-denaturing polyacrylamide gels (ETC, Germany) and run on a horizontal cooling electrophoresis system (Amersham Pharmacia, Germany) with the following conditions: 200 V, 10 mA, 10 W for 20' followed by 450 V, 30 mA, 20 W for 3:45' and 4:30' at a constant temperature of 10 °C. Gels were fixed and silver-stained using PlusOne DNA Silver Staining Kit (Amersham Pharmacia) following the manufacturer's recommendations. Distinctive single-strand bands were excised from the gel, eluted in 30 μ l TBE buffer and incubated for at least 3 hours. Subsequent reamplifications followed as described in the Material and Methods section.

References

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci U S A.* **86**, 2766–2770.

Sunnucks P, Wilson ACC, Beheregaray LB, Zenger K, French J, Taylor AC (2000) SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. *Mol Ecol* **9**, 1699–1710.

168

Table S1 - List of *Tamandua tetradactyla* samples (n=71) used in this study. Sample ID, biome of origin (according to IBGE (<u>http://www.ibge.gov.br/home</u>): Atlantic Forest - AF, Amazon Forest - AM, Caatinga - CA, Cerrado – CE and Pantanal - PT), country, Brazilian federal state abbreviation, locality/municipality, and reference institution where the voucher specimen, tissue or DNA is stored (Laboratório de Biologia Evolutiva e Conservação de Vertebrados, São Paulo – LABEC; Laboratório de Bioliversidade e Evolução Molecular, Belo Horizonte – LBEM; Museu de Biologia Professor Mello Leitão, Santa Teresa – MBML; Universidade Federal de Rondônia, Boa Vista – UNIR; Laboratório de Biologia Genômica e Molecular, Porto Alegre – LBGM).

| Sample ID | Biome | Country | Federal State ¹ | Locality/Municipality | Reference institution |
|-----------------|-------|---------|-------------------------------|-----------------------|-----------------------|
| M0659 | AF | Brazil | MG | Belo Horizonte | LBEM |
| M0714 | AF | Brazil | MG | Belo Horizonte | LBEM |
| M0716 | AF | Brazil | MG | Sto Antônio do Amparo | LBEM |
| M0661 | AF | Brazil | MG | Abre Campo | LBEM |
| M0666 | AF | Brazil | MG | Acaiaca | LBEM |
| M0703 | AF | Brazil | BA | Una | LBEM |
| M0691 | AF | Brazil | SP | Bauru | LBEM |
| M0694 | AF | Brazil | SP | Bauru | LBEM |
| M0973 | AF | Brazil | PR | Telemaco Borba | LBEM |
| M0974 | AF | Brazil | PR | Telemaco Borba | LBEM |
| M0975 | AF | Brazil | PR | Telemaco Borba | LBEM |
| TTGI01 | AF | Brazil | MG | Resplendor | LBEM |
| TTCA01 | AF | Brazil | MG | Barão de Cocais | LBEM |
| TTNA01 | AF | Brazil | MG | Belo Horizonte | LBEM |
| MBML2453 | AF | Brazil | ES | Santa Teresa | MBML |
| TMSJBV | AF | Brazil | SP | São João da Boa Vista | LABEC |
| TRS01 | AF | Brazil | RS | Vacaria | LABEC |
| TMIguape | AF | Brazil | SP | Iguape | LABEC |
| TMPauloA | AF | Brazil | SP | Caraguatatuba | LABEC |
| T1AL | AF | Brazil | AL | Maceió | LABEC |
| T2AL | AF | Brazil | AL | Maceió | LABEC |
| T3AL | AF | Brazil | AL | Rio Largo | LABEC |
| T4AL | AF | Brazil | AL | Maceió | LABEC |
| T5PE | AF | Brazil | PE | Recife | LABEC |

