

UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE MEDICINA VETERINÁRIA E ZOOTECNIA

PESQUISA DE GENOMAS VIRAIS EM PRIMATAS NÃO
HUMANOS DO NOVO MUNDO POR METAGENÔMICA

LEILA SABRINA ULLMANN

Botucatu-SP

Agosto/2014

UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE MEDICINA VETERINÁRIA E ZOOTECNIA

PESQUISA DE GENOMAS VIRAIS EM PRIMATAS NÃO
HUMANOS DO NOVO MUNDO POR METAGENÔMICA

LEILA SABRINA ULLMANN

Tese apresentada junto ao Programa
de Pós-Graduação em Medicina
Veterinária para obtenção do título de
Doutor.

Orientador: Prof. Adj. João Pessoa
Araújo Junior

Co-Orientador: Prof. Adj. Alexander W.
Biondo

Botucatu/SP

Agosto/2014

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. E TRAT. DA INFORMAÇÃO

DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: **ROSEMEIRE APARECIDA VICENTE**

Ullmann, Leila Sabrina.

Pesquisa de genomas virais em primatas não humanos do Novo Mundo por metagenômica/ Leila Sabrina Ullmann. - Botucatu, 2014

Tese (doutorado) – Faculdade de Medicina Veterinária e Zootecnia, Universidade Estadual Paulista, 2014

Orientador: João Pessoa Araújo Junior

Capes: 50502034

1. Primatas não humanos - Doenças. 2. Metagenômica viral. 3. Diagnóstico viral.

Palavras-chave: Primatas não humanos; Metagenômica viral; Diagnóstico viral.

Leila Sabrina Ullmann. Pesquisa de genomas virais em primatas não humanos do Novo Mundo por metagenômica. Defesa: 22/08/2014. Local: FMVZ/UNESP – Campus de Botucatu/SP.

COMISSÃO EXAMINADORA

Prof. Adj. João Pessoa Araújo Junior

Presidente e Orientador

Departamento de Microbiologia e Imunologia

Instituto de Biociências - UNESP - Botucatu/SP

Profa. Dra. Maria Inês de Moura Campos Pardini

Membro Titular

Hemocentro de Botucatu

Faculdade de Medicina de Botucatu - UNESP - Botucatu/SP

Prof. Ass. Dr. Deilson Elgui de Oliveira

Membro Titular

Departamento de Patologia

Faculdade de Medicina de Botucatu - UNESP - Botucatu/SP

Prof. Dr. Jean Carlos Ramos da Silva

Membro Titular

Departamento de Veterinária Preventiva

Universidade Federal Rural de Pernambuco - UFRPE

Prof. Dr. Walfrido Kuhl Svoboda

Membro Titular

Universidade Federal da Integração Latino Americana

Prof. Adj. Cesar Martins

Membro Suplente

Departamento de Morfologia

Instituto de Biociências - UNESP - Botucatu/SP

Prof. José Paes de Oliveira Filho

Membro Suplente

Departamento Clínica Veterinária

Faculdade de Medicina Veterinária e Zootecnia, Universidade Estadual Paulista

Profa. Ass. Dr. Márcia Guimarães da Silva

Membro Suplente

Departamento de Patologia

Faculdade de Medicina de Botucatu - UNESP - Botucatu/SP

DEDICATÓRIA

A Deus, pela bênção de eu poder cumprir mais uma etapa da minha formação, que faz parte dos meus planos para o futuro...

Dedico esta minha conquista ao meu pai, minha mãe e minha irmã pelo apoio incondicional e por entenderem que minha ausência apenas representa distância espacial, mas que estamos sempre juntos...

Dedico ainda, em especial, o cumprimento de mais esta etapa, a três pessoas que foram muito importantes na minha vida, mas que infelizmente não estão mais entre nós, três lindas flores que já se foram: meu querido primo Fábio, Vovô Álvaro e Vovó Calista Ullmann.

Usando as palavras do meu avô para expressar a gratidão, o amor e a saudade que sinto: “Para Deus, fazemos parte de um imenso jardim... E a cada dia, Ele colhe as flores mais bonitas para estarem junto Dele...” Vocês foram as três flores mais lindas que estiveram perto de mim neste jardim...

AGRADECIMENTOS

Tenho muito a agradecer as muitas pessoas que, de alguma forma me ajudaram nesta caminhada... Espero lembrar de cada um enquanto escrevo estes agradecimentos...

"I get by with a little help of my friends."

John Lennon and Paul McCartney

Posso dizer que eu faria uma alteração nesta frase: "I get by with a lot help of my friends." Sempre presentes, mesmo longe, os meus amigos são o alicerce de tudo o que eu consegui até hoje! Sempre prontos a me ajudar, sempre ali, mesmo que longe, estão sempre perto, moram no meu coração!

À minha família postiça, Seu Antônio, Dona Margarida, minha querida irmã Marianna e Claudinei, cunhado postiço, muito obrigada pelo apoio, incentivo e amizade de sempre... Ah, e Mari e Caixa, muito obrigada por me darem um sobrinho neste mês! Augusto, venha cheio de saúde e alegria!!!

Aos meus queridos amigos do Predinho, da época da faculdade, muito obrigada pela amizade, confiança, apoio e principalmente por vocês terem feito a diferença na minha vida a partir da graduação. Vocês são essenciais! Algumas pessoas entram na nossa vida e temos a certeza de que elas serão companheiras sempre, independente do momento de nossas vidas... Essas pessoas:, Gislaíne, Eduardo e Théo, muito obrigada mesmo!

Muito obrigada aos meus familiares por acreditarem em mim e entenderem que a minha ausência em reuniões familiares, as não visitas quando meus pais ainda moravam em Palotina e,

principalmente agora, que eles não moram mais no Paraná, são privações que preciso para completar esta etapa da minha vida. Tenham certeza que vocês fazem e sempre farão parte da minha vida, mesmo longe... A saudade é uma companheira aqui em São Paulo.

Ao professor, amigo e orientador de vida Dr. Alexander W. Biondo, muito obrigada pela paciência, apoio, amizade, dedicação, profissionalismo e por aguentar as minhas teimosias desde 2004! Saiba que você é um dos grandes exemplos da minha vida como um todo e um dos principais responsáveis pela profissional que sou. Quem diria que há quase quatro anos, estávamos na fila da Liga do Chopp discutindo sobre possível orientador de doutorado e hoje estou aqui, terminando esta etapa! Tenho certeza que fizemos a melhor escolha possível!

Muito obrigada à amiga Lucilene “Dulce”, minha querida R2 da residência. Encontrei uma frase para descrever como você faz a diferença: “Você faz do mundo um lugar melhor só por existir.” Aceite esta frase como sua porque é exatamente assim que me sinto em relação a você. Companheira de todas as horas, uma pessoa que enxerga somente o bem nas outras, que está sempre pronta a ajudar os outros e que serve de exemplo de pós-graduanda. Sinto falta de trabalhar com você no dia-a-dia...

As minhas colegas de casa Selene, Karen, Luciana e Pâmela, assim como a todos os estagiários que já passaram pela casa, muito obrigada por fazerem parte do Lar do Briôcko Seco, uma república com cara de casa de família!

Aos meus queridos colegas de pós-graduação da FMVZ e da área de Saúde Animal, Saúde Pública Veterinária e Segurança Alimentar (SASPVSA), muito obrigada pela convivência destes anos.

Aos colegas de laboratório, Cláudia Filoni, Cláudia Tozatto, Marianna, Taís, Raissa, Sueli e Flávio (que não estão mais no lab), Michelly, Vinícius, Duroc, Didier, Arianí e Vivian.

Aos funcionários do Departamento de Microbiologia e Imunologia muito obrigada pela ajuda e pelos momentos compartilhados no departamento! A Dona Anita, que trabalha para manter nosso departamento limpo para podermos nos dedicar as nossas pesquisas, saiba que a senhora é essencial! Lulinha, muito obrigada pelos mimos nossos de sextas à tarde! E pela alegria de sempre!! Ana, querida secretária, sempre pronta a nos ajudar, muito obrigada!

A todos os professores da Micro-Imuno, muito obrigada pelos ensinamentos em disciplinas e também no dia-a-dia... Em especial aos professores Maurício Sforcin, Sílvio Luís Oliveira e Josias, que foram meus orientadores no estágio docência.

Muito obrigada aos funcionários do Centro Nacional de Primatas (Belém-PA), em especial ao Dr. Muniz (égua!), do Parque Ecológico do Tietê (São Paulo-SP), em especial ao Mestre Ramiro Dias Neto, e do Refúgio Biológico Bela Vista (Foz do Iguaçu-PR), em especial ao Dr. Wanderlei de Moraes e ao Dr. Zalmir S. Cubas, pelo apoio, acolhida e profissionalismo durante a coleta das amostras.

Algumas pessoas além de companheiros de casa ou de trabalho se tornam verdadeiros amigos... Todos na mesma situação, longe da família, com muito trabalho e estresse na área acadêmica... Formamos uma família! Aos amigos Selene, Pâmela, João Marcelo ("brother"), Larissa (Lóris), Ramiro, Ricardo ("Curan"), Ramiro e Jacqueline Kurissio, muito obrigada! Mirela, muito obrigada pela amizade, companhia para os treinos de boxe,

pela suuuuper ajuda com os detalhes da minha partida para Davis e por estar sempre presente!

Ao Nicholas Días (“Bipolar”), pela alegria contagiante de sempre... Pelas risadas e cervejas... E também pela coragem ao enfrentar o câncer tão novinho... Obrigada pela amizade e pelo apoio!

A uma grande amiga e colega de laboratório, Camila Malossi, muito obrigada por fazer parte da minha vida! Você foi uma das surpresas boas do laboratório quando voltei de Davis! Sempre alegre, trabalhando sério, disposta a ajudar a todos, melhorando o sucesso do sequenciamento massivo, você é uma das pessoas que admiro e tenho orgulho de poder conviver diariamente. Tenho certeza que logo logo você alçará vôos mais altos, mas por enquanto temos você no LDM!

A minha querida amiga Mary Marcondes e aos seus filhos fofíssimos Lucas e Gabriel, muito obrigada por me albergarem na casa de vocês! Mary, muito obrigada pela amizade, pela TV, pela carona até Botucatu na volta de Davis e por entender exatamente o que eu estava sentindo...

A minha querida amiga Marina, muito obrigada por fazer parte da minha vida, aguentar minhas tagarelíces e histórias. Você é muito especial e saiba que tenho muito orgulho da pessoa que você é! Ao meu querido amigo Diego Nóbrega (“Toxa”), muito obrigada pela amizade construída neste último ano! Não imaginei que teria um amigo como você que eu pudesse chamar de irmão. Ah! E nunca imaginei que teria um amigo tão nerd como você... Hehehehe.

Aliás, Marina e Toxinha, muito obrigada por acreditarem na pessoinha aqui e tentarem... Parabéns pelo casamento de vocês, obrigada por deixarem a minha vida mais bela pela simples existência do amor de vocês e boa sorte no Canadá!!!

Muito obrigada Bruno Fantinatti e Erica Ramos pela ajuda com as análises das minhas sequências!!! Obrigada pela paciência e pela sempre disponibilidade de me ajudar! Estou aqui caso precisem de ajuda, só avisar!

Muito obrigada Ludmila Soares, por me fazer suar a camisa no treino funcional!!! Saiba que foi muuuuuuito bom encontrar uma personal trainer num churrasco, tomando cerveja... Soube que você entenderia a minha vidinha e me deixaria tomar umas de vez enquanto... Mesmo que isso atrapalhasse o resultado final..

The Davis part...

Tenho muito a agradecer a uma família que me recebeu de braços abertos: Ana Cláudia, Jairo, Catarina (kk) e Ana Sofia!!! Morei quase dois meses com vocês quando cheguei em Davis, vocês me ajudaram muito com tudo! A minha primeira lembrança é assistir desenho com as meninas... Quantas risadas, quantos momentos legais ao lado de vocês... Quanta saudade eu sinto de poder apenas cruzar o condomínio e encontrar as minhas pequeninas... Que já estão enormes! Não vejo a hora de reencontrá-las, já que estão de volta à terrinha! Ana, seja bem-vinda ao Brasil! Muita sorte, mais coragem e saiba que estou aqui para o que vocês precisarem!!! Amo vocês!

Daniel, meu melhor amigo de Davis!!! Uma amizade que começou com "burgers and brew", que teve muito companheirismo, gordices, trabalhos em cima da hora, cafés e vinhos, viagens com

direito a estrela cadente e tudo... Muito obrigada por fazer da minha vida em Davis mais alegre e dançante, por aguentar minhas rabugices e chatices... Por me ajudar com a saudade da família e dos amigos que ficaram no Brasil! Amo você! Saiba que estou aqui pra o que você precisar!!!

*Constituí uma família em Davis! Tenho quatro filhos!!! Daniel, Carol, Andrew e Filipe!!! Muito obrigada por participarem da minha vida, por fazerem dos nossos dias uma aventura! "Não quero que este dia acabe"... Vocês sabem como me sinto com vocês: "Safe and Sound". Adoro poder ter compartilhado do meu tempo em Davis com vocês!!! Espero revê-los logo!!! *Mom.*

Dri, Liza, Mariana e Carlos: muito obrigada pela convivência na Califórnia... Um período muito especial para todos nós!!! Isa e Chris (Isa, traduza o agradecimento pro Chris), como foi legal poder ter conhecido vocês!!! Lógico que nos tornaríamos amigos, nos conhecemos num dia de Oktoberfest! Muito obrigada pela amizade, pelas hoooooras de conversa e pelo mliehor Thanksgiving que eu poderia ter tido!

I'm really grateful that I had you Traci as my roomie... I loved to spend my year in Davis living with you and Riley! You have no idea how much I miss you guys... From the beginning when I was so slow on talking and finding the right words... Until the last day when we were talking right before I leave... You started crying and I was pretending I was so "strong" and not crying... You know how much you are important and how much I wish you happiness!!! I'm really glad that you're happy with your honey Adam!!!

Andrew and Steph, thank you so much for your friendship!!! I really loved to meet you guys and I'm very happy that you're a couple now!!! I hope we meet soon...

Dr. Christian Leutenegger, thank you so much for your attention, support and our talks. I learned a lot with you and your family, especially Iris! You know you have a special family!!!

An especial thank you for Dr. Pusterla for accepting me as his student and receiving me on your lab! Thanks for the opportunity to work with you and also with Samantha!!! I simply love you, Samantha!! It was a pleasure to help you at the lab, to share a lot of laughs and chocolate with you... Cara, thank you so much for your help! I hope we can meet soon and share some "happiness in a bottle"!!!

Dr. Delwart, thank you so much for receiving me on your lab! Thank you so much for your attention and for spending some time with me. I really appreciate the few days I spent on your lab and work with Linlin and Lark. I already told you that, but the first paper on viral metagenomics that I read, you were the author! So you can imagine how fascinating was to meet you personally.

Uma das partes mais difíceis para mim nesta tese foi escrever os agradecimentos... São muitas pessoas que me ajudaram e que infelizmente não poderei agradecer a todas especificamente aqui. Saibam que vocês são muito importantes. Muito obrigada!

Agradeço a FAPESP pela concessão da bolsa de doutorado, a CAPES pela bolsa de doutorado sanduíche, indispensável para minha manutenção em Botucatu e em Davis, e execução do projeto.

Bom, tenho de acrescentar alguns agradecimentos na versão final da minha tese... Muito obrigada as minhas duas cupidas Fabiana Ferreira de Souza e Luciana Ortega Xavier por me apresentarem ao agora, oficialmente namorado, João Netto. A data será sempre lembrada por todos os outros brasileiros como o

“dia que o Brasil perdeu de 7x1 para a Alemanha, na Copa do Mundo do Brasil”, mas para mim, este dia representa muito mais! Fui apresentada a uma pessoa que ocupa um espaço muito especial na minha vida. João Netto, espero que o que sentimos um pelo outro aumente a cada dia, como isso vem acontecendo. Saiba que você faz parte dos meus planos, é uma das minhas prioridades e que mesmo que o tal pós-doutorado em Cornell dê certo, ficaremos longe, mas perto... Vamos trilhar nosso caminho! Com respeito e diálogo alcançaremos nossos objetivos juntos!

Por último, mas não por isso menos especial, ao meu orientador Prof. Dr. João Pessoa Araújo Júnior, muito obrigada! Muito obrigada pela orientação, pela confiança, pelas cobranças, pela oportunidade de colocar a mão-na-massa do sequenciamento massivo e pelas pizzas! Obrigada por me deixar conviver com a sua família, por frequentar a sua casa em churrascos e pizzas.

Dizem que não podemos elogiar porque estraga... Mas é final de doutorado, então acho que posso... E não vai estragar! Prof. João Pessoa, eu só tenho a agradecer a oportunidade que você me deu de poder realizar o meu sonho de morar nos EUA. Muito obrigada pela orientação, por ir pra bancada mesmo quando eu não sabia como fazer determinada técnica, pela paciência com a garota enxaqueca aqui e por mostrar todos os dias que o Brasil tem sim pessoas competentes e que com parcerias conseguimos realizar projetos bem audaciosos! Espero poder contribuir para o seu crescimento profissional como você contribuí com o meu! Muito obrigada mesmo!!!

Sinceramente.

"Depois de algum tempo, você aprende a diferença, a sutil diferença, entre dar a mão e acorrentar uma alma. E você aprende que amar não significa apoiar-se, e que companhia nem sempre significa segurança. E começa a aprender que beijos não são contratos e presentes não são promessas. E começa a aceitar suas derrotas com a cabeça erguida e olhos adiante, com a graça de um adulto e não com a tristeza de uma criança.

E aprende a construir todas as suas estradas no hoje, porque o terreno do amanhã é incerto demais para os planos, e o futuro tem o costume de cair em meio ao vão. Depois de um tempo você aprende que o sol queima se ficar exposto por muito tempo. E aprende que não importa o quanto você se importe, algumas pessoas simplesmente não se importam... E aceita que não importa quão boa seja uma pessoa, ela vai feri-lo de vez em quando e você precisa perdoá-la, por isso. Aprende que falar pode aliviar dores emocionais.

Descobre que se levam anos para se construir confiança e apenas segundos para destruí-la, e que você pode fazer coisas em um instante das quais se arrependerá pelo resto da vida. Aprende que verdadeiras amizades continuam a crescer mesmo a longas distâncias. E o que importa não é o que você tem na vida, mas quem você tem na vida. E que bons amigos são a família que nos permitiram escolher. Aprende que não temos que mudar de amigos se compreendemos que os amigos mudam, percebe que seu melhor amigo e você podem fazer qualquer coisa, ou nada, e terem bons momentos juntos.

Descobre que as pessoas com quem você mais se importa na vida são tomadas de você muito depressa, por isso sempre devemos deixar as pessoas que amamos com palavras amorosas, pode ser a última vez que as vejamos. Aprende que as circunstâncias e os ambientes tem influência sobre nós, mas nós somos responsáveis por nós mesmos. Começa a aprender que não se deve comparar com os outros, mas com o melhor que pode ser.

Descobre que se leva muito tempo para se tornar a pessoa que quer ser, e que o tempo é curto. Aprende que não importa onde já chegou, mas onde está indo, mas se você não sabe para onde está indo, qualquer lugar serve. Aprende que, ou você controla seus atos ou eles o controlarão, e que ser flexível não significa ser fraco ou não ter personalidade, pois não importa quão delicada e frágil seja uma situação, sempre existem dois lados.

Aprende que heróis são pessoas que fizeram o que era necessário fazer, enfrentando as consequências. Aprende que paciência requer muita prática. Descobre que algumas vezes a pessoa que você espera que o chute quando você cai é uma das poucas que o ajudam a levantar-se.

Aprende que maturidade tem mais a ver com os tipos de experiência que se teve e o que você aprendeu com elas do que com quantos aniversários você celebrou. Aprende que há mais dos seus pais em você do que você supunha. Aprende que nunca se deve dizer a uma criança que sonhos são bobagens, poucas coisas são tão humilhantes e seria uma tragédia se ela acreditasse nisso.

Aprende que quando está com raiva tem o direito de estar com raiva, mas isso não te dá o direito de ser cruel. Descobre que só porque alguém não o ama do jeito que você quer que ame, não significa que esse alguém não o ama, contudo o que pode, pois existem pessoas que nos amam, mas simplesmente não sabem como demonstrar ou viver isso.

Aprende que nem sempre é suficiente ser perdoado por alguém, algumas vezes você tem que aprender a perdoar-se a si mesmo. Aprende que com a mesma severidade com que julga, você será em algum momento condenado. Aprende que não importa em quantos pedaços seu coração foi partido, o mundo não pára para que você o conserte. Aprende que o tempo não é algo que possa voltar para trás.

Portanto... Plante seu jardim e decore sua alma, ao invés de esperar que alguém lhe traga flores. E você aprende que realmente pode suportar... Que realmente é forte, e que pode ir muito mais longe depois de pensar que não se pode mais. E que realmente a vida tem valor e que você tem valor diante da vida!"

William Shakespeare

Safe and Sound (Capital Cities)

I could lift you up
I could show you what you want to see
And take you where you want to be
You could be my luck
Even if the sky is falling down
I know that we'll be safe and sound
We're safe and sound
I could fill your cup
You know my river won't evaporate
This world we still appreciate
You could be my luck
Even in a hurricane of frowns
I know that we'll be safe and sound
We're safe and sound (3x)
I could show you love
In a tidal wave of mystery
You'll still be standing next to me
You could be my luck
Even if we're six feet underground
I know that we'll be safe and sound
We're safe and sound
I could lift you up
I could show you what you want to see
And take you where you want to be
You could be my luck
Even if the sky is falling down
I know that we'll be safe and sound
I could lift you up
I could show you what you want to see
And take you where you want to be
You could be my luck
Even if the sky is falling down
I know that we'll be safe and sound
We're safe and sound (7x)

SUMÁRIO

CAPÍTULO 1

1. PREÂMBULO	2
2. INTRODUÇÃO	3
2.1. Primatas não humanos	6
2.1.1 Macaco-prego.....	6
2.1.2 Macacos-de-cheiro	7
2.1.3 Macacos-da-noite	8
2.1.4 Bugios.....	9
2.1.5 PNH e Saúde Pública	10
2.2. Metagenômica Viral	11
2.2.1 Amostragem e Enriquecimento Viral	12
2.2.2 Amplificação e preparo das bibliotecas	14
2.2.3 Sequenciamento.....	17
2.3 Vírus de importância veterinária descobertos por metagenômica ..	19
2.4 Bioinformática.....	25
3. OBJETIVOS.....	29
4. ARTIGO CIENTÍFICO 1.....	31
5. ARTIGO CIENTÍFICO 2.....	44
6. ARTIGO CIENTÍFICO 3.....	55
7. CONCLUSÃO	65
8. REFERÊNCIAS BIBLIOGRÁFICAS.....	66

ULLMANN, L.S. Pesquisa de genomas virais em primatas não humanos do Novo Mundo por metagenômica, 2014. 89p. Tese (Doutorado) - Faculdade de Medicina Veterinária e Zootecnia, Campus de Botucatu, Universidade Estadual Paulista “Júlio de Mesquita Filho”. Botucatu, São Paulo.

RESUMO

Zoonoses emergentes com origem em animais selvagens representam a mais significativa e crescente ameaça à saúde global. Os primatas não humanos contribuem com 20% das principais doenças humanas, são considerados o principal modelo experimental e também servem como sentinelas para diferentes doenças que acometem o homem. Os objetivos da presente tese foram desenvolver protocolos de metagenômica viral para detectar sequências de vírus a partir de plasma de primatas não humanos do Novo Mundo (*Saimiri sciuratus*, *Aotus ainsulatus*, *Alouatta guariba* e *Sapajus apella*) mantidos em cativeiro no Centro Nacional de Primatas (Belém, PA), no Refúgio Biológico Bela Vista (Foz do Iguaçu, PR) e no Parque Ecológico do Tietê (São Paulo, SP). A tecnologia Illumina (MiSeq) foi usada para o sequenciamento massivo e um fluxograma bioinformático foi desenvolvido para analisar as sequências obtidas. De forma geral, genomas correspondentes a diferentes famílias virais foram identificadas e ênfase foi dada ao torque teno vírus, ao vírus da hepatite G (tipos A e B), parvovírus e papilomavírus. Uma nova sequência de genoma completo do gênero *Hepacivirus*, família *Flaviviridae*, foi identificado em primatas do sul do Brasil. Os resultados demonstram a possibilidade de variedade de genomas virais presentes em primatas do Novo Mundo, considerados clinicamente saudáveis e destaca a importância de estudos que objetivam identificar possíveis vírus emergentes.

Palavras-chave: doenças infecciosas emergentes, primatas não humanos do Novo Mundo, metagenômica viral, sequenciamento massivo.

ULLMANN, L.S. Viral genomes research in New World non-human primates by metagenomics. Botucatu, 2014. 89p. Thesis (PhD) - Faculdade de Medicina Veterinária e Zootecnia, Campus de Botucatu, Universidade Estadual Paulista “Júlio de Mesquita Filho”. Botucatu, São Paulo.

ABSTRACT

Wild animals-borne emerging zoonoses represent the most significant and growing threat to global health. Non-human primates contribute with 20% of major human diseases, as they are considered the main experimental model and also can be considered sentinels for different diseases that affect humans. The aims of the present thesis were to develop a viral metagenomics protocol to detect sequences from viruses in plasma from captive New World non-human primate species (*Saimiri sciuratus*, *Aotus aenulatus*, *Alouatta guariba* e *Sapajus apella*) from National Primate Center (Belém, PA), Bela Vista Biological Sanctuary (BVBS, Foz do Iguaçu, PR), and Tietê Ecological Park (São Paulo, SP). The Illumina technology (MiSeq) was used for deep sequencing, and a bioinformatics pipeline was developed in order to analyze the sequences obtained. Overall, genomes corresponding to different viral families were identified from plasma, and emphasis was given to torque tenus virus, hepatitis G virus types A and B, parvovirus, and papillomavirus. A new whole-genome sequence from Hepacivirus genus, Flaviviridae family, was identified in monkeys from Southern Brazil. The results demonstrate the wide variety of viral genomes present on clinically healthy New World monkeys and highlight the importance of studies that aim to identify possible emerging viruses.

Key words: emerging infectious diseases, New World non-human primates, viral metagenomics, deep sequencing.

CAPÍTULO 1

1. PREÂMBULO

A presente tese foi escrita no formato publicação, de acordo com as normas do programa de Pós-Graduação em Medicina Veterinária, FMVZ, UNESP, Botucatu, São Paulo. No Capítulo 1 estão incluídos o preâmbulo, a introdução, a revisão bibliográfica e os objetivos gerais da tese. Os Capítulos 2, 3 e 4 correspondem, respectivamente ao primeiro, segundo e terceiro artigos científicos resultantes do presente estudo.

A apresentação da tese teve esta organização para facilitar o entendimento das técnicas empregadas e dos resultados encontrados. O Artigo Científico 1 é sobre metodologia para o sequenciamento massivo e foi realizado. O Artigo Científico 2 descreve os diversos genomas virais encontrados em amostras de plasma de primatas não humano do Novo Mundo encontrados pelas técnicas de metagenômica viral. Finalmente, o Artigo Científico 3 descreve um novo hepacivírus encontrado em amostras de plasma de primatas não humanos do Sul do Brasil.

Os artigos 2 e 3 encontram-se ainda em fase de elaboração. O novo hepacivírus identificado (Artigo 3) foi sequenciado a partir de um pool de amostras de plasma de Sapajus appela e de Alouatta guariba mantidos no Refúgio Biológico Bela Vista. Desta forma, as amostras terão de ser analisadas individualmente para a identificação do espécime infectado para posterior avaliação da função hepática do animal. A avaliação individual das amostras será possível, pois uma alíquota está armazenada no local onde foi coletada e a análise bioquímica do sangue total foi realizada com o sangue total coletado. Além disso, já entramos em contato com os virologistas responsáveis pela avaliação taxonômica da família Flaviviridae e os mesmos se propuseram a nos ajudar com a identificação final do novo genoma viral encontrado, que será realizada após a identificação da espécie infectada.

O hepacivírus identificado é o primeiro genoma viral encontrado pela análise dos dados obtidos e provavelmente teremos de analisar outros pools individualmente para caracterizar os diferentes genomas virais encontrados.

Diversas colaborações foram firmadas para que o presente estudo pudesse ser efetuado. As amostras foram coletadas em três locais diferentes, Centro Nacional de Primatas (Belém, PA), Refúgio Biológico Bela Vista (Foz do Iguaçu, PR) e Parque Ecológico Tietê (São Paulo, SP). O processamento das amostras foi realizado em parceria entre o Laboratório de Biologia Molecular (IBB, UNESP, Botucatu, SP), sob responsabilidade do Prof. Adj. João Pessoa Araújo Jr. e o Laboratório de Metagenômica Viral (Hemocentro da Universidade da Califórnia, São Francisco), sob responsabilidade do Dr. Eric Delwart. A FAPESP, a CAPES e o Laboratório IDEXX (Dr. Christian Leutenegger) foram os financiadores do presente projeto.

2. INTRODUÇÃO

A necessidade de maior produção de alimentos e construção de moradias decorrentes do aumento da população humana, as alterações ambientais antropogênicas e a prática de ecoturismo e eco-esportes propiciam o maior contato entre o homem e os animais silvestres, favorecendo a transmissão de importantes enfermidades infecciosas para o homem. O frequente contato entre homem e animais silvestres permite uma possível adaptação de determinado patógeno a um novo hospedeiro.

Entre 1940 e 2004, foram detectadas 335 doenças infecciosas emergentes (DIE) na população humana mundial, sendo que vírus e príons contribuem com 25,4% e bactérias e riquetsias com 54,3% (principalmente cepas resistentes a antimicrobianos). Os patógenos zoonóticos correspondem a maioria das DIE (60,3%) e destas, 71,8% tem origem em animais silvestres (JONES et al., 2008).

O Brasil é considerado um dos *hotspots* para a emergência de DIE devido à grande diversidade de espécies de animais silvestres, o que é, isoladamente, um importante preditor para a emergência de DIE, independente do crescimento da população humana, latitude e pluviosidade (JONES et al., 2008).

Desta forma, a busca por patógenos na fauna silvestre brasileira tem grande importância para a saúde pública em geral. Os primatas não humanos (PNH) são os animais filogeneticamente mais próximos ao ser humano e, apesar de contribuírem com apenas 0,5% das espécies de vertebrados, contribuem com 20% das principais doenças humanas (WOLFE et al., 2007) e servem ainda como sentinelas e modelos experimentais para diversas doenças.

Diversas DIE causadas por novos patógenos que acometeram o homem pela primeira vez e patógenos cuja incidência esteja aumentando ou se expandindo para áreas onde não haviam sido relatados vem ocorrendo no mundo todo. Muitos destes patógenos são conhecidos e técnicas diagnósticas específicas são encontradas em diversos laboratórios, facilitando o diagnóstico. Por outro lado, novos agentes, principalmente virais, tem sido descobertos por técnicas mais elaboradas, como as ferramentas de metagenômica.

A metagenômica é uma importante ferramenta para a descrição de novos micro-organismos. Devido à grande variedade e dificuldade de isolamento dos vírus e pelo fato das últimas pandemias terem sido causadas por vírus, a metagenômica viral tem crescente utilização para detectar sequências de ácidos nucleicos virais provenientes de diferentes biomas e amostras biológicas. Diversos protocolos podem ser empregados, mas a etapa fundamental é o enriquecimento viral, com a eliminação de contaminantes, como DNA mitocondrial e RNAm do hospedeiro, bactérias e outras formas de ácido nucleico presentes na amostra.

Devido à proximidade do homem com os primatas não humanos e à diversidade de espécies que habitam o Brasil, a presente tese teve como objetivo a identificação de vírus em PNH mantidos em cativeiro em três diferentes centros brasileiros. Apesar de serem animais clinicamente saudáveis, o metagenoma viral identificado nestes animais pode servir como comparação em estudos futuros com PNH doentes.

A presente tese é um estudo inédito brasileiro, realizado com primatas não humanos do Novo Mundo e técnicas de metagenômica viral, que visa em

última estância a conservação de PNH do Novo Mundo, mas também tem importante papel na identificação de novos genomas virais, uma vez que alguns dos genomas identificados podem pertencer a um possível vírus emergente que afete não somente a população de primatas do Novo Mundo mas também o homem. Além disso, as técnicas de metagenômica podem servir como ferramenta de triagem em estudos epidemiológicos quando as técnicas diagnósticas clássicas falham na identificação do patógeno causador de determinada enfermidade.

2. REVISÃO DE LITERATURA

2.1. Primatas não humanos

Os primatas constituem um dos grupos com maior variedade entre os mamíferos, com aproximadamente 280 espécies, tamanho variado, com a maior parte das espécies habitando os trópicos. O Brasil apresenta a maior variedade de espécies de primatas do mundo pela localização geográfica na região neotropical e também devido ao vasto território brasileiro, com diferentes biomas, que comporta espécies de diferentes hábitos (PISSINATTI et al., 2010).

A ordem Primates é subdividida em duas subordens, Prossimi e Antropoidea. Os primeiros são encontrados em Madagascar, no continente africano e em países do sudoeste asiático. A subordem Antropoidea é subdividida em Platyrrhini (Novo Mundo) e Catarrhini (Velho Mundo) (PISSINATTI et al., 2010).

Os Platyrrhini são encontrados na região neotropical, compreendida do sul do México até o norte da Argentina, com cinco famílias: Callitrichidae, Cebidae, Aotidae, Pitheciidae e Atelidae (PISSINATTI et al., 2010) e 18 gêneros: *Cebuella*, *Mico*, *Callithrix*, *Saguinus*, *Leontopithecus*, *Callimico*, *Saimiri*, *Cebus*, *Aotus*, *Callicebus*, *Pithecia*, *Chiropotes*, *Cacajao*, *Alouatta*, *Ateles*, *Lagothrix*, *Oreonax* e *Brachyteles*. Todas as espécies estão incluídas na lista da “Conventional on International Trade in Endangered Species of Wild Fauna and Flora” (CITES), apêndice II, que indica que há algum grau de vulnerabilidade, apesar de nem todas estarem ameaçadas de extinção (VERONA & PISSINATTI, 2007).

A seguir decorre a descrição de algumas espécies de primatas, as quais foram amostradas nesse estudo.

2.1.1 Macaco-prego

Os macacos-prego (*Sapajus apella*, Linnaeus, 1758) (ALFARO et al., 2012) habitam quase toda a região neotropical, encontrados em florestas de

clima temperado e tropical, desde Honduras até o norte da Argentina, compreendendo o leste da Colômbia, Venezuela, Guianas, leste do Peru, Brasil, Bolívia e Paraguai (AURICCHIO, 1995). Apresentam o hábitat mais diversificado: todos os estratos arbóreos de florestas chuvosas inundáveis ou não, florestas primárias, secundárias, caatingas, palmeirais e manguezais; com hábitos essencialmente arborícola e diurno (REIS et al., 2011).

Macacos-prego medem entre 32-56 cm de comprimento, pesam até 4 Kg e podem viver até 11 anos. Apresentam pelagem de coloração castanha, que varia com a idade, principalmente em machos dominantes. Vivem em bandos de oito a 16 indivíduos, com um macho dominante. São animais gregários, sedentários e formam associações com outros primatas, formando territórios diferentes dos originais de cada espécie. Sendo onívoros, possuem dieta que inclui frutos (60%), sementes (7%), castanhas, flores, gomas, néctar, fungos, seiva, ovos, insetos, aracnídeos, pequenos vertebrados, ostras e caranguejos, mas não folhas (KINDLOVITS & KINDLOVITS, 2009).

2.1.2 Macacos-de-cheiro

Os macacos-de-cheiro (*Saimiri sciureus*, Linnaeus, 1758) pertencem à mesma família que o macaco-prego. Vivem em diversos habitats de climas temperado e tropical, desde florestas primárias, secundárias, secas, tropicais de terra firme até as temporariamente inundadas, de altitudes entre 0 e 2.000 m (AURICCHIO, 1995). Apresentam, desta forma, ampla distribuição geográfica, habitando desde a Costa Rica até o chaco paraguaio. As florestas de galeria, que crescem ao longo das margens dos rios, são o hábitat preferido desta espécie de primata neotropical (REIS et al., 2011).

Macacos-de-cheiro medem cerca de 35 cm de comprimento, pesam aproximadamente 1 Kg e podem viver entre 15-20 anos. A pelagem varia de acordo com a parte do corpo, podendo ser verde-oliva a alaranjada nas partes superiores e branco-amareladas nas inferiores. São sedentários, gregários, com hábito arborícola e diurno. Vivem em grandes grupos (30-50 indivíduos), são sociáveis e misturam-se com outros primatas, assim como os macacos-prego. Esta associação entre diferentes espécies e até gêneros de primatas é

considerada uma forma de proteção contra predadores e melhora a eficiência da alimentação dos primatas (KINDLOVITS & KINDLOVITS, 2009; REIS et al., 2011).

Assim como o macaco-prego, os macacos-de-cheiro são onívoros, com dieta que inclui frutos, (diâmetro menos que 1 cm) (90%), néctar (6%), sementes (2%), flores (1%), insetos (larvas), moluscos terrestres e pequenos vertebrados como lagartixas, pererecas e filhotes de aves (KINDLOVITS & KINDLOVITS, 2009).

2.1.3 Macacos-da-noite

Os macacos-da-noite (*Aotus infulatus*, Olfers, 1818) habitam florestas de clima temperado e tropical do sul do Panamá, Guianas, Peru, Equador, Oeste do Brasil e norte da Argentina, desde as florestas tropicais úmidas até as capoeiras. Medem entre 24-37 cm, pesam 1 Kg e vivem por 25 anos, com relato de 7 anos de vida em cativeiro. Diferente das outras amostradas no presente estudo é a única espécie de PNH com hábito noturno (DIAS et al., 2011). Apesar da visão adaptada à ambientes escuros é capaz de diferenciar cores. Possuem olhos grandes, com pelagem de tonalidade cinzenta-acastanhada, com destaque para a face com áreas de pêlos brancos nos supercílios e bochechas e três listras negras na fonte (KINDLOVITS & KINDLOVITS, 2009).

Os macacos-da-noite abrigam-se em árvores ocas ou escondido entre os ramos durante o dia, mas saem à procura de alimento durante a noite. Não apresentam preferência por nenhum estrato arbóreo ou tipo de vegetação (AURICCHIO, 1995). Andam sempre em pares, com hábito monogâmico, mas vivem em grupos de aproximadamente 12 indivíduos. São gregários, sedentários e onívoros, e a dieta é composta de folhas (30%), frutos (60%), invertebrados (larvas e pupas de lepidópteros, besouros e aranhas) e, eventualmente, morcegos e ovos de pássaros. Há relato de que estes primatas não ingerem água devido ao hábito noturno, o que diminui a desidratação (KINDLOVITS & KINDLOVITS, 2009). Também por razão de hábito noturno, os

macacos-da-noite são muito utilizados em pesquisas sobre a visão (PISSINATTI et al., 2010).

2.1.4 Bugios

A última espécie amostrada são os bugios (*Alouatta guariba*, Humbolt, 1812), encontrados desde o sul do México até o norte da Argentina. No Brasil são encontrados no Amazonas, Pará, Maranhão, Paraíba, Minas Gerais, Ceará, Mato Grosso, Goiás, Minas Gerais, São Paulo, Bahia, Espírito Santo, Rio de Janeiro, Paraná e Rio Grande do Sul. Habitam florestas montanhosas úmidas ou vegetação mais aberta, como cerrado, caatinga e florestas de araucárias, em altitudes de 0 a 1.200 metros (KINDLOVITS & KINDLOVITS, 2009).