| T6PE | AF | Brazil | Brazil PE | | LABEC |
|----------|----|---------------|---------------|------------------------|-------|
| T7PE | AF | Brazil | PE | Goiana | LABEC |
| T8PE | AF | Brazil | PE | Igarassu | LABEC |
| T10PE | AF | Brazil | PE Igarassu | | LABEC |
| T11PB | AF | Brazil | PB | João Pessoa | LABEC |
| T12PB | AF | Brazil | PB | João Pessoa | LABEC |
| Btte010 | AF | Brazil | RS | Vacaria | LBGM |
| TTAMZ01 | AM | Brazil | PA | Marabá | LABEC |
| TMPar | AM | Brazil | AM | Parintins | LABEC |
| TMAM | AM | Brazil | PA | Oriximiná | LABEC |
| UFROM302 | AM | Brazil | RO | Porto Velho | UNIR |
| UFROM303 | AM | Brazil | RO | Porto Velho | UNIR |
| TTMT1* | AM | Brazil | MT | Pontes e Lacerda | LABEC |
| TTMT2* | AM | Brazil | MT | Pontes e Lacerda | LABEC |
| TTMT3 | AM | Brazil | MT | Pontes e Lacerda | LABEC |
| TMPeru1 | AM | Peru | | | LABEC |
| TMPeru2 | AM | Peru | | | LABEC |
| TMPeru3* | AM | Peru | | | LABEC |
| 13H47 | AM | French Guyana | | Kourou | LABEC |
| M0658 | CE | Brazil | MG | Mariana | LBEM |
| M0667 | CE | Brazil | GO | São Simão | LBEM |
| TTTA01 | CE | Brazil | MT | Nova Xavantina | LBEM |
| TTTA02 | CE | Brazil | MT | Nova Xavantina | LBEM |
| TTJE01 | CE | Brazil | GO | Catalão | LBEM |
| TTPAN8 | CE | Brazil | MS | Paranaíba | LABEC |
| TTPAN9 | CE | Brazil | MS | Paranaíba | LABEC |
| TMUnimat | CE | Brazil | MT | Nova Xavantina | LABEC |
| Btte008 | CE | Brazil | MS | Anaurilândia | LBGM |
| Btte006 | CE | Brazil | MS | Bataguaçu | LBGM |
| M0688 | CE | Brazil | SP | Lorena | LBEM |
| TTBA01 | CE | Brazil | BA | Luís Eduardo Magalhães | LABEC |
| M0692 | CE | Brazil | SP | Lorena | LBEM |
| TOC015 | AM | Brazil | MA | Imperatriz | LABEC |
| TOC065 | AM | Brazil | MA | Imperatriz | LABEC |
| TOC001 | AM | Brazil | MA | Imperatriz | LABEC |
| T13CE | CA | Brazil | CE | Fortaleza | LABEC |
| T16CE | CA | Brazil | CE | Ipiapaba | LABEC |
| T17CE | CA | Brazil | CE | Sobral | LABEC |
| TTPAN1 | PT | Brazil | MS Aquidauana | | LABEC |
| TTPAN2 | РТ | Brazil | MS Aquidauana | | LABEC |
| TTPAN3 | PT | Brazil | MS | Aquidauana | LABEC |
| TTPAN4 | PT | Brazil | MS Aquidauana | | LABEC |
| TTPAN5 | РТ | Brazil | MS | Aquidauana | LABEC |

| TTPAN7 | PT | Brazil | MS | Aquidauana | LABEC |
|----------|----|--------|----|------------|-------|
| TTPAN11* | PT | Brazil | MS | Aquidauana | LABEC |
| TTPAN12* | PT | Brazil | MS | Aquidauana | LABEC |
| TTPAN13* | PT | Brazil | MS | Aquidauana | LABEC |

¹Names of Brazilian federal states, in order of appearance: Minas Gerais (MG), Bahia (BA), São Paulo (SP), Paraná (PR), Espírito Santo (ES), Rio Grande do Sul (RS), Alagoas (AL), Pernambuco (PE), Paraíba (PB), Pará (PA), Amazonas (AM), Rondônia (RO), Goiás (GO), Mato Grosso (MT), Mato Grosso do Sul (MS), Ceará (CE).

*Samples that did not yield enough coverage through pyrosequencing, and were excluded from further analyses.