Os bugios pesam até sete quilogramas, vivem por até 20 anos, apresentam corpo robusto e pelagem longa, mais volumosa na mandíbula e nos lados da face, que forma uma vasta barba sob a face negra e que esconde o grande volume do osso hióide. Este osso, mais desenvolvido nos machos, serve como caixa de ressonância para a sua voz, o que lhes confere uma potente vocalização, podendo ser ouvida a aproximadamente 5 Km de distância (AURICCHIO, 1995). Costumam vocalizar em grupo em certas horas do dia (alvorecer e entardecer) para informar sua posição a outros grupos que estão próximos. Vivem em grupos de quatro a seis indivíduos, mas o grupo pode ter até 15 espécimes. Os machos compreendem 30% do grupo, 40% são fêmeas e 30% são filhotes, são comandados por um macho adulto, o “capelão”. São animais lentos, que passam horas parados nas árvores (KINDLOVITS & KINDLOVITS, 2009), diferentemente das outras espécies estudadas.

A dieta dos bugios é composta por folhas (40-60%), em especial de folhas novas e frutos (manga, figo e embaúba), que comem antes das outras espécies, pois possuem a capacidade de detoxificar as defesas químicas de muitas plantas, evitando intoxicações, e aumentando a variedade de alimentos à sua disposição (AURICCHIO, 1995). Apresentam estômago compartimentalizado e um grande intestino, o que corrobora com sua preferência por folhas e pecíolos. A mucosa oral apresenta coloração preta

devido aos pigmentos das folhas utilizadas na alimentação. Alto teor de fibras deve ser oferecido a estes primatas quando em cativeiro (KINDLOVITS & KINDLOVITS, 2009).

2.1.5 PNH e Saúde Pública

A composição da dieta e o hábitat em que as espécies vivem influencia a ocorrência de determinado micro-organismo nos animais, seja patógeno ou apenas contaminante encontrado no alimento.

Diversas viroses foram identificadas em PNH neotropicais, incluindo: herpesviroses em macaco-de-cheiro, *Aotus*, *Ateles*, *Cebus*, *Saguinus* e calitriquídeos; paramixovirose em calitriquídeos; sarampo em calitriquídeos, *Aotus* e *Saimiri*; encefalomiocardite em *Leontopithecus*, *Saguinus*, *Saimiri*, *Aotus*, *Cebus*, entre outros; hepatite dos calitriquídeos; febre de Mayaro em *Alouatta*, *Pithecia* e *Saguinus*; e todos são considerados susceptíveis à raiva (VERONA e PISSINATTI, 2007). Diferentes graus de susceptibilidade são encontrados nas espécies de primatas neotropicais, e algumas servem como sentinela para importantes zoonoses como a febre amarela. Os bugios, por exemplo, são altamente susceptíveis a esse vírus, e morrem de forma aguda, indicando e até predizendo a ocorrência de surtos em humanos.

As espécies *Alouatta caraya*, *Cebus negritus* e *Cebus cay* foram utilizadas como “animais sentinela” em uma avaliação de PNH de vida livre na região do município de Porto Rico-PR, alto do rio Paraná entre 2004 e 2006. Ainda, uma tropa de cavalos de fazenda, que frequentava local próximo a área de captura foi analisada. A soroprevalência encontrada para o vírus da encefalite de Saint Louis foi de 11,6% em *A. caraya*, 12,5% em *C. negritus* e 30,7% e *C. cay* pela inibição da hemaglutinação; e de 39,1% em cavalos. Tentou-se o isolamento viral em cultura celular de *Aedes albopictus*, porém sem sucesso (SVOBODA, 2007).

Pela importância dos PNH na saúde pública, as epizootias presentes nestes animais foram incluídas na relação nacional de notificação compulsória desde 2006, em cumprimento à Portaria nº 5 de 21/02/2006 da Secretaria de Vigilância em Saúde/MS, e um programa de vigilância nacional foi implantado.

A contínua pandemia do HIV, os recentes surtos de patógenos como SARS e influenza H5N1, a atual pandemia do H1N1 e as consequências globais que ainda devem ser determinadas e demonstram a nossa contínua vulnerabilidade a doenças infecciosas emergentes, reforçando a importância de estudos em animais silvestres, responsáveis pela emergência de mais da metade das últimas doenças infecciosas que acometeram o homem.

2.2. Metagenômica Viral

Os vírus são considerados a forma de vida mais abundante do planeta e menos de 1% não são cultiváveis (KUPROVIC & BAMFORD, 2011). Dessa forma, o estudo e a descoberta de vírus desconhecidos eram prejudicados pela limitação das técnicas de cultura celular e microscopia eletrônica. Na década passada, um dos acontecimentos mais notáveis da microbiologia veio auxiliar esse ramo da virologia. Com o desenvolvimento da metagenômica observou-se o estabelecimento da metagenômica viral, definido como o estudo das sequências de ácidos nucleicos de vírus não cultiváveis de diferentes biomas, baseada em diversos métodos de concentração, purificação, extração, sequenciamento e bioinformática (THURBER et al., 2009). A aplicação da metagenômica para a descoberta viral teve ampla expansão após o advento do sequenciamento massivo (LAW et al., 2013).

A metagenômica viral teve início em 2002 com a publicação de dois artigos descrevendo duas comunidades virais marinhas não-cultiváveis em San Diego (BREITBART et al., 2002) e emergiu como uma ferramenta útil para aumentar o acesso à genética e fisiologia de organismos não cultiváveis. Diversos protocolos tem sido desenvolvidos para capturar a diversidade microbiana inexplorada para superar barreiras existentes de estimação das espécies (SINGH et al., 2009).

A metagenômica é uma ferramenta poderosa para acessar a diversidade microbiana complexa presente em amostras ambientais como solo, sedimento ou água. Possibilita a descoberta de biomoléculas com propriedades biotecnológicas valiosas (SIMON e DANIEL, 2009), em estudos de medicina

forense (BRENIG et al., 2010) e para a descoberta de novos agentes virais causadores de doenças no homem e nos animais (BLOMSTRÖM, 2011). Atualmente, a rápida redução do custo das técnicas de sequenciamento de nova geração tem aumentado consideravelmente o desenvolvimento de estudos metagenômicos. Estima-se que no futuro a metagenômica será usada da mesma maneira que o gene rRNA 16S para descrever perfis de comunidades microbianas (THOMAS et al., 2012).

Um amplo estudo para descrever o microbioma humano associado com saúde e doença foi recentemente publicado (NELSON et al., 2010). Os avanços tecnológicos utilizados nos estudos metagenômicos tem proporcionado o aumento do número e dimensão dos projetos, como o projeto *Sorcerer II Global Ocean Sampling* (GOS) (RUSCH et al., 2007; BIRS et al., 2009) e a comparação metagenômica de 45 microbiomas e 42 viromas distintos (DINSDALE et al., 2008).

2.2.1 Amostragem e Enriquecimento Viral

A primeira etapa do estudo metagenômico é a obtenção da amostra, que deve representar a população da qual foi colhida (WOOLEY et al., 2010). O ácido nucléico extraído deve representar todas as células presentes na amostra e quantidades suficientes de ácidos nucléicos de alta qualidade devem ser obtidas para a produção da biblioteca e sequenciamento (THOMAS et al., 2012).

Na maior parte das amostras, o ácido nucléico viral constitui apenas uma pequena porção de todos os ácidos nucléicos presentes. Desta forma, uma das tarefas mais importantes e complexas da metagenômica viral é preparar a amostra de tal forma que os ácidos nucléicos do hospedeiro e bactérias, considerados contaminantes, sejam eliminados enquanto o ácido nucléico viral é preservado (DELWART, 2007, THOMAS et al., 2012, LAW et al., 2013).

O capsídeo presente em alguns vírus tem como função proteger o ácido nucléico viral, o que pode ser considerado vantajoso em estudos metagenômicos, pois as nucleases podem ser utilizadas para o tratamento das amostras, degradando o ácido nucléico do hospedeiro, mas não afetando o ácido nucléico viral, que permanece viável. Após a extração dos ácidos

nucléicos virais, quando o DNA ou o RNA estão livres (fora do capsídeo), pode-se fazer o tratamento com DNases quando se busca RNA vírus e com RNases quando se procura por DNA vírus (DELWART, 2007). O tratamento das amostras com nucleases pode não eliminar totalmente o ácido nucléico contaminante, principalmente em amostras como tecidos. Ainda, é importante enfatizar que há alguns vírus mais sensíveis ao tratamento com nucleases, e há, portanto, um risco de degradação do ácido nucléico viral (BLOMSTRÖM, 2011).

Quando o propósito é separar os vírus de outros micro-organismos, a filtração com membranas de poro entre 145-160 nm pode ser inicialmente utilizada para remover bactérias, eucariontes e agregados grandes. Porém, devido à existência de vírus maiores que o tamanho do poro da membrana, como o poxvírus, que tem 140-260 nm de diâmetro, estes vírus seriam eliminados da amostra, juntamente com outros contaminantes (DELWART, 2007). A ultra-centrifugação por densidade de gradiente de cloreto de cério (THURBER et al., 2009) e em colchão de sacarose 30% é outra maneira de separar as partículas virais (DELWART, 2007).

O processo de fracionamento das amostras deve ser checado para garantir que o enriquecimento suficiente do ácido nucléico alvo tenha sido alcançado e que mínima contaminação tenha restado (THOMAS et al., 2012). Este é um fator importante, pois as técnicas de sequenciamento massivo utilizadas nos estudos metagenômicos amplificam aleatoriamente as bibliotecas, ou seja, quanto menos ácido nucléico contaminante maior a chance de detecção viral.

A purificação do ácido nucléico viral antes da amplificação é etapa crucial nos estudos metagenômicos e os métodos citados acima podem ser empregados na mesma amostra, uma vez que quando se desconhece a partícula viral presente na amostra, a combinação de dois ou mais métodos pode aumentar a sensibilidade do estudo.

Anticorpos de pacientes convalescentes de doença respiratória e gastrointestinal foram utilizados para capturar os vírus presentes nas amostras biológicas. Sabe-se que mesmo após algum tempo da infecção, IgGs são encontradas no soro de pacientes em convalescência. Estas imunoglobulinas

podem ser ligadas a *beads* magnéticas e utilizadas para a detecção do vírus na amostra. Esta estratégia de enriquecimento viral apresenta a vantagem de selecionar vírus, que tem maior chance de serem os causadores de determinada doença (MUNNINK et al., 2013).

2.2.2 Amplificação e preparo das bibliotecas

Diversas técnicas de amplificação podem ser utilizadas para a identificação de novos genomas virais e, por ser um método de amplificação, alguns problemas como reagentes contaminados, formação de quimeras e viés na amplificação podem ocorrer. O impacto destes problemas dependerá da quantidade e do tipo de material inicial e do número necessário de ciclos de amplificação para produção de quantidades suficientes de ácidos nucleicos (THOMAS et al., 2012). A escolha da técnica vai depender do objetivo do estudo.

A amplificação dos ácidos nucleicos pode ser realizada de maneira independente da sequência para demonstrar a verdadeira composição genética da amostra. Esta etapa é capaz de multiplicar simultaneamente diversos genomas virais, incluindo vírus altamente divergentes e completamente novos, e, portanto, garantir sua descoberta e caracterização genética (BLOMSTRÖM, 2011).

A amplificação independente da sequência com um *primer* (*sequence-independent single-primer amplification* - SISPA) surgiu há quase duas décadas com o objetivo de identificar ácido nucleico viral a partir de sequências desconhecidas presentes nas amostras em pequenas quantidades. O método envolve a digestão do DNA por endonuclease, seguida de ligação direcional de um adaptador assimétrico, ou *primer*, nas duas porções finais da molécula de DNA. As sequências finais comuns ao adaptador permitem a amplificação do DNA em subsequente reação de PCR usando um único *primer* complementar. Devido à baixa complexidade do genoma viral, a digestão enzimática produz grande quantidade de um limitado número de fragmentos. Após a amplificação, bandas discretas no gel de agarose são observadas e podem ser então sequenciadas e identificadas (BEXFIELD e KELLAM, 2011). Pode ser utilizada para a identificação de genomas virais novos com alta divergência, porém ácidos nucleicos contaminantes podem ser amplificados (DELWART, 2007).

A SISPA foi utilizada para a detecção do vírus da hepatite C a partir de biopsias hepáticas. O cDNA foi obtido a partir do RNA com a utilização da enzima de transcrição reversa Superscript III e o *primer* senso utilizado na SISPA. RNA e DNA foram então incubados com a Klenow para obtenção da segunda fita de DNA e a reação de PCR foi realizada com o *primer* reverso da SISPA (DALY et al., 2011). Dois circovírus-símile foram identificados a partir de amostra de fezes pela técnica de SISPA (CASTRIGNANO et al., 2013).

Uma alternativa de tecnologia de amplificação é a PCR randômica (rPCR), que não requer a digestão do DNA ou cDNA nem a ligação a adaptadores (BLOMSTRÖM, 2011). Este método utiliza dois *primers* diferentes e duas reações de PCR. O único *primer* usado na primeira PCR tem uma sequência definida no final 5', seguido de uma sequência hexamérica ou heptamérica degenerada no final 3'. A segunda PCR é então realizada com um *primer* específico complementar à região 5' do primeiro *primer*, possibilitando a amplificação dos produtos formados na primeira reação. A técnica de PCR randômico tem sido utilizada para a detecção de DNA e RNA vírus e é uma das técnicas mais utilizadas atualmente na identificação de novos vírus (BEXFIELD e KELLAM, 2011).

A amplificação por deslocamento (*rolling circle amplification* - RCA) é outra técnica eficiente que utiliza *primers* randômicos em conjunto com uma polimerase por deslocamento. Uma das polimerases de deslocamento de alta fidelidade é a Phi29 (fago *Bacillus subtilis*) com alta processividade, podendo incorporar até 20.000 bases antes de se destacar do alvo, além de atividade 5' - 3' exonuclease (BEXFIELD e KELLAM, 2011). Utilizando-se *primers* randômicos e phi29, alvos circulares são amplificados eficientemente por uma reação denominada de amplificação por círculo rolante, que cria longos concatâmeros de múltiplas cópias do alvo. Pequena quantidade de material é necessária para a reação, e além de alvos circulares, alvos lineares também podem ser amplificados por esta técnica (BLOMSTRÖM, 2011).

Um novo método eficiente e altamente preciso foi desenvolvido para o sequenciamento massivo de RNA vírus. Apesar do protocolo ser laborioso (aproximadamente 5 dias de trabalho), resulta em sequências com taxa de erro mais baixas e facilita a identificação de variações raras de baixa frequência.

Brevemente, o RNA é fragmentado, circularizado e selecionado por tamanho (85 -100 nucleotídeos). A transcrição reversa é realizada pela reação de círculo rolante, que resulta em um cDNA-*tandem*, que são clonados para gerar a segunda fita de DNA, já contendo os adaptadores necessários para o sequenciamento (ACEVEDO & ANDINO, 2014).

O PCR degenerado utiliza *primers* para anelar em regiões parcialmente conservadas de vírus relacionados. O objetivo principal é criar um equilíbrio na tentativa de cobrir todas as variantes possíveis dentro de uma família viral, criando-se uma grande variedade de *primers*. Esta técnica é utilizada para a detecção de vírus, inclusive novos vírus, pela existência suficiente de homologia entre as famílias virais (BEXFIELD e KELLAM, 2011). Diversos vírus produzem sua própria polimerase e existe uma semelhança tanto na sequência como na estrutura destas proteínas, pois exercem uma importante função. Desta forma, as DNA e RNA polimerases podem ser utilizadas na identificação de agentes virais pela proximidade com uma família viral conhecida (SONNTAG et al., 1995; LAKSHMINARAYAN et al., 2003). Dois novos herpesvírus foram identificados de *Macaca nestrina* e *Macaca mulatta* (ROSE et al., 1997) e dois radinovírus (*Chlorocebus* radinovírus 1 e 2), em macacos verdes africanos (GREENSILL et al., 2000), com *primers* degenerados que amplificam a região conservada do gene da DNA polimerase dos vírus da família Herpesviridae.

O PCR longo foi utilizado para amplificar o ácido nucléico mitocondrial de amostras de sangue, gerando fragmentos de aproximadamente 8 Kb em comprimento. As bibliotecas foram preparadas com o kit Nextera-XT e sequenciadas com a tecnologia Illumina (KING et al., 2014). O PCR de longo alcance foi também utilizado para a detecção de DNA mitocondrial a partir de diferentes amostras como swabs de bochecha, cultura celular e DNA previamente extraído. Diferentes *primers* foram utilizados para a obtenção de fragmentos de aproximadamente 8,5 Kb (McELHOE et al., 2014).

Muitas vezes a amplificação inicial do ácido nucléico viral não é necessária para os estudos metagenômicos. Devido à variedade de ácidos nucléicos virais, diferentes metodologias para o preparo da biblioteca podem ser empregados. Vírus dsDNA podem ser mais facilmente preparados para o

sequenciamento massivo em comparação com vírus RNA, porém o enriquecimento viral é prejudicado pela dificuldade na separação do DNA viral e do hospedeiro.

Se tratando de amostras de vírus RNA ou ssRNA, o DNA complementar deve ser obtido a partir do RNA para posterior obtenção da segunda fita de DNA. Algumas enzimas como a Klenow e T4 DNA Polimerase, entre outras polimerases, tem sido utilizadas na obtenção da dsDNA (DALY et al., 2011; CASTRIGNANO et al., 2013; MUNNINK et al., 2013). O tratamento com a RNase H pode ser utilizado para degradar a fita de RNA inicial, previamente a obtenção da dsDNA (JONES et al., 2005). De acordo com a tecnologia utilizada, quando se busca vírus ssDNA, a segunda fita de DNA também deve ser obtida.

Dependendo da tecnologia de sequenciamento massivo utilizada, diferentes kits para o preparo das bibliotecas estão disponíveis no mercado. Dentre eles, o Nextera-XT (Illumina) é indicado para o preparo de genomas pequenos, amplicons e plasmídeos e requer apenas 1 ng de dsDNA como *input*. Esta técnica consiste de uma etapa de tagmentação enzimática que adiciona adaptadores que permitirão a amplificação por PCR e adição dos primers para o sequenciamento (McELHOE et al., 2014).

2.2.3 Sequenciamento

Diferentes técnicas de sequenciamento tem sido frequentemente utilizadas para a identificação de ácidos nucleicos. Uma técnica é construir biblioteca viral *shotgun* e sequenciar os fragmentos por meio de métodos como o Sanger, conhecido como “método de terminação de cadeia” ou “sequenciamento didesoxi” (BEXFIELD e KALLAM, 2011; BLOMSTRÖM, 2011). A técnica é baseada na síntese de uma fita complementar dependente da polimerase na presença de 2'-desoxinucleotídeos (dNTPs) e 2',3'-didesoxinucleotídeos, que servem como finalizadores não reversíveis da síntese (BEXFIELD e KALLAM, 2011).

A técnica de Sanger cria sequências de alta qualidade e pode ser empregada para produzir sequências de leitura de até quase 1.000 nt (BLOMSTRÖM, 2011). Uma possível limitação da técnica é a necessidade de

clonagem viral em bactérias antes do sequenciamento (BEXFIELD e KALLAM, 2011). A técnica de Sanger é altamente laboriosa em comparação com seu rendimento e a utilização de novas técnicas de alta transferência vem substituindo o sequenciamento de Sanger em estudos de metagenômica viral (BLOMSTRÖM, 2011).

O sequenciamento com alta taxa de transferência ou sequenciamento massivo não necessita da etapa de clonagem e produz maior quantidade de dados comparado ao sequenciamento de Sanger. Para a metagenômica viral, este fato permite a detecção de vírus com menor número de cópia inicial (BLOMSTRÖM, 2011). Também conhecidos como “tecnologias de sequenciamento de nova geração”, os métodos de sequenciamento massivo incluem o instrumento baseado em pirosequenciamento 454 (Roche Applied System), o analisador de genoma (Illumina) e o sistema SoLiD (Thermo Fhiesher) (BEXFIELD e KELLAM, 2011). Os avanços destas técnicas e a redução do custo tem aumentado o número e a dimensão dos projetos de sequenciamento metagenômicos (SIMON e DANIEL, 2011).