 Table S2 – Fusion primer name and composition (adaptor Lib-A sequence, internal library key, barcode for individual identification

(multiplex identifiers, MIDs) and specific primer sequence (forward: JF1eV, reverse: YML10) used for 454 pyrosequencing.

| Fusion Primer Name | Adaptor | Key | Barcode (MID) | Specific Primer Sequence |
|------------------------|-----------------------|------|---------------|-----------------------------|
| Forward | | | | |
| SS_MHC_JF1eV_LibA_A001 | CGTATCGCCTCCCTCGCGCCA | TCAG | ACGAGTGCGT | GAGTGTCATTTYGAGAACGGGACSGAG |
| SS_MHC_JF1eV_LibA_A002 | CGTATCGCCTCCCTCGCGCCA | TCAG | AGACGCACTC | GAGTGTCATTTYGAGAACGGGACSGAG |
| SS_MHC_JF1eV_LibA_A003 | CGTATCGCCTCCCTCGCGCCA | TCAG | CGTGTCTCTA | GAGTGTCATTTYGAGAACGGGACSGAG |
| SS_MHC_JF1eV_LibA_A004 | CGTATCGCCTCCCTCGCGCCA | TCAG | CTCGCGTGTC | GAGTGTCATTTYGAGAACGGGACSGAG |
| SS_MHC_JF1eV_LibA_A005 | CGTATCGCCTCCCTCGCGCCA | TCAG | TACAGATCGT | GAGTGTCATTTYGAGAACGGGACSGAG |
| SS_MHC_JF1eV_LibA_A006 | CGTATCGCCTCCCTCGCGCCA | TCAG | TGATACGTCT | GAGTGTCATTTYGAGAACGGGACSGAG |
| SS_MHC_JF1eV_LibA_A007 | CGTATCGCCTCCCTCGCGCCA | TCAG | CATACTCTAC | GAGTGTCATTTYGAGAACGGGACSGAG |
| SS_MHC_JF1eV_LibA_A008 | CGTATCGCCTCCCTCGCGCCA | TCAG | CGAGAGATAC | GAGTGTCATTTYGAGAACGGGACSGAG |
| SS_MHC_JF1eV_LibA_A009 | CGTATCGCCTCCCTCGCGCCA | TCAG | ATACGACGTA | GAGTGTCATTTYGAGAACGGGACSGAG |
| SS_MHC_JF1eV_LibA_A010 | CGTATCGCCTCCCTCGCGCCA | TCAG | TCTGTCTCGC | GAGTGTCATTTYGAGAACGGGACSGAG |
| SS_MHC_JF1eV_LibA_A011 | CGTATCGCCTCCCTCGCGCCA | TCAG | CGACACTATC | GAGTGTCATTTYGAGAACGGGACSGAG |
| Reverse | | | | |
| SS_MHC_YML10_LibA_B001 | CTATGCGCCTTGCCAGCCCGC | TCAG | TCTCTATGCG | TCGCCGCTGCACTGTGAACGTCTC |
| SS_MHC_YML10_LibA_B002 | CTATGCGCCTTGCCAGCCCGC | TCAG | TCACGCGAGA | TCGCCGCTGCACTGTGAACGTCTC |
| SS_MHC_YML10_LibA_B003 | CTATGCGCCTTGCCAGCCCGC | TCAG | TAGAGACGAG | TCGCCGCTGCACTGTGAACGTCTC |
| SS_MHC_YML10_LibA_B004 | CTATGCGCCTTGCCAGCCCGC | TCAG | TCGTCGCTCG | TCGCCGCTGCACTGTGAACGTCTC |
| SS_MHC_YML10_LibA_B005 | CTATGCGCCTTGCCAGCCCGC | TCAG | ATAGATAGAC | TCGCCGCTGCACTGTGAACGTCTC |
| SS_MHC_YML10_LibA_B006 | CTATGCGCCTTGCCAGCCCGC | TCAG | TAGTCGCATA | TCGCCGCTGCACTGTGAACGTCTC |
| SS_MHC_YML10_LibA_B007 | CTATGCGCCTTGCCAGCCCGC | TCAG | TGTACTACTC | TCGCCGCTGCACTGTGAACGTCTC |
| SS_MHC_YML10_LibA_B008 | CTATGCGCCTTGCCAGCCCGC | TCAG | ACGACTACAG | TCGCCGCTGCACTGTGAACGTCTC |
| SS_MHC_YML10_LibA_B009 | CTATGCGCCTTGCCAGCCCGC | TCAG | CGTAGACTAG | TCGCCGCTGCACTGTGAACGTCTC |
| SS MHC YML10 LibA B010 | CTATGCGCCTTGCCAGCCCGC | TCAG | CTATAGCGTA | TCGCCGCTGCACTGTGAACGTCTC |