As três tecnologias diferem no método de amplificação, na quantidade de dados gerados e na forma de leitura. A plataforma 454 utiliza para a amplificação clonal a PCR em emulsão (ePCR), enquanto a Illumina usa a PCR por pontes. Ainda, a plataforma 454 utiliza o pirosequenciamento, a Illumina utiliza terminadores reversíveis e o SOLiD utiliza ligação e probes cliváveis. A maior capacidade de *throughput* é feita pelo aparelho SOLiD (10-20 Gb/corrída), seguida pelo Illumina (3-6 Gb/corrída) e pelo 454 (0,4-0,6Gb/corrída). Apesar do menor *throughput*, a plataforma 454 realiza leituras mais longas (até 400 nt), seguida pelo Illumina (aproximadamente 100 nt) e SOLiD (aproximadamente 50 nt). Todas as técnicas apresentam vantagens e desvantagens e a escolha do sistema ideal vai depender do tipo de estudo a ser realizado (BLOMSTRÖM, 2011).

A plataforma 454 System Roche faz a PCR por emulsão (ePCR) para a amplificação clonal, onde fragmentos de DNA randômicos são ligados a *beads*, que são depositadas em poços de uma placa *picotitre*. Em seguida, o pirosequenciamento é feito individualmente. O pirosequenciamento envolve a adição sequencial dos quatro desoxinucleotídeos trifosfato à reação. Caso haja

complementariedade, são incorporados pela polimerase. Quando ocorre a polimerização, há a liberação de pirofosfatos, que são convertidos por duas reações enzimáticas e produzem luz. A luz é detectada em paralelo via dispositivo ligado à carga (*charge-coupled device* – CCD), convertida na sequência da fita. Esta técnica apresenta relativo baixo custo, comprimento de leitura entre 600 e 800.000 bp, 10 ng de DNA para iniciar a reação e capacidade de *throughput* de até 500 Mbp (THOMAS et al., 2012).

Já a plataforma Illumina/Solexa tem como princípio a imobilização de fragmentos de DNA randômico em uma superfície, ou seja, a amplificação clonal ocorre em superfície sólida. A amplificação ocorre em *clusters* de fragmentos de DNA idênticos, que subsequentemente são sequenciados com finalizadores removíveis, conhecida como sequenciamento por síntese. Esta técnica também apresenta baixo custo, tem grande aplicação na metagenômica, utilizada na geração de “rascunhos” de genoma, 20 ng de DNA podem ser usados como amostra inicial, mas a plataforma disponível há mais tempo, a HiSeq, apresenta tempo de corrida maior que o da plataforma 454. Esta desvantagem está sendo corrigida pela nova versão da tecnologia Illumina, o MiSeq, que também pode ser utilizado para sequenciamentos teste (THOMAS et al., 2012).

2.3 Vírus de importância veterinária descobertos por metagenômica

Estudos metagenômicos podem auxiliar na identificação de novos agentes causadores de infecções no homem e nos animais, principalmente quando patógenos conhecidos foram excluídos como causadores.

Um novo bornavirus foi identificado em um papagaio-do-peito-roxo (*Amazona vinacea*) e duas araras canindé (*Ara ararauna*) que foram diagnosticadas com a doença da dilatação pró-ventricular. A doença causa inflamação do sistema nervoso central, periférico e autônomo e é associada com disfunção gastrointestinal. De acordo com o histórico das aves e pelas características histológicas, realizou-se PCR randômica seguida de sequenciamento massivo (pirosequenciamento) a partir de amostras de RNA extraídas de cérebro. Após a identificação de bornavírus, realizou-se PCR para

confirmação da presença do vírus. A presença do vírus nas amostras coloca-o como possível agente causador, pois não foi possível concluir se o vírus induz a doença da dilatação pró-ventricular ou se uma resposta imune adaptativa ocorre em associação com a doença (HONKAVOURI et al., 2008).

O vírus da hepatite dos perus foi identificado a partir do pirosequenciamento de RNA a partir de fígado de aves com hepatite. A presença do vírus em fígado, bile, intestino, soro e swab cloacal foi obtida por PCR em tempo real, com *primers* específicos ao vírus identificado pelo sequenciamento massivo. A hibridização *in situ* com probes virais e imunoistoquímica também confirmaram a presença de ácido nucléico e proteínas virais no fígado dos perus doentes (HONKAVOURI et al., 2011).

A fibropapilomatose é uma doença neoplásica debilitante que afeta tartarugas marinhas no mundo todo. A causa da doença é considerada desconhecida, mas sugere-se que um alfaherpesvírus seja o agente etiológico. Partículas virais foram purificadas a partir de fibropapiloma com uma combinação de seleção por tamanho, centrifugação por densidade e tratamento com nuclease. O DNA extraído foi amplificado por SISPA e a identificação do tornovírus 1 de tartaruga marinha foi obtida. Verificou-se ainda que o vírus circula como quasi-espécie, pois variantes semelhantes foram encontradas nos animais com as lesões (NG et al., 2009).

A doença de corpo de inclusão é uma doença infecciosa fatal de serpentes e é caracterizada por alterações comportamentais e infecções secundárias. Com a utilização da metagenômica, dois arenavírus foram identificados. O vírus identificado apresenta a organização genômica semelhante a dos arenavírus, porém codificam uma glicoproteína mais semelhante a dos filovírus. O vírus foi isolado em células de jiboia. A nucleoproteína viral foi encontrada em inclusões citoplasmáticas grandes (STENGLEIN et al., 2012).

Martas (*Mustela lutreola*) com desordem neurológica de uma fazenda da Dinamarca apresentaram tremor, andar cambaleante e ataxia. A inoculação de homogeneizado cerebral de animais doentes em sádios resultou na reprodução de sinais clínicos. A amplificação randômica seguida de sequenciamento massivo resultou na identificação de um astrovírus, uma das primeiras

caracterizações destes vírus em sistema nervoso central (BLOMSTRÖM et al., 2010).

Três leões marinhos (*Otaria byronia*) do Zoológico Municipal de Kansas morreram sem causa aparente. A necropsia revelou alterações respiratórias e a pesquisa de intoxicação química não detectou pesticidas, componentes orgânicos tóxicos ou níveis anormais de minerais na água das piscinas ou no conteúdo gástrico dos animais. Foram negativos também para West Nile e nenhum vírus foi isolado. A metagenômica viral foi realizada a partir de fragmentos congelados de pulmão. O enriquecimento viral foi realizado por centrifugação inicial para remoção das células do hospedeiro, filtração em membranas de 0,45 µm, concentração por gradiente de Césio e tratamento com DNase. O DNA viral foi amplificado pela RCA com kits comerciais, clonado e sequenciado. O anelovírus identificado foi considerado causador da morte dos animais e foi posteriormente identificado em leões marinhos de vida livre (NG et al., 2009).

Um anelovírus semelhante ao identificado em leões marinhos e gatos foi encontrado em focas (*Phoca vitulina richardsii*) após a morte de 21 animais. Este é o segundo anelovírus identificado em pinípedes, sugerindo que estes vírus circulam em mamíferos marinhos (NG et al., 2011) e reforça a importância da investigação dos casos de morte súbita, muito comuns em animais silvestres.

A PCR com *primers* degenerados foi utilizada para identificar o herpesvírus-2 de mustelídeos encaminhados a um centro de reabilitação após vazamento de óleo. A identificação de lesões herpéticas, úlceras e placas orais com inclusões intranucleares eosinofílicas) em lontras marinhas (*Enhydra lutris kenyoni*) pode ser resultado de reativação viral associado ao estresse causado pela exposição ao óleo, captura e reabilitação. Os autores sugerem que o herpesvírus identificado é comum e geralmente não causa sintomatologia clínica significativa nas lontras (TSENG et al., 2012).

A associação de hibridização e RCA resultou na identificação de um novo poliomavírus associado a tumor cerebral maligno em dez guaxinins de vida livre (*Procyon lotor*) dos estados da Califórnia e Oregon. Os tumores estavam localizados no lobo frontal e trato olfatório dos animais e suspeitou-se

de um agente emergente e infeccioso pela dispersão dos animais (DELA-CRUZ et al., 2013).

A doença de Theiler é uma hepatite aguda de equinos que é associada com a administração de produtos de sangue equino. Sequenciamento massivo foi utilizado para identificar um flavivírus, do gênero *Pegivirus*, denominado de vírus associado a doença de Theiler (TDAV). Um estudo epidemiológico com equinos de outros locais confirmou a associação do vírus e a hepatite aguda que ocorre nos equinos. A inoculação experimental do vírus indica que a viremia não influencia diretamente na lesão hepática (CHANDRIANI et al., 2013).

A metagenômica viral foi utilizada para a identificação de um novo orthobunyavirus, denominado Schmallerberg, incriminado como causador de uma enfermidade em bovinos na Alemanha e na Holanda, em 2011. Os animais apresentavam sinais como febre, diminuição da produção de leite e diarreia, porém as viroses endêmicas clássicas foram excluídas como causadoras da enfermidade. Após a identificação do vírus como sendo um orthobunyavírus pelo estudo metagenômico, foi desenvolvida uma nova técnica de PCR em tempo real para analisar amostras de outros animais e verificar a prevalência do vírus para confirmar o Schmallerberg como o vírus causador deste surto (HOFFMANN et al., 2012).

Um grande projeto de ecologia, saúde e conservação de primatas é realizado no Parque Nacional Kibale, na Uganda. O parque tem alta variedade e densidade de primatas não humanos, que tem contato com as pessoas e com os animais domésticos (SIBLEY et al., 2014). Diversos vírus foram identificados nos primatas que habitam o parque, conforme descrito abaixo.

Um colobus vermelho (*Procolobus rufomitatus tephrosceles*) de vida livre foi capturado, mesmo estando sadio, para colheita de amostra para metagenômica viral, que identificou duas variantes do vírus da síndrome hemorrágica dos símios (VSHS). Resumidamente, a amostra foi centrifugada, filtrada em membrana de 0,45 µm e o RNA extraído foi tratado com DNase e após o preparo da biblioteca foi realizado o pirosequenciamento. Esta foi a primeira detecção do VSHS em primata de vida livre e destaca-se a importância deste estudo não somente pela detecção deste vírus, mas também

pela sintomatologia causada, semelhante à febre amarela e ebola, importantes infecções humanas, ressaltando-se ainda a importância dos primatas como reservatórios de patógenos emergentes (LAUCK et al., 2011).

Duas novas variantes do VSHS também foram identificadas em quatro guenons-da-cauda-vermelha (*Cercopithecus ascanius*) do mesmo parque descrito acima. As duas variantes apresentam características muito diferentes dos outros VSHS já descritos e os autores sugerem a reclassificação dos Aterivírus, sugerindo a inclusão do gênero Simartevirus (LAUCK et al., 2013a).

Um vírus novo, divergente e semelhante ao da hepatite G foi identificado por sequenciamento massivo em três de nove colobus preto-e-branco (*Colobus guereza*). O vírus foi denominado de guereza hepacivírus (GHV) e se assemelha filogeneticamente a hepacivírus identificados em morcegos africanos e roedores norte-americanos (LAUCK et al., 2013b).

Amostras de plasma de nove colobus preto-e-branco (*Colobus guereza*) foram analisadas por PCR randômico e sequenciamento massivo para verificar a variedade de vírus da imunodeficiência dos símios. Alíquotas de 1 mL de plasma foram centrifugadas e filtradas para remoção de resíduos celulares, seguida de extração do RNA, tratamento com DNase, obtenção de cDNA e dsDNA e preparo das bibliotecas para sequenciamento. Com o intuito de sequenciar os genomas completos dos SIV, quatro reações com *amplicons* de aproximadamente 2,5 Kb foram padronizadas a partir das sequências obtidas pelo *de novo assembly* das leituras do sequenciamento. Duas novas variantes do SIV foram encontradas, SIVKcol-1 e SIVKcol-2, em três e quatro animais, respectivamente e nenhum dos animais estava co-infectado (LAUCK et al., 2013c).

Uma nova linhagem do vírus da imunodeficiência dos símios (SIV), denominada SIVrtg_Kib, foi encontrada em guenons-da-cauda-vermelha (*Cercopithecus ascanius schmidtii*) ainda no mesmo parque. Esta espécie pertence ao gênero *Cercopithecus*, de grande importância, pois compreende a maior parte das espécies de primatas que albergam o SIV. Amostras de 12 animais foram coletadas, de grupos diferentes que habitam o parque. Como o SIV já havia sido relatado em outras espécies deste gênero no parque, as amostras de plasma foram triadas para SIV/HIV. Por sequenciamento massivo,

três amostras foram positivas e apenas duas amostras foram positivas pela sorologia, indicando infecção aguda. Por outro lado, a partir de duas amostras sorologicamente positivas não foi possível a detecção do vírus pelo sequenciamento. Realizou-se então RT-PCR na tentativa de detectar um fragmento de 400 pb nas duas amostras de plasma. O produto obtido foi sequenciado pelo método de Sanger, confirmando a presença do SIVrtg_Kib (LAUCK et al., 2014).

Outras variantes do SIV (LOCATELLI et al., 2008; GOLDBERG et al., 2009; LEENDERTZ et al., 2010), os vírus da célula T linfotrófica dos símios e o *simian foamy virus* (GOLDBERG et al., 2009) foram encontrados em colobus vermelhos (*Piliocolobus badius badius*). Muitos estudos com primatas não humanos do velho mundo tem sido conduzidos devido à importância desses animais como transmissores do HIV para o homem. A metagenômica viral tem ampla aplicabilidade na identificação de novos vírus, principalmente os relacionados ao HIV, pelo fato de já serem conhecidos. A grande variedade do SIV encontrada nas diferentes espécies pode ser diferenciada pelas técnicas de metagenômicas.

O sequenciamento massivo de amostras de colobus vermelhos, colobus-da-cauda-vermelha e babuínos verde-oliva (*Papio anubis*) do parque Kibale revelou a existência de três pegivírus, espécie-específicos, denominados SPgV_{krc}, SPgV_{krtg} e SPgV_{kbab}, de acordo com a espécie e local onde foram encontrados. Estes são os primeiros pegivírus detectados em colobus e demonstram a variedade de vírus encontrados nestas espécies ameaçadas (SIBLEY et al., 2014).

De acordo com a revisão acima, observa-se a importância da metagenômica viral na descoberta de novos vírus em diferentes espécies de animais, o que auxilia na identificação precoce de possíveis vírus emergentes que possam causar importantes surtos nos animais e também no ser humano, uma vez que a maior parte das doenças infecciosas humanas tem origem zoonótica.

Destaca-se ainda a relevância dos vírus encontrados em amostras de animais silvestres, principalmente os primatas não-humanos, considerados os principais transmissores de agentes infecciosos emergentes ao ser humano. A

identificação de novos flavivírus em primatas do Velho Mundo (LAUCK et al., 2013b; SIBLEY et al., 2014) destaca a importância dos vírus da família Flaviridae nos PNH. O Brasil é considerado como santuário de arboviroses (BEDIN, 2007), dentre elas as flavivirose. No presente estudo, o primeiro estudo de metagenômica viral realizado no Brasil, foi identificado um novo hepacivírus em PNH do Novo Mundo, o primeiro identificado em PNH do Novo Mundo.

2.4 Bioinformática

Os recentes avanços tecnológicos das técnicas de sequenciamento de nova geração permitiram aumento massivo no número de sequências desconhecidas, sem similaridade com as sequências atualmente disponíveis, trazendo juntamente novos desafios e dificuldades na área de bioinformática, que devem ser superados para completar o princípio da metagenômica (SCHOLZ et al., 2012).

Devido à grande diversidade dos genomas virais em banco de dados públicos, a maioria das sequências não tem similaridade significativa com nenhum gene conhecido. Para a maioria dos dados de metagenômica viral com genes longos, 30% ou menos das sequências apresentam similaridade com um gene previamente identificado, e muitos desses não estão associados com nenhuma função (SCHOENFELD et al., 2009). Além disso, os genomas do conjunto de dados são geralmente incompletos, com alguns casos onde há apenas pequena porção de fragmentos pequenos pertencentes a cada genoma (WOOLEY et al., 2010).

Normalmente, o fluxograma de análise de metagenomas começa pelo controle de qualidade e pré-processamento dos dados brutos. O principal objetivo é criar um banco de dados de alta qualidade que represente os genomas presentes na amostra e sua abundância. O controle de qualidade inclui o tamanho, o conteúdo GC, o escore de qualidade, o número de bases ambíguas “N” e a distribuição da complexidade das sequências (FANCELO et al., 2012).

A presença de sequências contaminantes (*primers* e material genômico de organismos que não vírus) nos metagenomas brutos tem relevância nos estudos de metagenômica viral. A filtração deve ser realizada para se obter apenas sequências de interesse e limitar montagens erradas, além de reduzir o tamanho do banco de dados que será analisado (FANCELLO et al., 2012).

A identificação taxonômica é um dos principais objetivos da metagenômica viral e pode ser considerada uma tarefa difícil, especialmente com as sequências curtas (50-500 nt) que podem ser menos informativas e dificultarem a classificação. A classificação taxonômica pode ser realizada com métodos de similaridade ou composição. Os métodos de similaridade são baseados em busca pelo BLAST e pelo fato da maior parte dos metagenomas serem compostos por micro-organismos desconhecidos, buscas com *E-value* estridente podem gerar poucas sequências classificáveis; por outro lado, buscas com *E-value* menos estridente podem resultar em classificações incorretas. Já os métodos de composição são baseados na composição nucleotídica, são computacionalmente mais rápidos e podem ser usados para a classificação de sequências altamente divergentes das disponíveis em bancos de dados, porém dependem no tamanho do fragmento e tem menor acurácia que os métodos baseados na similaridade (FANCELLO et al., 2012).

A montagem de dados metagenômicos é complexa devido a alguns fatores: a presença de diversos genomas, *contigs* não-espécie específicos, regiões genômicas conservadas que são compartilhadas por espécies distantemente relacionadas, a alta frequência de polimorfismos e variação genômica mesmo para subespécies, regiões repetidas e as diferentes taxas de cobertura entre as espécies devido à diferente abundância das espécies dentro de determinada amostra (FANCELLO et al., 2012).

Após a montagem, muitos pesquisadores utilizam programas como BLASTn e BLASTx para a busca em banco de dados na tentativa de encontrar homologias com vírus conhecidos. Embora estes métodos sejam utilizados para detectar vírus conhecidos, assim como vírus altamente divergentes, eles geralmente falham na descoberta de novos genomas virais, pois não mostram similaridade a nenhuma sequência depositada nos bancos de dados (BLOMSTRÖM, 2011). Estima-se que os métodos de busca baseados no

BLAST retornem entre 10-90% dos resultados como desconhecidos, devido principalmente ao pequeno tamanho das sequências produzidas pelo sequenciamento massivo, o fato dos bancos de dados não serem completos e a erros de sequenciamento (FANCELO et al., 2012).

A maioria dos dados metagenômicos inclui um número de sequências que não pode ser identificado por homologia em bancos de dados. Para se encontrar novos genomas virais, outras ferramentas devem ser desenvolvidas (BLOMSTRÖM, 2011). Os programas disponíveis atualmente permitem a comparação do metagenoma de interesse com outros genomas e metagenomas em diversos níveis, incluindo o gene, família da proteína, entre outros, e todos incluem variantes de ferramentas específicas ao metagenoma (KUNIN et al., 2008).

Muitos programas e plataformas tem sido desenvolvidos para auxiliar tanto na análise dos dados como na visualização dos resultados. Scholz et al. (2012) citam em sua revisão nove ferramentas que podem ser utilizadas para anotação e análise dos dados, dez para a montagem do metagenoma e onze para o mapeamento e alinhamento das sequências.

A maior dificuldade a ser sobreposta pelos estudos metagenômicos é o grande volume de dados gerados e a capacidade de análise destes dados, ou seja, o tempo de análise (KUNIN et al., 2008). Desta forma, novos programas de bioinformática devem ser desenvolvidos na tentativa de facilitar a comparação dos dados gerados, assim como para padronizar a forma de publicação, armazenamento, divulgação dos dados gerados e para analisar a estrutura e função das sequências encontradas.