Table S3 – Comparison of putative MHC alleles obtained for ten individuals by using the SSCP gel or the 454 pyrosequencing approach. N_1 and N_2 are the total number of putative alleles found in each individual using SSCP and NGS, respectively. All MHC alleles detected by SSCP were confirmed by NGS. Additional MHC alleles detected by NGS only are marked in bold.

| Samples | N_1 | N_2 | | | Putative Alleles | | |
|----------|-------|-------|-------------|-------------|------------------|-------------|-------------|
| TMIguape | 4 | 4 | TateDRB*04 | TateDRB*40 | TateDRB*39d | TateDRB*17 | |
| 13H47 | 8 | 9 | TateDRB*04 | TateDRB*44c | TateDRB*44a | TateDRB*39b | TateDRB*38 |
| | | | TateDRB*01a | TateDRB*43 | TateDRB*11 | TateDRB*29 | |
| MBML2453 | 4 | 6 | TateDRB*40 | TateDRB*39a | TateDRB*01a | TateDRB*44c | TateDRB*01a |
| | | | TateDRB*25 | | | | |
| TMPauloA | 4 | 6 | TateDRB*37b | TateDRB*36 | TateDRB*30 | TateDRB*39b | TateDRB*25 |
| | | | TateDRB*40 | | | | |
| T2AL | 3 | 4 | TateDRB*04 | TateDRB*05 | TateDRB*29 | TateDRB*40 | |
| T7PE | 4 | 5 | TateDRB*04 | TateDRB*05 | TateDRB*40 | TateDRB*29 | TateDRB*17 |
| T17CE | 4 | 4 | TateDRB*40 | TateDRB*04 | TateDRB*42 | TateDRB*53 | |
| TMAM | 5 | 6 | TateDRB*04 | TateDRB*01a | TateDRB*44a | TateDRB*31 | TateDRB*57 |
| | | | TateDRB*56 | | | | |
| TTPAN1 | 4 | 8 | TateDRB*40 | TateDRB*01b | TateDRB*39c | TateDRB*39d | TateDRB*03 |
| | | | TateDRB*25 | TateDRB*44a | TateDRB*49 | | |
| TTPAN3 | 5 | 7 | TateDRB*40 | TateDRB*28 | TateDRB*39a | TateDRB*01a | TateDRB*45 |
| | | | TateDRB*25 | TateDRB*39d | | | |

Figure S1 – Sharing of MHC class II DRB exon 2 alleles among individuals within and between different biomes. Biomes are denoted by outer circle color: Atlantic Forest (blue, n=29), Amazon Forest (green, n=9), Cerrado (grey, n=16), Caatinga (violet, n=3) and Pantanal (red, n=8). Each ribbon represents one DRB amino acid allele (N=60). Each interruption in the outer circle represent a different individual (N=65).



Figure S2 – Network of MHC class II DRB exon2 alleles (nucleotide level) with nodes proportional to frequency of individuals carrying the allele, colored by the biomes in which it occurs: Atlantic Forest (blue, n=29), Amazon Forest (green, n=9), Cerrado (grey, n=16), Caatinga (violet, n=3) and Pantanal (red, n=8). Interruptions in lines represent the presence of more than ten mutations.