No sentido de se padronizar a descrição das investigações genômicas, formou-se o *Genomic Standards Consortium* (GSC), que foi responsável pela divulgação de um guia contendo as informações mínimas que devem ser incluídas em publicação de sequência genômica (MIGS) (FIELD et al., 2008). Neste sentido, uma fonte de monitoramento centralizado de projetos de genoma e metagenômica foi fundado em 1997, *Genome OnLine Database* (GOLD). Em setembro de 2009, informações de mais de 5.800 projetos de sequência, incluindo 200 projetos metagenômicos, estavam inseridos no banco de dados. O projeto GOLD continua em expansão baseado no princípio de

fornecer informações de metadados de forma compreensível, relacionado ao MIGS (LIOLIOS et al., 2010).

3. OBJETIVOS

3.1 Padronizar um protocolo enzimático *in house* para a síntese da segunda fita de DNA de amostras virais.

3.2 Comparar o uso do protocolo completo do kit Nextera® XT e a combinação do kit Nextera® XT e kit Kapa Library Quantification para o preparo de bibliotecas para sequenciamento com o MiSeq.

3.3 Identificar novos genoas virais a partir de plasmas de primatas não humanos de diferentes regiões do Brasil por técnicas de metagenômica viral.

CAPÍTULO 2

4. ARTIGO CIENTÍFICO 1

Artigo científico formatado de acordo com as normas do *Journal of Virological Methods*.

Clinical Sample Preparation for NGS (Illumina MiSeq)

ULLMANN, L.S.; TOZATO, C.C.; MALOSSI, C.D.; CRUZ, T.F.; CAVALCANTE, R.V.; KURISSIO, J.K.; CAGNINI, D.Q.; RODRIGUES, M.V.; ARAÚJO JR, J.P.*

Laboratory of Animal and Human Virology, Department of Microbiology and Immunology, Biosciences Institute, UNESP – Univ Estadual Paulista, Botucatu, 18618-970, São Paulo, Brazil.

*Corresponding Author: Laboratory of Animal and Human Virology, Department of Microbiology and Immunology, Bioscience Institute, UNESP – Univ Estadual Paulista, Distrito de Rubião Junior, s/n, Botucatu, Sao Paulo, 80035-050, Brazil. Tel.: +55 (14) 3880-0422, E-mail: jpessoa@ibb.unesp.br.

HIGHLIGHTS

- Massive sequencing systems can be used for whole viral genome sequencing.
- dsDNA can be obtained with RNase H from cDNA and random primer for ssDNA viruses.
- Nextera[®] XT + Kapa is a valuable alternative for library preparation.

ABSTRACT

Sequence-independent methods for viral discovery have been widely used for whole genome sequencing of viruses. Different protocols for viral enrichment, library preparation and sequencing are available and the costs are decreasing. Accordingly, the aim of the present study was to evaluate an in house enzymatic protocol for double-stranded DNA (dsDNA) synthesis and also compare the use of full protocol Nextera[®] XT and the combined Nextera[®] XT and Kapa library quantification for MiSeq Sequencing. Two RNA viruses (canine distemper virus and dengue virus) and one ssDNA virus (porcine circovirus type 2) were tested with the developed protocols. The method for dsDNA synthesis employed showed satisfactory results and can be used on the laboratory, mainly if the enzymes used are already used in the lab. Library preparation with the combined kits (Nextera[®] XT and Kapa) yielded more reads and coverage of the genomes, probably because the normalization step of the Nextera[®] XT does not recover small fragments. Using Kapa quantification, libraries can be diluted or concentrated in order to increase the coverage.

Key words: double-stranded DNA, library preparation, massive sequencing, MiSeq.

Number of pages: 12; **Number of Tables:** 1

Viral metagenomics emerged as an important tool for virus discovery and whole genome sequencing (Barzon et al., 2011). More than 99% of viruses are considered non-cultivable and have high mutation rate (Thurber et al., 2009) enhancing the importance of non-sequence dependent detection methods such as massive sequencing.

A wide variety of protocols have been employed for viral enrichment from different sample types. Different library preparation methods can be used to increase the likelihood of the target nucleic acid detection. Advances in sequencing technology generate more data, and innumerable bioinformatics programs can be used to analyze the data generated.

Briefly, library preparation for Illumina MiSeq sequencing includes enzymatic tagmentation, adaptors and indexes addition, pooling, denaturation and sequencing. The sequencing-by-synthesis reversible terminator approach of the Illumina MiSeq sequencer uses single stranded DNA amplified by bridge for cluster formation. During the sequencing, the reversible terminator is imaged and modified to allow the incorporation of the next base (McEhloe et al., 2014).

Accordingly, the aim of the present study was to evaluate an in house method of double-strand DNA preparation from ssDNA virus and RNA viruses for viral massive sequencing with Illumina MiSeq system from different biological samples. Moreover, we described an association with Nextera[®] XT and quantitative PCR (qPCR) to enhance the number of reads per sample.

One cultured sample positive for PCV2, one cultured and one urine positive samples for canine distemper virus (CDV), and six positive serum samples for dengue virus were used to evaluate the efficiency of the method

used to obtain the dsDNA. Positive samples for CDV and dengue virus were also used to compare the use the Nextera® XT protocol and the combination of Nextera® XT and the qPCR library quantification.

Total RNA Purification Kit (Norgen, Thorold, Canada) was used to purify RNA from the samples, according to the manufacturer's instructions. The RNAs were then quantified by Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, USA) with the RNA Assay kit (Invitrogen, Carlsbad, USA). Samples with low concentration (including < 0.5 ng/mL - too low to be quantified) were concentrated in the Speed Vac Concentrator 5301 (Eppendorf, Hamburg, Germany) to a final volume of 10 µL. RNA samples were then treated with DNase I (Sigma-Aldrich, Saint Louis, USA). Following, the cDNA was obtained with the Improm-II™ RT kit (Promega, Fitchburg, USA), according to the manufacturer's instructions.

The DNA from the cultured PCV2 was purified with the Illustra Blood GenomicPrep Mini Spin Kit (Ge Healthcare, Little Chalfont, UK) and then quantified with Qubit® 2.0 Fluorometer with the DNA Assay kit.

Before the second strand of DNA was obtained, PCR was performed to verify the presence of the targets (PCV, CDV, and dengue virus) in the samples. A multiplex nested PCR was used to verify the positivity of the samples to DENV (Bronzoni et al., 2005) and quantitative PCR was used to detect CDV (Elia et al., 2006) and PCV2 (Ladekjaer-Mikkelsen et al., 2002).

After positive result in PCR, dsDNA was obtained with the following mix: T4 DNA Polymerase (5 U/µL) (Thermo Scientific, San Jose, USA), T4 DNA Ligase (5 U/µL), 10X T4 DNA Ligase Buffer (10 µL) (Thermo Scientific, San Jose, USA), dNTP mix (10 mM) (Promega, Madison, USA), 10 µL of the cDNA

and nuclease-free water qsp 100 μ L. RNase H (5 U/mL) (New England BioLabs, Ipswich, USA) was used for cDNA samples, since the enzyme cleavages the hybrid DNA and the ssRNA fragments can be used as primers, and random primer (250 ng/ μ L) for ssDNA samples. The reaction was incubated at 12 °C for 60 minutes and 22 °C for 60 minutes (modified from Okayama and Berg, 1982). Double-stranded DNA was then purified with Agencourt[®] AMPure[®] XP beads (Beckman Coulter Inc., Indianapolis, USA) and quantified using Qubit[®] 2.0 Fluorometer using the dsDNA High Sensitive kit (Invitrogen, Carlsbad, USA).

As the maximum concentration of 0.2 ng/ μ L is required as input dsDNA for Nextera[®] XT, samples with higher concentrations were diluted in order to fulfill the manufacturer's instructions. On the other hand, dsDNA samples with lower concentrations were concentrated either to 0.2 ng/ μ L or adjusted to 5 μ L, maximum volume that can be added in the reaction, to increase the initial DNA input and to improve sequencing coverage.

Samples were submitted to Nextera[®] XT (Illumina, Inc., San Diego, USA) to library preparation for deep sequencing. For this purpose, two different approaches were used. In the first one, full protocol of the Nextera[®] XT, with 3 μ L of the tagmentation enzyme, was used in three runs that were performed at the lab. In the second approach, Nextera[®] XT protocol was followed accordingly to the manufacturer's instructions until the third step (PCR Cleaning), also with 3 μ L of the tagmentation enzyme. After that, libraries were then quantified using Kapa library quantification, according to the manufacturer's instructions.

In order to increase the samples' coverage, before library pooling, samples with concentration lower than 1 nM were adjusted to 5 μ L. Libraries

with concentration higher than 1nM were diluted to 1 nM. The volume of 5 μ L of each library was pooled and concentrated to 4 nM (4x), then denatured accordingly to the Preparing DNA Libraries for Sequencing on the MiSeq® guide. Libraries were sequenced with MiSeq® System (Illumina, Inc., San Diego, USA) using MiSeq® Reagent kits (150 and 500 cycles).

Geneious R6 (Biomatters, Auckland, New Zealand) and CLC Genomics Workbench (CLC bio, Aarhus, Denmark) were used to analyze the sequences. The deposited reference genomes from NCBI used were NC_001921, NC_001477, DQ_923523, for CDV, dengue, and PCV2, respectively.

Samples of CDV (urine and culture) and Dengue virus (serum and culture) were prepared to massive sequencing by two different protocols: Nextera® XT full protocol and a combination of Nextera® XT and KAPA quantification kit. Porcine circovirus type 2 sample was only prepared by the combined Nextera® XT and KAPA quantification protocol (Table 1). The Kapa product was submitted to electrophoresis presenting the average size of 200 bp.

Deep sequencing generated from 28,026 to 865,212 reads per sample with Nextera® XT full protocol and ranged from 272,703 to 1,581,902 reads per sample with the combined Nextera® XT and KAPA quantification protocol. The combination of Nextera® XT and Kapa quantification resulted in more reads generated and more reads assembled to the reference genome. Also, cultured samples yielded more reads than the samples that were directly sequenced (Table 1).

The combined protocol resulted in 3 times more reads than Nextera® XT for CDV from urine, 2 times more reads for cultured CDV, and 11x more reads

for dengue virus. The number of assembled reads was 12x higher with the combined protocol for CDV from urine, about 5x higher for cultured CDV and about 4x higher for dengue virus (Table 1).

Table 1. Sample type, dsDNA concentration, library quantification total number of reads and reads assembled of CDV, DENV, and PCV2 using Illumina MiSeq Reagent kit.

Nextera-XT								
Sample	Type of Sample	dsDNA Quantification	Library Quantification (KAPA)	# of Cycles	# of reads	HQ%	# of reads Assembled	% Used Reads
CDV	Urine	< 0.5 ng/mL	-	2x250 cycles	109,228	99.6	5,264	4.819277108
CDV	Culture	< 0.5 ng/mL	-	2x250 cycles	865,212	99.7	139,349	16.10576367
Denv2	Serum	0.292ng/μL	-	2x150 cycles	153,178	98.6	8,766	5.72275392
Denv2	Serum	< 0.5 ng/mL	-	2x150 cycles	59,424	97.4	29,712	50
Denv4	Serum	< 0.5 ng/mL	-	2x150 cycles	28,026	99.6	2,573	9.180760722
Combination of Nextera-XT and KAPA								
Sample	Type of Sample	dsDNA Quantification	Library Quantification (KAPA)	# of Cycles	# of reads	HQ%	# of reads Assembled	% Used Reads
CDV	Urine	< 0.5 ng/mL	0.11 nM	2x250 cycles	348,361	99.9	65,728	18.86778371
CDV	Culture	< 0.5 ng/mL	12.4 nM	2x250 cycles	1,581,902	99.6	790,965	50.00088501
Denv1	Serum	< 0.5 ng/mL	6.64 nM	1x150 cycles	354,286	-	139,792	39.45738753
Denv1	Serum	< 0.5 ng/mL	0.20 nM	1x150 cycles	276,188	-	19,124	6.924268976
Denv1	Serum	< 0.5 ng/mL	0.43 nM	1x150 cycles	498,080	-	154,528	31.02473498
PCV2	Culture	< 0.5 ng/mL	5.71 nM	1x150 cycles	272,703	100	5,717	2.096419915

We used ssRNA and ssDNA viruses to set up a different method for dsDNA synthesis for further library preparation and massive sequencing with Illumina Inc. system. Double-stranded DNA was obtained with a mix containing RNase H to cleave the hybrid DNA from cDNA synthesis and the RNA fragments were used as primer for the second strand synthesis. RNase H enzyme has been used in the mix for dsDNA synthesis previously (Jones et al., 2005), but different ligase and polymerase were used.

Cultured sample from CDV yielded more reads than urine (Table 1) with both protocols probably due to higher viral load. Culture can be used prior to massive sequencing as a viral enrichment but it can influence on the sequence because mutations can occur during the culture process. Also, culture can select some genomes that can be more readily adapted to the cell type and this can influence on the diversity of the sequences obtained.

The combination of Nextera[®] XT and Kapa quantification generated more reads for all samples tested and different attempts can be made in order to increase the number of reads per run. Although the use of Kapa quantification could increase costs the gain compensates.

Due to the small size of the viral genomes we added 3 μ L of the tagmentation enzyme. In previous tests (data not showed) it was observed that the addition of the recommended volume of the tagmentation enzyme (5 μ L) and using the incubation period, resulted in smaller fragments, that could be lost during the bead purification step influencing final library quality. Another approach could be to keep the enzyme volume and to increase the amount of DNA input (McElhoe et al., 2014).

Purification after PCR with Agencourt[®] AMPure[®] beads can also decrease library quantification. The PCR clean-up is used in both protocols tested, it is a critical step since the attempts to circumvent the DNA loss were not satisfactory. In brief, the amount of beads (0.5x to 1.0x per reaction), ethanol concentration (80% to 70%), and final dry before eluting DNA (15 to 10 minutes) were modified without success (McElhoe et al., 2014).

Normalization during Nextera[®] XT is a step used to equalize the quantity of all libraries. Most of the samples tested had lower quantification than is detectable by dsDNA fluorescent dye quantitation (< 0.5 ng/mL) before library preparation, what could lead to a loss of fragments during normalization step reflecting on the cluster formation during sequencing. We used qPCR quantification after PCR clean up to increase the clusterable DNA, amplifying dsDNA and also ssDNA, considered target for bridge amplification, as recommended by McElhoe et al. (2014).

The usage of 1x150 or 2x250 cycles presented similar coverage since the DNA fragments had an average size of 200 bp. Furthermore, MiSeq[®] Reagent kit of 1x150 cycles was used on the last runs because there is a decrease in the quality of reads in the end of the fragments and also from read 1 and read 2, when pair-ended sequencing is performed (McElhoe et al., 2014).

The method of dsDNA preparation from ssDNA (PCV2) also showed satisfactory results, because the addition of random primers enabled the formation of dsDNA. Thus, this methodology can be used to another ssDNA viruses that are present in samples with high viral load. Furthermore, the sample preparation for sequencing without the amplification step conserves the diversity present in the original sample compared with the pre-amplified product

by PCR or RCA (Rolling Circle Amplification). Therefore, preserving the diversity of the sample is important for the determination of quasispecies by next-generation sequencing.

Double-stranded DNA mix containing RNase H or random primer, T4 DNA Polymerase and T4 DNA Ligase successfully produced dsDNA from ssRNA and dsDNA viruses from different samples. Nextera[®] XT and Kapa quantification combined yielded more reads per sample with higher coverage of the viruses sequenced than the Nextera[®] XT full protocol.

Acknowledgements

Funding for Dr. Leila Ullmann's research fellowship was provided by the State of São Paulo Research Foundation (FAPESP).

References

- Barzon, L., Lavezzo, E., Militello, V., Toppo, S., Palù, G. 2011. Application of next-generation sequencing technologies to diagnostic virology. *International J. Mol. Sci.* 12, 7861-7884.
- Bronzoni, R.V.M., Baleotti, F.G., Nogueira, R.M.R., Nunes, M., Figueiredo, L.T. 2005. Duplex Reverse Transcription-PCR followed by Nested PCR assays for detection and identification of Brazilian Alphaviruses and Flaviviruses. *J. Clin. Microbiol.* 43, 2, 696-702.
- Elia, G., Decaro, N., Martella, V., Cirone, F., Lucente, M.S., Lorusso, E., Trani, L., Buonavoglia, C. 2006. Detection of canine distemper virus in dogs by real-time RT-PCR. *J. Virol. Methods* 36, 1-2, 171-6.

Jones, M.S., Kapoor, A., Lukashov, V.V., Simmonds, P., Hecht, F., Delwart, E. 2005. New DNA Viruses Identified in Patients with Acute Viral Infection Syndrome. *J. Virol.* 79, 13, 8230-8236.

Ladekjaer-Mikkelsen, A.S., Nielsen, J., Stadejek, T., Storgaard, T., Krakowka, S., Ellis, J., McNeilly, F., Allan, G., Botner, A., 2002. Reproduction of postweaning multisystemic wasting syndrome (PMWS) in immunostimulated and non-immunostimulated 3-week-old piglets experimentally infected with porcine circovirus type 2 (PCV2). *Vet. Microbiol.* 89, 2-3, 97-114.

McElhoe, J.A., Holland, M.M., Makova, K.D., Su, M.S-W., Paul, I.M., Baker, C.H., Faith, S.A., Young, B. 2014. Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. *Forensic Sci Int Genet* (in press).

Okayama, H., BERG, P. 1982. High-efficiency cloning of full-length cDNA. *Mol. Cell Biol.* 2, 161-70.

Thurber, V.B., Haynes, M., Breitbart, M., Wegley, L., Rowher, F. 2009. Laboratory procedures to generate viral metagenomes. *Nature* 4, 4, 470-479.

CAPÍTULO 3

5. ARTIGO CIENTÍFICO 2

Artigo científico formatado de acordo com as normas do *Journal of Virology*.

VIRAL METAGENOMICS FROM NEW WORLD NON-HUMAN PRIMATES

Ullmann, L.S.¹; Li, L.²; Cubas, Z.S.³; Morais, W.³; Dias-Neto, R.⁴, Soares, L.⁴, Muniz, JAPC⁵, Castro, P.⁵; Fantinatti, B.⁶; Deng, X.²; Pusterla, N.⁷; Leutenegger, C.⁸; Biondo, A.W.⁹; Delwart, R.²; Araujo Jr., J.P.¹

¹Laboratory of Animal and Human Virology, Department of Microbiology and Immunology, Biosciences Institute, UNESP – Univ Estadual Paulista, Botucatu, 18618-970, São Paulo, Brazil.

²Blood Systems Research Institute, University of California, San Francisco, CA, USA, 94118.

³Bela Vista Biological Sanctuary, Itaipu Binacional Hydroelectric Power Plant, Foz do Iguaçu, PR, Brazil, 85866-900.

⁴ Tietê Ecological Park, São Paulo, SP, Brazil, 03719-000.

⁵ National Primate Center, Ananindeua, PA, Brazil, 67020-380.

⁶Integrative Genomic Laboratory, Department of Morphology, Biosciences Institute, UNESP – Univ Estadual Paulista, Botucatu, 18618-970, São Paulo, Brazil.

⁷Department of Veterinary Medicine and Epidemiology, University of California, Davis, CA, USA, 95616

⁸Idexx Reference Laboratories, West Sacramento, CA, USA, 95605.

⁹Department of Veterinary Medicine, Federal University of Parana, Curitiba, PR, Brazil, 80035-050.

ABSTRACT

Wild animals-borne emerging zoonoses represent the most significant and growing threat to global health. Non-human primates contribute with 20% of the main human diseases, considered the main experimental model and also can

be considered sentinels for different diseases that affect humans. Accordingly the aim of the present study was to identify new viral genomes in plasma from wild-caught non-human primates from different parts of Brazil and from four species: squirrel monkey (*Saimiri sciuratus*), night monkey (*Aotus aenulatus*), Howley monkey (*Alouatta caraya*), and capuchin (*Sapajus apella*), considered clinically healthy during sampling. Deep sequencing was performed with MiSeq® System and a bioinformatics pipeline was developed to analyze the sequences obtained. Preliminary results show that a wide variety of viral genomes were present in the samples highlighting torque tenus virus, hepatitis G virus types A and B, parvovirus, and papillomavirus.

Key words: viral metagenomics, deep sequencing, non-human primates.

Viral metagenomics emerged as an important tool for viral discovery mainly after the advances on sequence-independent techniques known as high-throughput sequencing. Traditional methods such as cell culture, electron microscopy and inoculation in experimental animal models are based on comparison with known viruses; molecular biology including PCR and Sanger sequencing are also based on known sequences and although some viruses were discovered by the use of these techniques, viral metagenomics circumvent the need of a previous known sequence, allows the discovery of new viral genomes within complex samples and are exponentially increasing (1).

Zoonotic pathogens contribute with about 60% of the emerging infectious diseases (EID) and 71.8% are wildlife-borne. The variety of wildlife species by itself is an EID predictor, independently of human population, rainfall or latitude (2). Brazil is considered a hotspot for EID emergence due to the broad variety of wildlife species variety and the large territory that the country occupies in South America.

A project involving the ecology, health and conservation of primates has been conducted at the Kibale National Park, Uganda. A high variety and density of NHP is found at the park and the animals have contact with humans and domestic animals present at the park. Viral metagenomics was used to identify simian hemorrhagic fever virus (SHFV) in red colobus (*Procolobus rufomitratus*

tephrosceles) (3) and African red-tailed guenons (*Cercopithecus ascanius*) (4), *guereza* hepatitis virus (GHV) (5) and simian immunodeficiency virus (SIV) (6) were identified in black-and-white colobus (*Colobus guereza*), another variant of SIV was identified in red-tailed guenons (7), and pegiviruses were identified in red colobus, red-tailed guenons and olive baboon (*Papio anubis*) (8). Viral metagenomics studies identified other variants of SIV (9-11) and simian foamy virus (10) in red colobus.

Due to the phylogenetically proximity of humans and non-human primates (NHP) and the diversity of NHP inhabiting the Brazilian territory viral metagenomics were used to identify new viral genomes in New World NHP species from Brazil.

Ninety-six samples from four species of NHP were collected in the National Primate Center (NPC) (Belem, PA) from 52 squirrel monkey (*Saimiri sciureus*) and 24 night monkey (*Aotus ainfulatus*), Tiete Ecological Park (TEP) (São Paulo, SP) from Howley monkey (*Alouatta caraya*), and Bela Vista Biological Sanctuary (BVBS) (Foz do Iguaçu, PR) from 3 Howley monkey and 12 capuchin (*Sapajus apella*) (Table 1). The animals were clinical healthy during sampling and were manually and chemically restrained for whole blood sampling.

Table 1. Blood plasma pooling from New World non-human primates accordingly to the place and time of sampling.

Sample Pooling	Place/Sampling Date	Sampled Species
Sample 1	NPC/June 2011	<i>Saimiri sciureus</i>
Sample 3	NPC/June 2011	<i>Saimiri sciureus</i>
Sample 8	NPC/June 2011	<i>Saimiri sciureus</i>
Sample 9	TEP/December 2011	<i>Alouatta caraya</i>
Sample 10	NPC/November 2011	<i>Saimiri sciureus</i> and <i>Aotus infulatus</i>
Sample 11	NPC/November 2011	<i>Saimiri sciureus</i> and <i>Aotus infulatus</i>
Sample 20	NPC/November 2011	<i>Saimiri sciureus</i> and <i>Aotus infulatus</i>
Sample 21	NPC/November 2011	<i>Saimiri sciureus</i> and <i>Aotus infulatus</i>
Sample 22	NPC/November 2011	<i>Saimiri sciureus</i> and <i>Aotus infulatus</i>
Sample 23	BVBS/ May 2012	<i>Alouatta caraya</i> and <i>Sapajus apella</i>

Primate plasma samples (300 μ L each) were centrifuged at 12,000 $\times$ g for 5 minutes at 8°C and the supernatants were filtered through a 0.45 μ M filter (Merck Millipore, Darmstadt, Germany) to remove possible host cellular debris and bacteria. The filtrates were treated with a nuclease mixture of 14U turbo DNase (Ambion, Thermo Fisher Scientific, Waltham, USA), 3U Baseline-ZERO (Epicentre, Illumina, San Diego, USA) and 20U RNase One (Promega, Madison, USA) in 1x DNase buffer (Ambion) at 37°C for 1.5 hour to reduce background nucleic acids from the host cells and bacteria.

Viral nucleic acids protected by viral capsids, were then extracted from ~200 μ L resulting solutions by Maxwell 16 Viral Total Nucleic Acid Purification Kit (Promega) following the manufacture's instruction. Viral cDNA synthesis was performed by incubation of 10 μ L extracted viral nucleic acids with 100 pmol of a primer containing a fixed 20 bp sequence plus a random nonamer at the 3' end at 85°C for 2 min. Then, 200U SuperScript III reverse transcriptase (Invitrogen – Thermo Fisher Scientific), 0.5mM of each deoxynucleoside triphosphate (dNTP), 10mM dithiothreitol, and 1x first-strand extension buffer were added to the mixture and incubated at 25°C for 10 min, followed by 50°C incubation for 1 hour. The 2nd strand DNA synthesis was performed by incubation with 50 pmol of random primer at 95°C for 2 min, 4 °C for 2 min, and then with 5U Klenow Fragment (New England Biolabs, Beverly, MA, USA) at 37°C for 1 hour. The resulting products were either put into library preparation directly or PCR amplified by using 5 μ L of the RT-Klenow DNA products and 2.5 μ M primer consisting of the fixed 18 bp portion of the random primer with 1U AmpliTaq Gold DNA polymerase (Life Technologies - Thermo Fisher Scientific), 2.5mM MgCl₂, 0.2mM dNTPs, and 1 $\times$ PCR Gold buffer in a reaction volume of 50 μ L. Temperature cycling was performed as follows: 1 cycle of 95°C for 5 min, 30 cycles of denaturing at 95°C for 30 sec, 55°C for 30 sec, 72°C for 1.5 min. An additional extension for 10 min at 72°C was added to the end of the run.

The PCR products were purified twice by Agencourt AMPure XP beads (Beckman Coulter, Villepinte, France) with 0.8:1 ratio of beads to sample. Purified products were checked by agarose gel electrophoresis and Nanodrop spectrophotometer (Thermo Scientific). The pre-amplification products were

used as input DNA for TruSeq library preparation following the manufacture's instruction. The quality of the libraries was assessed by 2100 Bioanalyzer (Agilent Technologies, ADS Agilent, Santa Rosa, CA, USA) and the quantification was estimated by KAPA Library Quant Kit (Kapa Biosystems, Woburn, MA, USA) following the manufacturer's instructions. Libraries were pooled and sequenced with MiSeq® v2 Reagent kit (2x250 cycles) with MiSeq® System (Illumina®, San Diego, USA).

The resulting libraries of single-stranded DNA fragments were sequenced using the MiSeq Illumina platform and in two multiplexed runs using 2x250 cycle MiSeq Reagent Kit v2 (Illumina). Paired-end reads of 250 bp generated by MiSeq were demultiplexed using vendor software from Illumina. A virus discovery pipeline running on a 32-nodes Linux cluster was developed to process the data. Human host reads and bacterial reads were subtracted by mapping the reads to human reference genome hg19 and bacterial RefSeq genomes release 59 using bowtie2. Reads were considered duplicates if 5bp to 55bp from 5 prime end was identical. One random copy of duplicates was kept. Low sequencing quality tails were trimmed using Phred quality score 10 as the threshold. Adaptor and primer sequences were trimmed using the default parameters of VecScreen. The filtered reads were *de novo* assembled using a *de novo* sequence assembler SOAPdenovo2, ABySS, meta-Velvet and CAP3. The assembled contigs, along with singlets were aligned to a viral proteome database (consisting of all annotated full or near full viral genomes) using BLASTx. The significant hits to virus were then aligned to a non-virus-non-redundant (NVNR) universal proteome database using BLASTx. Hits with more significant E-value to NVNR than to virus were removed. A web-based graphical user interface was developed to present users with the virus hits, along with taxonomy information and processing meta-information.

Different viral families were identified on each sample pool sequenced (Figure 1). Overall, *Anelloviridae* was the family with more viral hits among all the samples and only on the pools 9, 10, 11, and 23 *Flaviviridae*, *HERV*, *Flaviviridae*, and *Parvoviridae* families had more hits than *Anelloviridae*. The most important virus species and number of hits are presented on Table 2.

Table 2. Viral genomes hits per plasma pool sequenced from New World non-human primates and the number of hits per virus.

Sample Pools	Viral Sequences Hits (Number of hits)
Sample 1	TTV(10000+), GBV A (12), GBV B (2), GBV C (1), Parvovirus (5), Chipmunk Parvovirus (7), Circovirus (36), Ebola Virus (2)
Sample 3	TTV (10000+), GBV A (3) GBV B (2), Parvovirus (66), Simian-Human Immunodeficiency Virus (9)
Sample 8	TTV (10000+), GBV A (10), GBV B (5), Papillomavirus (21), Parvovirus (96), Picornaviridae (2)
Sample 9	TTV (4500+) , GBV A (10000+), GBV C (23), Parvovirus (75), Papillomavirus (14)
Sample 10	TTV(10000+), GBV A (753), GBV B (2), Parvovirus (57), Rhadinovirus (16), Circovirus (50)
Sample 11	TTV (49), GBV A (3000), GBV B (2), Parvovirus (15)
Sample 20	TTV (1000+), GBV A (722), papillomavirus (7), Parvovirus (22)
Sample 21	TTV (2000+), GBV A (8), GBV B (2), Parvovirus (21)
Sample 22	TTV (30000+), GBV A (1), Parvovirus (37)
Sample 23	TTV(2000+) , GBV A (3), GBV B (3669), Parvovirus (10000+)

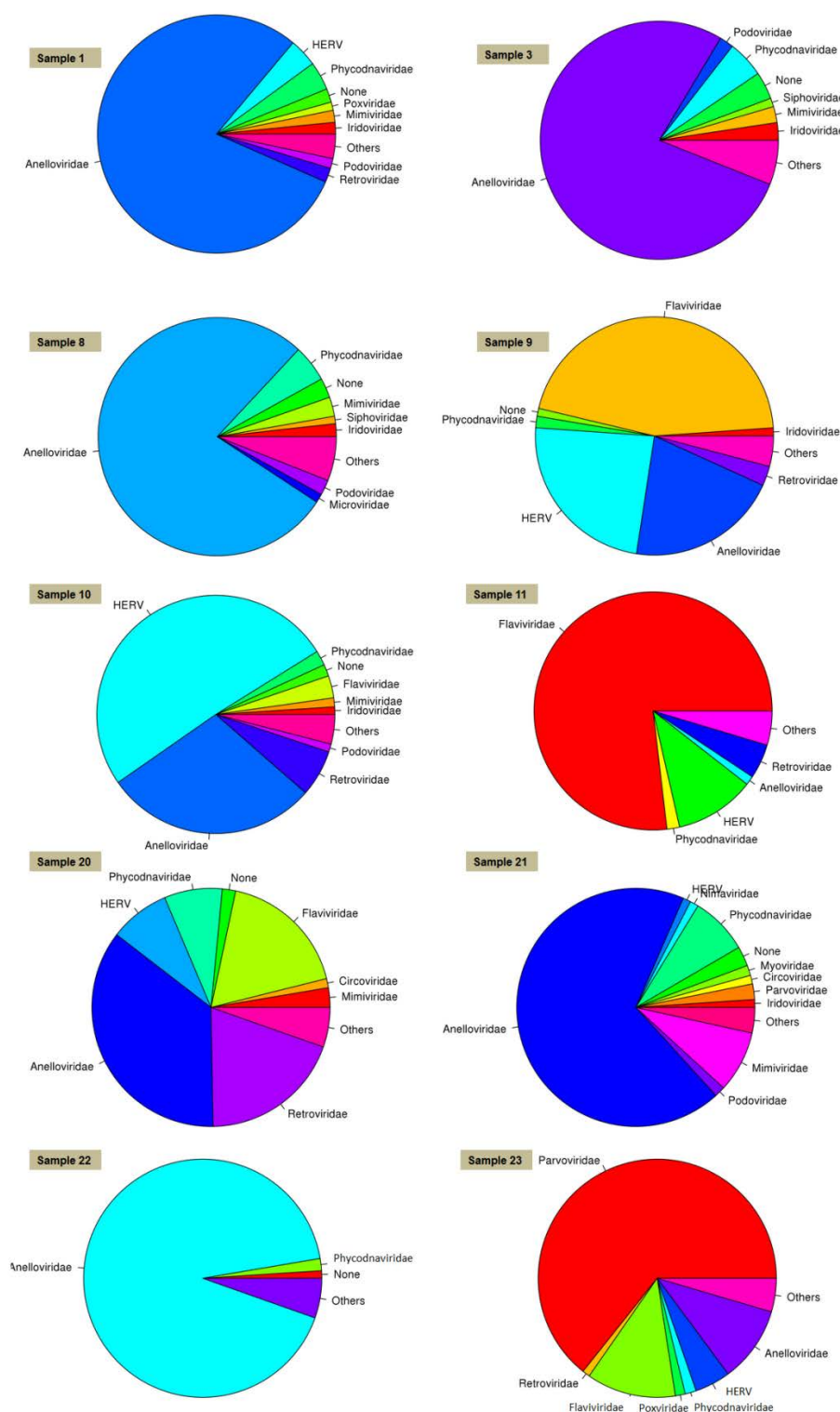


Figure 1. Graphical representation of the contig numbers of viral families identified on each plasma pool sequenced from New World non-human primates.

Hits for Torque teno virus TTV were found in all pools sequenced with a high number of hits in all pools but Sample 11. TTV is a small, nonenveloped virus with circular, single-stranded DNA genome classified within the *Anellovirus* genus. Many TTV sequences were identified in different species of NHP such as chimpanzee (*Pan troglodytes*), Japanese macaque (*Macaca fuscata*), cotton-top tamarin (*Saguinus oedipus*), douroucoulis (*Aotus trivirgatus*), and tupaia (*Tupaia belangeri chinensis*) and also in domestic animals and human. A wide variety of divergent sequences have been identified classified either in about 39 genotypes or five phylogenetic groups (12).

Reads that match with *Flaviviridae* family were found in all pools sequenced. Hepatitis G virus B (GBV-B) that belongs to the *Hepacivirus* genus, the same as Hepatitis C virus (HCV), was identified in six pools (Table 2). A new hepacivirus (*guereza* Hepacivirus) was identified in a black-and-white colobus (5), demonstrating the virus is present in both New and Old World monkeys.

Hepatitis G virus types A and C (GBV-A and GBV-C) that are within the *Pegivirus* genus were identified in all the pools. Three new variants of pegiviruses were recently identified in red colobus, red-tailed guenons and olive baboon from Kibale National Park, Uganda (8).

Other interesting viruses were found as parvovirus, papillomavirus, circovirus and simian-human immunodeficiency virus and their sequences are under analysis in attempt to obtain the whole-genome sequence. A new hepacivirus was identified on Sample 23 and both nucleotide and amino acid sequences are under analysis for full characterization accordingly to the International Committee on Taxonomy of Viruses (ICTV).

The majority of the studies are conducted with Old World monkeys, mainly resulting in new SIV variants, and few is known about the New World monkeys, which enhances the importance of more investigation about the viruses harbored by these animals. This is the first viral metagenomics study conducted with NHP from Brazil and although the animals sampled were from captive and considered healthy when sampled, the results can serve for comparison with other viral metagenomics studies from both sick and healthy animals.

ACKNOWLEDGEMENTS

Funding for Dr. Leila Ullmann's research fellowship was provided by the São Paulo State Research Foundation (FAPESP), Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), and IDEXX Laboratories. We kindly thank the help given by the Bela Vista Sanctuary and Itaipu Binacional, National Primate Center and Tiete Ecological Park personal for the sample collection and storage.

REFERENCES

1. **Temmam S, Davoust B, Berenger J-M, Raoult D, Desnues C.** 2014. Viral metagenomics on animals as a tool for the detection of zoonoses prior to human infection? *Int. J. Mol. Sci.* 15:10377-10397.
2. **Jones KE, Patel NG, Levy MA, Storeugard A, Balk D, Gittleman JL, Daszak P.** 2008. Global trends in emerging infectious diseases. *Nature.* 451(21):990-994.
3. **Lauck M, Hyeroba D, Tumukunde A, Weny G, Lank SM, Chapman CA, O'Connor DH, Friedrich TC, Goldberg TL.** 2011. Novel, divergent simian hemorrhagic fever viruses in a wild Ugandan red colobus monkey discovered using direct pyrosequencing. *PLoS One.* 6(4):19056.
4. **Lauck M, Sibley SD, Hyeroba D, Tumukunde A, Weny G, Chapman CA, Ting N, Switzer WM, Kuhn JH, Friedrich TC, O'Connor DH, Goldberg TL.** 2013. Exceptional simian hemorrhagic fever virus diversity in a wild African Primate Community. *J. Virol.* 87(1): 688-691.
5. **Lauck M, Sibley SD, Lara J, Purdy MA, Khudyakov Y, Hyeroba D, Tumukunde A, Weny G, Switzer WM, Chapman CA, Hughes AL, Friedrich TC, O'Connor DH, Goldberg TL.** 2013. A novel Hepacivirus with an unusually long and intrinsically disordered NS5A protein in a wild Old World primate. *J. Virol.* 87(6):8971-8981.
6. **Lauck M, Switzer WM, Sibley SD, Hyeroba D, Tumukunde A, Weny G, Taylor B, Shankar A, Ting N, Chapman CA, Friedrich TC, Goldberg TL, O'Connor DH.** 2013. Discovery and full genome characterization of two highly divergent simian immunodeficiency viruses infecting black-and-white colobus monkeys (*Colobus guereza*) in Kibale National Park, Uganda. *Retrovirol.* 10:107.
7. **Lauck M, Switzer WM, Sibley SD, Hyeroba D, Tumukunde A, Weny G, Shankar A, Greene JM, Ericson AJ, Zheng H, Ting N, Chapman CA, Friedrich TC, Goldberg TL.** 2014. Discovery and full genome

characterization of a new SIV lineage infecting red-tailed guenons (*Cercopithecus ascanius schmidtii*) in Kibale National Park, Uganda. *Retrovirol.* 11:55.

8. **Sibley SD, Lauck M, Bailey AL, Hyeroba D, Tumunkunde A, Weny G, Chapman CA, O'Connor DH, Goldberg TL, Friedrich TC.** 2014. Discovery and Characterization of Distinct Simian Pegiviruses in Three Wild African Old World Monkey Species. *PLoS One.* 9(6):98569.
9. **Locatelli S, Liegeois F, Lafay B, Roeder AD, Bruford MW, Formenty P, Noe R, Delaporte E, Peeters M.** 2008. Prevalence and genetic diversity of simian immunodeficiency virus infection in wild-living red colobus monkeys (*Piliocolobus badius badius*) from the Taï forest, Côte d'Ivoire SIVwrc in wild-living western red colobus monkeys. *Infection Gen. Evol.* 8:1-14.
10. **Goldberg TL, Sintasath DM, Chapman CA, Cameron KM, Karesh WB, Tang S, Wolfe ND, Rwegom IB, Ting N, Switzer WM.** 2009. Coinfection of Ugandan Red Colobus (*Procolobus [Piliocolobus] rufomitratu tephrosceles*) with Novel, Divergent Delta-, Lenti-, and Spumaretroviruses. *J. Virol.* 83(21):11318-11329.
11. **Leendertz SAJ, Junglen S, Hedemann C, Goffe A, Calvignac S, Boesch C, Leendertz FH.** 2010. High Prevalence, Coinfection Rate, and Genetic Diversity of Retroviruses in Wild Red Colobus Monkeys (*Piliocolobus badius badius*) in Taï National Park, Côte d'Ivoire. *J. Virol.* 84(15):7427-7436.
12. **Okamoto H.** 2009. TT viruses in animals. *Curr. Trop. Microbiol. Immunol.* 331:35-52.

CAPÍTULO 4

6. ARTIGO CIENTÍFICO 3

Artigo científico formatado de acordo com as normas do *Journal of Virology*.

NEW HEPACIVIRUS IN NON-HUMAN PRIMATES FROM SOUTHERN BRAZIL

Ullmann, L.S.¹; Li, L.²; Cubas, Z.S.³; Morais, W.³; Fantinatti, B.⁴; Deng, X.²; Pusterla, N.⁵; Leutenegger, C.⁶; Biondo, A.W.⁷; Delwart, R.²; Araujo Jr., J.P.¹

¹Laboratory of Animal and Human Virology, Department of Microbiology and Immunology, Biosciences Institute, UNESP – Univ Estadual Paulista, Botucatu, 18618-970, São Paulo, Brazil.

²Blood Systems Research Institute, University of California, San Francisco, CA, USA, 94118.

³Bela Vista Biological Sanctuary, Itaipu Binacional Hydroelectric Power Plant, Foz do Iguaçu, PR, Brazil, 85866-900.

⁴Integrative Genomic Laboratory, Department of Morphology, Biosciences Institute, UNESP – Univ Estadual Paulista, Botucatu, 18618-970, São Paulo, Brazil.

⁵Department of Veterinary Medicine and Epidemiology, University of California, Davis, CA, USA, 95616

⁶Idexx Reference Laboratories, West Sacramento, CA, USA, 95605.

⁷Department of Veterinary Medicine, Federal University of Parana, Curitiba, PR, Brazil, 80035-050.

ABSTRACT

Flavivirus, *Pestivirus*, *Hepacivirus*, and *Pegivirus* genera belong to the *Flaviviridae* family. Hepatitis C virus and Hepatitis G virus (GBV-B) are within the *Hepacivirus* genus, with variants identified in non-human primates, wild rodents, dogs, and bats. Brazil is considered a sanctuary of flaviviruses and important diseases such as dengue and yellow fever occur throughout the country. As non-human primates contribute to important human diseases, the

study of the viruses present in these animals represents an important tool for detection of possible emergent viruses. Thus, viral metagenomics was performed with plasma samples from 10 non-human primates (*Alouatta guariba* and *Sapajus apella*) from Southern Brazil. Libraries were pooled and sequencing with MiSeq® System with 2x250 cycles reagent kit. Geneious R7 and Mega 6.0 were used for analysis. Out of 2,181,312 generated reads, 3,819 had hit to GBV-B (NC001655) and were used for de novo assembly, resulting in 10 contigs. One contig had 8,885 bp and was used as reference for concatenation with the whole raw data from the pool sequenced, resulting in a consensus sequence of 9,708 bp (HQ 99.1%). Amino acid analysis resulted on the viral polyprotein with a CDS of 2,830 aa. The sequence showed 63% identity with 39 gaps compared with the deposited sequence (BAK24073). Phylogenetic analysis with neighbor-joining tree (1,000 rounds of bootstrap) resulted in a monophyletic group that shares ancestry with GBV-B. Results suggest a new species of Hepacivirus similar to GBV-V.

Key words: viral metagenomics, hepacivirus, non-human primates.

Zoonotic pathogens contribute with 60.3% of the emerging infectious diseases (EID) and 71.8% are originated from wildlife. Brazil is considered a hotspot for the EID due to the wide diversity of wildlife species, what is per se, an important predictor for a disease emergence, independently of the human population growth, latitude and rainfall (1). Non-human primates (NHP) are the closest phylogenetically animals to humans and contribute with 20% of human diseases (2) and act as sentinels and experimental models for human diseases.

Hepatitis G virus B (GBV-B) was first described in tamarins (*Saguinus* spp.) and other New World monkeys after the inoculation of an acute-phase plasma from a physician with unexplained hepatitis (3). GBV-B and hepatitis C virus compound the *Hepacivirus* genus, *Flaviviridae* family. Genera *Pestivirus*, *Flavivirus* and *Pegivirus* (recently included) also belong to this viral family (4).

Non-primate hepaciviruses were identified in dogs (5), horses (6), rodents (7), and bats (8). A divergent and similar to hepatitis G virus was

identified in a black-and-white colobus (*Colobus guereza*), denominated *guereza* Hepacivirus (GHV) (9).

Accordingly, a new hepacivirus was identified in captive New World monkeys (*Alouatta guariba* and *Sapajus apella*), Southern Brazil. The virus, non-human primate hepacivirus (NHP-HV), shares common ancestry with GBV-B. Although Amazon region is considered the more likely region for viral emergency, NHP-HV was discovered in NHP from Southern Brazil, what highlights the importance of studies that aim to identify new viruses that can emerge and disseminate in wildlife as well human population.

Three howley monkeys (*Alouatta guariba*) and seven capuchins (*Sapajus apella*) were sampled at the Bela Vista Biological Sanctuary, Foz do Iguacu, Paraná, Brazil. After physical and chemical restraintment, whole blood was collected with EDTA, plasma was obtained by centrifugation and kept under refrigeration during the transportation to the laboratory.

Primate plasma samples (300 µL each) were centrifuged at 12,000xg for 5 minutes at 8°C and the supernatants were filtered through a 0.45 µm filter (Merck Millipore, Darmstadt, Germany) to remove possible host cellular debris and bacteria. The filtrates were treated with a nuclease mixture of 14U turbo DNase (Ambion, Thermo Fisher Scientific, Waltham, MA, USA), 3U Baseline-ZERO (Epicentre, Illumina, San Diego, USA) and 20U RNase One (Promega, Madison, USA) in 1x DNase buffer (Ambion) at 37°C for 1.5 hour to reduce background nucleic acids from the host cells and bacteria.

Viral nucleic acids protected by viral capsids, were then extracted from ~200 µL resulting solutions by Maxwell 16 Viral Total Nucleic Acid Purification Kit (Promega) following the manufacture's instruction. Viral cDNA synthesis was performed by incubation of 10 µL extracted viral nucleic acids with 100 pmol of a primer containing a fixed 20 bp sequence plus a random nonamer at the 3' end at 85°C for 2 min. Then, 200U SuperScript III reverse transcriptase (Invitrogen – Thermo Fisher Scientific), 0.5mM of each deoxynucleoside triphosphate (dNTP), 10mM dithiothreitol, and 1x first-strand extension buffer were added to the mixture and incubated at 25°C for 10 min, followed by 50°C incubation for 1 hour. The 2nd strand DNA synthesis was performed by incubation with 50 pmol of random primer at 95°C for 2 min, 4 °C for 2 min, and

then with 5U Klenow Fragment (New England Biolabs, Beverly, USA) at 37°C for 1 hour. The resulting products were either put into library preparation directly or PCR amplified by using 5 µL of the RT-Klenow DNA products and 2.5 µM primer consisting of the fixed 18 bp portion of the random primer with 1U AmpliTaq Gold DNA polymerase (Life Technologies - Thermo Fisher Scientific), 2.5mM MgCl₂, 0.2mM dNTPs, and 1× PCR Gold buffer in a reaction volume of 50 µL. Temperature cycling was performed as follows: 1 cycle of 95°C for 5 min, 30 cycles of denaturing at 95°C for 30 sec, 55°C for 30 sec, 72°C for 1.5 min. An additional extension for 10 min at 72°C was added to the end of the run.

The PCR products were purified twice by Agencourt AMPure XP beads (Beckman Coulter, Villepinte, France) with 0.8:1 ratio of beads to sample. Purified products were checked by agarose gel electrophoresis and Nanodrop spectrophotometer (Thermo Scientific). The pre-amplification products were used as input DNA for TruSeq library preparation following the manufacture's instruction. The quality of the libraries was assessed by 2100 Bioanalyzer (Agilent Technologies, ADS Agilent, Santa Rosa, USA) and the quantification was estimated by KAPA Library Quant Kit (Kapa Biosystems, Woburn, USA) following the manufacturer's instructions. Libraries were pooled and sequenced with MiSeq® v2 Reagent kit (2x250 cycles) with MiSeq® System (Illumina®, San Diego, USA).

The resulting libraries of single-stranded DNA fragments were sequenced using the MiSeq Illumina platform and in two multiplexed runs using 2x250 cycle MiSeq Reagent Kit v2 (Illumina). Paired-end reads of 250 bp generated by MiSeq were demultiplexed using vendor software from Illumina. A virus discovery pipeline running on a 32-nodes Linux cluster was developed to process the data. Human host reads and bacterial reads were subtracted by mapping the reads to human reference genome hg19 and bacterial RefSeq genomes release 59 using bowtie2. Reads were considered duplicates if 5bp to 55bp from 5 prime end was identical. One random copy of duplicates was kept. Low sequencing quality tails were trimmed using Phred quality score 10 as the threshold. Adaptor and primer sequences were trimmed using the default parameters of VecScreen. The filtered reads were *de novo* assembled using

a *de novo* sequence assembler SOAPdenovo2, ABySS, meta-Velvet and CAP3. The assembled contigs, along with singlets were aligned to a viral proteome database (consisting of all annotated full or near full viral genomes) using BLASTx. The significant hits to virus were then aligned to a non-virus-non-redundant (NVNR) universal proteome database using BLASTx. Hits with more significant E-value to NVNR than to virus were removed. A web-based graphical user interface was developed to present users with the virus hits, along with taxonomy information and processing meta-information.

Geneious R7 was used to generate the contigs and the polyprotein and MEGA 6.0 was used for phylogenetic analysis. Polyproteins from 26 viruses representing the four genera from *Flaviviridae* family, as well as one poliovirus and one togavirus (outgroup), available on GenBank, were aligned with ClustalW. The evolutionary history was inferred by using the Maximum Likelihood method based (10, 11).

Deep sequencing from plasma of ten New World monkeys resulted in 2,181,312 reads with 3,819 similar to GBV-B (NC_001655). De novo assembly of the 3,819 reads resulted in 10 contigs and one had 8,885 bp in length assembled from 3,690 reads. The consensus of this contig was used as reference for concatenation with raw data from the pool sequenced, resulting in a consensus sequence of 9,708 bp (HQ 99.1%). This consensus was then searched with BLASTx and CDS was identified by comparing the amino acids from the sequences that matched with those from the NHP-HV, resulting in a polyprotein of 2,830 aa. The sequence showed 63% identity with 39 gaps compared with the deposited sequence (BAK24073).

Phylogenetic analysis constructed from the polyprotein alignment demonstrated relationships among *Flaviviridae* and a common ancestral with GVB-B (AF179612.1, NP056931.1, and BAK24073.1), although the polyproteins were in different branches within the tree (Figure 1).

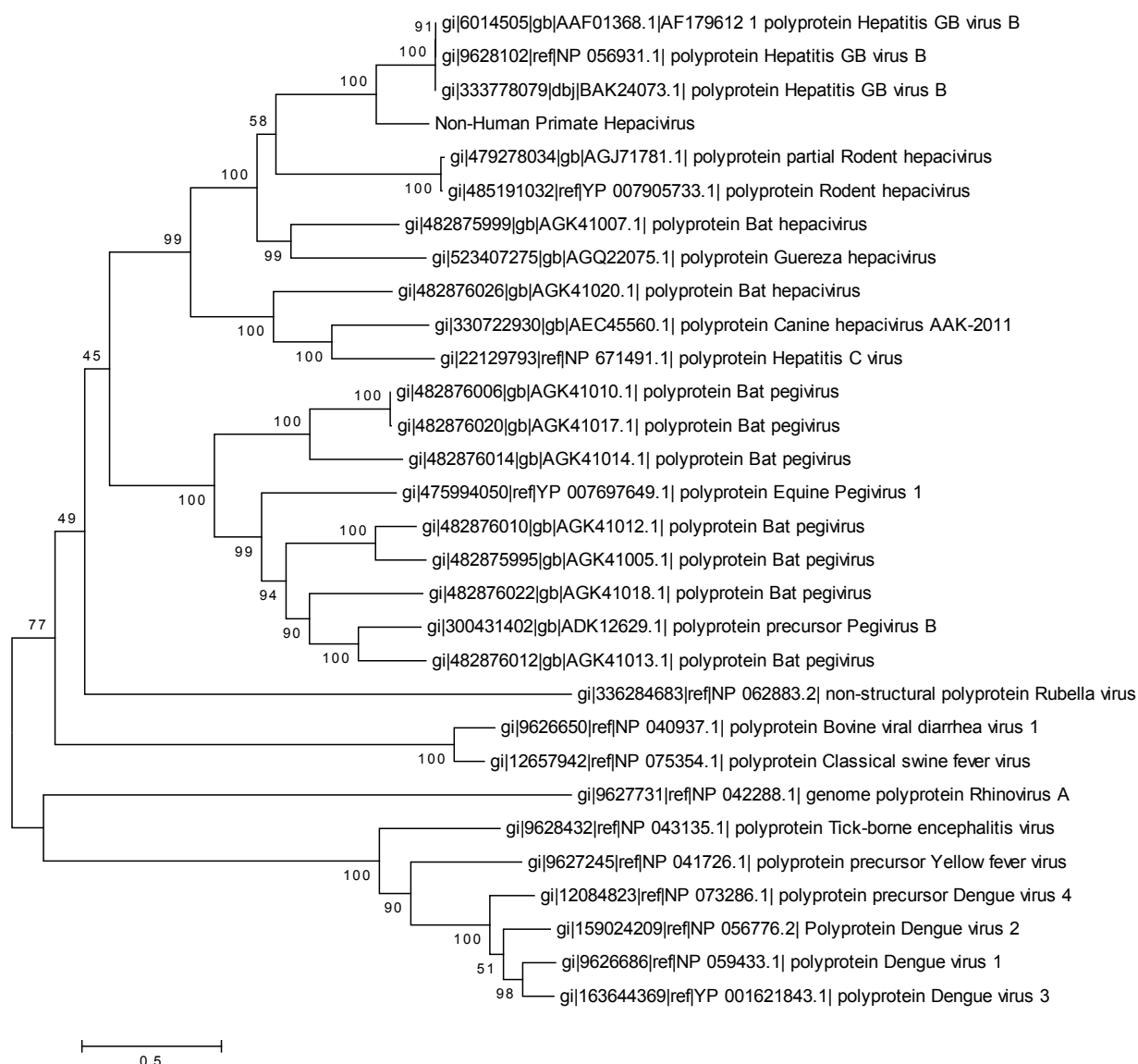


Figure 1. Molecular phylogenetic analysis by Maximum Likelihood method. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The tree with the highest log likelihood (-69585.1847) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 30 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 1438 positions in the final dataset.

Hepatitis G virus type A belongs to *Pegivirus* genus from *Flaviviridae* family and was detected in Old World monkeys (12). Although GBV-B was recovered from tamarins after inoculation (13), the virus was not recovered from wild-caught New World NHP until now (12).

Although the genome was sequenced as a sample pool from two different NHP species, howley monkey and capuchin, they were healthy during sampling, suggesting the virus is adapted to the host, as well as the black-and-white colobus (*Colobus guereza*) that hosts *guereza* Hepacivirus (9). As tamarins that were experimentally infected developed acute but self-resolving subacute hepatitis (14) the monkeys from the present study could have developed a self-resolving hepatitis when they were first infected even before they were caught from wildlife.

The NHP-HV detected in the present study represents the first hepacivirus detected in naturally infected wild-caught New World monkeys and *Guereza* hepacivirus was the first hepacivirus detected in naturally infected Old World monkeys (9). Recently, new hepacivirus were identified in dogs (5), horses (6), rodents (7), bats (8), and Old World monkeys (9), demonstrating the wide variety of hosts this viruses infect. In the present study we identified a new hepacivirus that infect other species of New World monkeys.

This is the first hepacivirus identified in natural infected wild-caught New World monkeys from Brazil. The present study is part of a viral metagenomics study with samples from other New World monkeys. Hits for GBV-B-like virus were detected in squirrel monkeys (*Saimiri sciuratus*) and night monkeys (*Aotus aenulatus*) from the National Primate Center, North Brazil (data been processed).

Studies involving samples from a wider variety of New World primates' species should be encouraged in an attempt to identify the natural host of hepacivirus and other possible viruses from *Flaviviridae* family.

ACKNOWLEDGEMENTS

Funding for Dr. Leila Ullmann's research fellowship was provided by the São Paulo State Research Foundation (FAPESP), Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), and IDEXX Laboratories. We kindly thank the help given by the Bela Vista Sanctuary and Itaipu Binacional personal for the sample collection and storage.

REFERENCES

1. **Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P.** 2008. Global trends in emerging infectious diseases. *Nature*. 451(21):990-994.
2. **Wolfe ND, Dunavan CP, Diamond J.** Origins of major human infectious diseases. *Nature*. 447(17):279-283.
3. **Deinhardt F, Holmes A, Capps R, Popper H.** 1967. Studies on the transmission of human viral hepatitis to marmoset monkeys. I. Transmission of disease, serial passages, and description of liver lesions. *J. Exp. Med.* 125:673-688.
4. **Stapleton J, Fong S, Muerhoff A, Bukh J, Simmonds P.** 2011. The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus *Pegivirus* within the family *Flaviviridae*. *J. Gen. Virol.* 92:233-246.
5. **Kapoor A, Simmonds P, Gerold G, Qaisar N, Jain K, Henriquez JA, Firth C, Hirschberg DL, Rice CM, Shields S, Lipkin WI.** 2011. Characterization of a canine homolog of hepatitis C virus. *PNAS*. 108(28):11608-11613.
6. **Kapoor A, Simmonds P, Cullen JM, Scheel TKH, Medina JL, Giannitti F, Nishiuchi E, Brock KV, Burbelo PD, Rice SM, Lipkin WI.** 2013. Identification of a Pegivirus (GB Virus-Like Virus) that infects horses. *J. Virol.* 87(12):7185-7190.
7. **Kapoor A, Simmonds P, Scheel TKH, Hjelle B, Cullen JM, Burbelo PD, Chauhan LV, Duraisamy R, Leon MS, Jain K, Vandegrift KJ, Calisher CH, Rice CM, Lipkin WI.** 2013. Identification of rodent homologs of Hepatitis C virus and Pegiviruses. *mBio*. 4(2):00216-13.
8. **Quan P-L, Firth C, Conte JM, Williams SH, Zambrana-Torrel CM, Anthonya SJ, Ellison JA, Gilbert AT, Kuzmin IV, Niezgoda M, Osinubic MOV, Recuenco S, Markotter W, Breiman RF, Kalemba L, Malekani J, Lindblade KA, Rostal MK, Ojeda-Flores R, Suzan R, Davis LB, Blau DM, Ogunkoyak AB, Castillo DAA, Moran D, Ngam S, Akaibe D, Agwanda B, Briese T, Epstein JH, Daszak P, Rupprecht CE, Holmes EC, Lipkin WI.** 2013. Bats are a major natural reservoir for hepaciviruses and pegiviruses. *PNAS*. 110(20):8194-8199.

9. **Lauck M, Sibley SD, Lara J, Purdy MA, Khudyakov Y, Hyeroba D, Tumukunde A, Weny G, Switzer WM, Chapman CA, Hughes AL, Friedrich TC, O'Connor DH, Goldberg TL.** 2013. A novel Hepacivirus with an unusually long and intrinsically disordered NS5A protein in a wild Old World primate. *J. Virol.* 87(6):8971-8981.
10. **Jones DT, Taylor WR, Thornton JM.** 1992. The rapid generation of mutation data matrices from protein sequences. *CABIOS.* 8:275-282.
11. **Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S.** 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30:2725-2729.
12. **Adams NJ, Prescott LE, Jarvis LM, Lewis JCM, McClure MO, Smith DB, Simmonds P.** 1998. Detection in chimpanzees of a novel flavivirus related to GB virus-C/hepatitis G virus. *J. Gen. Virol.* 79:1871-1877.
13. **Bukh J, Appgar CL, Govindarajan S, Purcell RH.** 2001. Host Range Studies of GB Virus-B Hepatitis Agent, the Closest Relative of Hepatitis C Virus, in New World Monkeys and Chimpanzees. *J. Med. Virol.* 65:694-697.
14. **Akari H, Iwasaki Y, Yoshida T, Iijima S.** 2009. Non-human primate surrogate model of hepatitis C virus infection. *Microbiol. Immunol.* 53:53-57.

CAPÍTULO 5

7. CONCLUSÃO

De acordo com os resultados obtidos no presente estudo, pode-se concluir:

1. O protocolo enzimático utilizado para a obtenção da segunda fita de DNA foi eficiente e pode ser empregado no preparo de bibliotecas a partir de diferentes amostras biológicas.

2. A combinação do Nextera® XT e Kapa Library Quantification resultou em mais sequências virais que a utilização do protocolo completo do Nextera® XT.

3. Diversas sequências de ácidos nucleicos virais foram geradas pela metagenômica viral a partir das amostras de primatas não-humanos do Novo Mundo sequenciadas.

4. Foi identificado um novo hepacivírus em amostras de PNH do Novo Mundo que pode ser caracterizado como uma nova espécie.

8. REFERÊNCIAS BIBLIOGRÁFICAS

ACEVEDO, A.; ANDINO, R. Library preparation for highly accurate sequencing of RNA viruses. *Nature Prot.*, v.9, n.7, p.1760-1769, 2014.

ALFARO, J.W.L.; SOUSA E SILVA JR, J.; RYLANDS, A. How different are robust and gracile Capuchin monkeys? An argument for the use of *Sapajus* and *Cebus*. *Am. J. Primatol.*, v.74, p.276-286, 2012.

ANDRADE, A.; ANDRADE, M.C.R.; MARINHO, A.M.; FERREIRA-FILHO, J. *Biologia, Manejo e Medicina de Primatas Não Humanos na Pesquisa Biomédica*. Rio de Janeiro, RJ: Editora Fiocruz, 2010, 472p.

AURICCHIO, P. *Primatas do Brasil*. Terra Brasilis Edit., São Paulo, 1995, 198p.

BEDIN, F. Le Brésil, une terre d'élection pour les arboviroses? *Med. Trop.*, v.67, p.281-287, 2007.

BEXFIELD, N.; KELLAM, P. Metagenomics and the molecular identification of novel virus. *Vet. J.*, v.190, n.2, p.191-198, 2011.

BIERS, E.J.; SUN, S.; HOWARD, E.C. Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. *App. Environ. Microbiol.*, v.75, p.2221-2229, 2009.

BLOMSTRÖM, A.-L. Viral metagenomics as an emerging and powerful tool in veterinary medicine. *Vet. Quarterly*, v.31, n.3, p.107-114, 2011.

BLOMSTRÖM, A.-L.; WIDÉN, F.; HAMMER, A.-S.; BELÁK, S.; BERG, M. Detection of a novel astrovirus in brain tissue of mink suffering from shaking mink syndrome by use of viral metagenomics. *J. Clin. Microbiol.*, v. 48, n.12, p.4392-4396, 2010.

BREITBART, M; HEWSON, I.; FELTS, B.; MAHAFFY, J.M.; NULTON, J.; SALAMON, P.; ROHWER, F.; Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.*, v.85, p.6220-6223, 2003.

BRENIG, B.; BECK, J.; SCHÜTZ, E. Shotgun metagenomics of biological stains using ultra-deep DNA sequencing. *Forensic Sci. Internat.: Genetics*, v.4, p.228-231, 2010.

CASTRIGNANO, S.B.; NAGASSE-SUGAHARA, T.K.; KISIELIUSB, J.J.; UEDA-ITOB, M.; BRANDAO, P.E.; CURTI, S.P. Two novel circo-like viruses detected in human feces: complete genome sequencing and electron microscopy analysis. *Virus Res.*, v.178, p.364-373, 2013.

CHANDRIANI, S.; SKEWES-COX, P.; ZHONG, W.; GANEM, D.E.; DIVERS, T.J.; VAN BLARICUM, A.J.; TENNANT, B.C.; KISTLER, A.L. Identification of a previously underscribed divergent virus from the Flaviviridae family in an outbreak of equine serum hepatitis. *Proc. Natl. Acad. Sci.*, p.1-9, 2013.

DALY, G.M.; BEXFIELD, N.; HEANEY, J.; STUBBS, S.; MAYER, A.P.; PALSER, A.; KELLAM, P.; DROU, N.; CACCAMO, M.; TILEY, L.; ALEXANDER, G.J.M.; BERNAL, W.; HEENEY, J.L. A Viral Discovery Methodology for Clinical Biopsy Samples Utilising Massively Parallel Next Generation. *PLoS One*, v.6, n.12, p.28879, 2011.

DELA-CRUZ, F.N.; GIANTINI, F.; LI, L.; WOODS, L.W.; DEL VALLE, L.; DELWART, E.; PESAVENTO, P.A. Novel polyomavirus associated with brain tumors in free-ranging raccoons, Western United States. *Emerg. Infect. Dis.*, v.19, n.1, p.77-84, 2013.

DELWART, E.R. Viral metagenomics. *Rev. Med. Virol.*, v.17, p.115-131, 2007.

DINSDALE, E.A.; EDWARDS, R.A.; HALL, D.; ANGLY, F.; BREITBART, M.; BRULE, J.M.; FURLAN, M.; DESNUES, C.; HAYNES, M.; LI, L.; McDANIEL, L.; MORAN, M.A.; NELSON, K.E.; NILSSON, C.; OLSON, R.; PAUL, J.; BRITO, B.R.; RUAN, Y.; SWAN, B.K.; STEVENS, R.; VALENTINE, D.L.; THURBER, R.V.; WEGLEY, L.; WHITE, B.A. ROHWER, F. Functional metagenomic profiling of nine biomes. *Nature*, v.452, p.629-632, 2008.

FANCELLO, L.; RAOULT, D.; DESNUES, C. Computational tools for viral metagenomics and their application in clinical research. *Virology*, v.434, p.162-174, 2012.

FIELD, D.; GARRITY, G.M.; SANSONE, S.A.; STERK, P.; GRAY, T.; KYRPIDES, N.; HIRSCHMAN, L.; GLÖCKNER, F.O.; KOTTMANN, R.; ANGIULI, S.; WHITE, O.; DAWYNDT, P.; THOMSON, N.; GIL, I.S.; MORRISON, N.; TATUSOVA, T.; MIZRACHI, I.; VAUGHAN, R.; COCHRANE, G.; KAGAN, L.; MURPHY, S.; SCHRIML, L. GENOMIC STANDARDS CONSORTIUM. Meeting report: the fifth Genomic Standards Consortium (GSC) workshop. *Omics*, v.12, n.3, p.109-113, 2008.

GOLDBERG, T.L.; SINTASATH, D.M.; CHAPMAN, C.A.; CAMERON, K.M.; KARESH, W.B.; TANG, S.; WOLFE, N.D.; RWEGOM I.B.; TING, N.; SWITZER, W.M. Coinfection of Ugandan Red Colobus (*Procolobus [Piliocolobus] rufomitratu tephrosceles*) with Novel, Divergent Delta-, Lenti-, and Spumaretroviruses. *J. Virol.*, v.83, n.21, p.11318-11329, 2009.

GREENSILL, J.; SHELDON, J.A.; RENWICK, N.M.; BEER, B.E.; NORLEY, S.; GOUDSMIT, J.; SCHULZ, T.F. Two distinct gamma-2 Herpesvirus in African

Green monkeys: a second gamma-2 herpesvirus lineage among Old World Primates? *J. Virol.*, v.74, n.3, p.1572-1577, 2000.

HOFFMANN, B.; SCHEUCH, M.; HÖPER, D.; JUNGBLUT, R.; HOLSTEG, M.; SCHIRRMAYER, H.; ESCHBAUMER, M.; GOLLER, K.V.; WERNIKE, K.; FISCHER, M.; BREITHAUPT, A.; METTENLEITER, T.C.; BEER, M. Novel Orthobunyavirus in cattle, Europe, 2011. *Emerg. Infect. Dis.*, v.18, n.3, p.469-472, 2012.

HONKAVOURI, K.S.; SHIVAPRASAD, H.L.; BRIESE, T.; STREET, C.; HIRSCHBERG, D.L.; HUTCHISON, S.K.; LIPKIN, W.I. Novel picornavirus in turkey poult with hepatitis, California, USA. *Emerg. Infect. Dis.*, v.17, n.3, p.480-487, 2011.

HONKAVOURI, K.S.; SHIVAPRASAD, H.L.; WILLIAMS, B.; QUAN, P.-L.; HORNIG, M.; STREET, C.; PALACIOS, G.; HUTCHINSON, S.K.; FRANCA, M.; EGHOLM, M.; BRIESE, T.; LIPKIN, W.I. Novel borna virus in Psittacine birds with proventricular dilatation disease. *Emerg. Infect. Dis.*, v.14, n.12, p.1883-1886, 2008.

JONES, K.E.; PATEL, N.G.; LEVY, M.A.; STOREUGARD, A.; BALK, D.; GITTLEMAN, J.L.; DASZAK, P. Global trends in emerging infectious diseases. *Nature*, v.451, n.21, p.990-994, 2008.

JONES, M.S.; KAPOOR, A.; LUKASHOV, V.V.; SIMMONDS, P.; HECHT, F.; DELWART, E. New DNA viruses identified in patients with acute viral infection syndrome. *J. Virol.*, v.79, n.13, p.8230-8236, 2005.

KINDLOVITS, A.; KINDLOVITS, L.M. *Clinica e Terapeutica em Primatas Neotropicaes*. Rio de Janeiro: L. F. Livros, 2ed., 2009, 535p.

KING, J.; LaRUE, B.L.; NOVROSKI, N.M.; STOLJAROVA, M.; SEO, S.B.; ZENG, X.; WARSHAUER, D.H.; DAVIS, C.P.; PARSON, W.; SAJANTILA, A.; BUDOWLE, B. High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic Sci. Internat.: Genetics*, v.12, p.128-135, 2014.

KRUPOVIC, M.; BAMFORD, D.H. Revealing virus-host interplay. *Science* v.333, p.45-46, 2011.

KUNIN, V.; COPELAND, A.; LAPIDUS, A.; MAVROMATIS, K.; HUGENHOLTZ, P. A Bioinformatic's guide to metagenômico. *Microbiol. Mol. Biol. Rev.*, v.71, n.4, p.557-578, 2008.

LAUCK, M.; HYEROBA, D.; TUMUKUNDE, A.; WENY, G.; LANK, S.M.; CHAPMAN, C.A.; O'CONNOR, D.H.; FRIEDRICH, T.C.; GOLDBERG, T.L.

Novel, divergent simian hemorrhagic fever viruses in a wild Ugandan red colobus monkey discovered using direct pyrosequencing. PLoS One, v.6, n.4, p.19056, 2011.

LAUCK, M.; SIBLEY, S.D.; HYEROBA, D.; TUMUKUNDE, A.; WENY, G.; CHAPMAN, C.A.; TING, N.; SWITZER, W.M.; KUHN, J.H.; FRIEDRICH, T.C.; O'CONNOR, D.H.; GOLDBERG, T.L. Exceptional simian hemorrhagic fever virus diversity in a wild African Primate Community. J. Virol., v.87, n.1, p.688-691, 2013a.

LAUCK, M.; SIBLEY, S.D.; LARA, J.; PURDY, M.A.; KHUDYAKOV, Y.; HYEROBA, D.; TUMUKUNDE, A.; WENY, G.; SWITZER, W.M.; CHAPMAN, C.A.; HUGHES, A.L.; FRIEDRICH, T.C.; O'CONNOR, D.H.; GOLDBERG, T.L. A novel Hepacivirus with an unusually long and intrinsically disordered NS5A protein in a wild Old World primate. J. Virol., v.87, n.6, p.8971-8981, 2013b.

LAUCK, M.; SWITZER, W.M.; SIBLEY, S.D.; HYEROBA, D.; TUMUKUNDE, A.; WENY, G.; TAYLOR, B.; SHANKAR, A.; TING, N.; CHAPMAN, C.A.; FRIEDRICH, T.C.; GOLDBERG, T.L.; O'CONNOR, D.H. Discovery and full genome characterization of two highly divergent simian immunodeficiency viruses infecting black-and-white colobus monkeys (*Colobus guereza*) in Kibale National Park, Uganda. Retrovirol., v.10, p.107, 2013c.

LAUCK, M.; SWITZER, W.M.; SIBLEY, S.D.; HYEROBA, D.; TUMUKUNDE, A.; WENY, G.; SHANKAR, A.; GREENE, J.M.; ERICSEN, A.J.; ZHENG, H.; TING, N.; CHAPMAN, C.A.; FRIEDRICH, T.C.; GOLDBERG, T.L. Discovery and full genome characterization of a new SIV lineage infecting red-tailed guenons (*Cercopithecus ascanius schmidtii*) in Kibale National Park, Uganda. Retrovirol., v.11, p.55, 2014.

LAW, J.; JOVEL, J.; PATTERSON, J.; FORD, G.; O'KEEFE, S.; WANG, W.; MENG, B.; SONG, D.; ZHANG, Y.; TIAN, Z.; WASILENKO, S.T.; RAHBARI, M.; MITCHELL, T.; JORDAN, T.; CARPENTER, E.; MASON, A.L.; WONG, G.K-S. Identification of hepatotropic viruses from plasma using deep sequencing: a next generation diagnostic tool. PLoS One, v.8, n.4, p.60595, 2013.

LEENDERTZ, S.A.J.; JUNGLEN, S.; HEDEMANN, C.; GOFFE, A.; CALVIGNAC, S.; BOESCH, C.; LEENDERTZ, F.H. High Prevalence, Coinfection Rate, and Genetic Diversity of Retroviruses in Wild Red Colobus Monkeys (*Piliocolobus badius badius*) in Taï National Park, Côte d'Ivoire. J. Virol., v.84, n.15, p.7427-7436, 2010.

LIOLIOS, K.; CHEN, I.-M.A.; MAVROMATIS, K.; TAVERNARAKIS, N.; HUGENHOLTZ, P.; MARKOWITZ, V.M.; KYRPIDES, N.C. The Genome On

Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their association metadata. Nucl. Ac. Res., v.38, p.346-353, 2010.

LOCATELLI, S.; LIEGEOIS, F.; LAFAY, B.; ROEDER, A.D.; BRUFORD, M.W.; FORMENTY, P.; NOE, R.; DELAPORTE, E.; PEETERS, M. Prevalence and genetic diversity of simian immunodeficiency virus infection in wild-living red colobus monkeys (*Piliocolobus badius badius*) from the Tai forest, Côte d'Ivoire SIVwrc in wild-living western red colobus monkeys. Infection Gen. Evol., v.8, p.1-14, 2008.

McELHOE, J.A.; HOLLAND, M.M.; MAKOVA, S., M.S.-W.; PAUL, I.M.; BAKER, C.H.; FAITH, S.A.; YOUNG, B. Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. Forensic Sci. Internat.: Genetics, 2014 (in press).

MUNNINK, B.B.O.; FARSANI, S.M.J.; DEIJS, M.; JONKERS, J.; VERHOEVEN, J.T.P.; IEVEN, M.; GOOSSENS, H.; DE JONG, M.D; BERKHOUT, B.; LOENS, K.; KELLAM, P.; BAKKER, M.; CANUTI, M.; COTTEN, M.; VAN DER HOEK, L. Antologous antibody capture to enrich immunogenic viruses for viral discovery. PLoS One, v.8, n.11, p.785454, 2013.

NELSON, K.E.; WEINSTOCK, G.M.; HIGHLANDER, S.K. et al. A catalog of reference genomes from the human microbiome. Science, v.262, p.994-999, 2010.

NG, T.F.F.; MANIRE, C.; BORROWMAN, K.; LANGER, T.; EHRHART, L.; BREITBART, M. Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. J. Virol., v.83, n.6, p.2500-2509, 2009.

NG, T.F.F.; SUEDEMEYER, W.K.; WHEELER, E.; GULLAND, F.; BREITBART, M. Novel anellovirus discovered from a mortality event of captive California sea lions. J. Gen. Virol., v.90, p.1256-1261, 2009.

NG, T.F.F.; WHEELER, E.; GREIG, D.; WALTZEK, T.B.; GULLAND, F.; BREITBART, M. Metagenomic identification of a novel anellovirus in Pacific harbor seal (*Phoca vitulina richardsii*) lung samples and its detection in samples from multiple years. J. Gen. Virol., v.92, p.1318-1323, 2011.

REIS, N.R.; PERACCHI, A.L.; PEDRO, W.A.; LIMA, I.P. Mamíferos do Brasil. Londrina, PR, Londrina: Nelio dos Reis, 2011, 439p.

ROSE, T.M.; STRAND, K.B.; SCHULTS, E.R.; SCHAEFER, G.; RANKIN, G.W.; THOULES, M.E. Identification of two homologs of the Kaposi's Sarcoma-Associated Herpesvirus (Human Herpesvirus 8) in retroperitoneal fibromatosis of different macaque species. J. Virol., v.71, n.5, p.4138-4144, 1997.

RUSCH, D.B.; HALPEM, A.L.; SUTTON, G. et al. The *Sorcerer II* global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol., v.7, 2007.

SCHOENFELD, T.; LILES, M.; WOMMACK.; POLSON, S.W.; GODISKA, R.; MEAD, D. Functional viral metagenomics and the next generation of molecular tools. Trends Microbiol., v.18, n.1, p.20-29, 2009.

SCHOLZ, M.B.; LO, C.-C.; CHAIN, P.S.G. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. Curr. Opin. Biotech., v.23, n.9, p.9-15, 2012.

SIBLEY, S.D.; LAUCK, M.; BAILEY, A.L.; HYEROBA, D.; TUMUNKUNDE, A.; WENY, G.; CHAPMAN, C.A.; O'CONNOR, D.H.; GOLDBERG, T.L.; FRIEDRICH, T.C. Discovery and Characterization of Distinct Simian Pegiviruses in Three Wild African Old World Monkey Species. PLoS One, v.9, n.6, p.98569, 2014.

SIMON, C.; DANIEL, R. Achievements and new knowledge unraveled by metagenomic approaches. Appl. Microbiol. Biotechnol., v.85, p.265-276, 2009.

SINGH, J.; BEHAL, A.; SINGLA, N.; JOSHI, A.; BIRBIAN, N.; SINGH, S.; BALI, V.; BATRA, N. Metagenomics: concept, methodology, ecological inference and recent advances. Biotechnol. J., v.4, p.480-494, 2009.

STENGLEIN, M.D.; SANDERS, C.; KISTLER, A.L.; RUBY, J.G.; FRANCO, J.Y.; REAVILL, D.R.; DUNKER, F.; DeRISI, J.L. Identification, characterization, and *in vitro* culture of highly divergent arenaviruses from boa constrictors and annulated tree boas: candidate etiological agents for snake inclusion body disease. mBio, v.3, n.4, p.00180-12, 2012.

SVOBODA, W.K. Vigilância de epizootias em primatas não humanos (PNH) como instrumento de monitoramento de arbovirose e outras viroses de interesse em saúde pública. 135p. Tese (Doutorado em Ciência Animal, área de concentração em Sanidade Animal) – Universidade Estadual de Londrina, Londrina-PR, 2007.

THOMAS, T.; Gilbert, J.; MEYER, F. Metagenomics - a guide from sampling to data analysis. Microbial Inform. Experiment. v.2, n.3, p.1-12, 2012.

THURBER, R.V.; HAYNES, M.; BREITBART, M.; WEGLEY, L.; ROWHER, F. Laboratory procedures to generate viral metagenomes. Nat. Prot., v.4, n.4, p.470-483, 2009.

TSENG, M.; FLEETWOOD, M.; REED, A.; GILL, V.A.; HARRIS, R.K.; MOELLER, R.B.; LIPSCOMB, T.P.; MAZET, J.A.K.; GOLDSTEIN, T. Mustelid

herpesvirus-2, a novel herpes infection in Northern sea otters (*Enhydra Lutris Kenyoni*). J. Wildl. Dis., v.48, n.1, p.181-185, 2012.

VERONA, C.E.S.; PISSINATTI, A. Primates – Primatas do Novo Mundo (Sagui, Macaco-Prego, Macaco-Aranha, Bugio). In: In: CUBAS, Z.S.; SILVA, J.C.R.; CATÃO-DIAS, J.L. Tratado de Animais Selvagens – Medicina Veterinária. Roca: São Paulo, 2006, 358-377p.

WOLFE, N.D.; DUNAVAN, C.P.; DIAMON, J. Origins of major human infectious diseases. Nature, v.447, n.17, p.279-283, 2007.

WOOLEY, J.C.; GODZIK, A.; FRIEDBERG, I. A primer on metagenomics. PLoS Comput. Biol., v.6, n.2, 2010.