U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

LOYOLA UNIVERSITY HEALTH SYSTEM

Stritch School of Medicine

Insights into the European rabbit immune system:

evolutionary history, diversity and expression of genes playing key roles in Myxoma virus infection and host-resistance

Ana Cristina Lemos de Matos

Tese de Doutoramento apresentada à
Faculdade de Ciências da Universidade do Porto

Programa Doutoral em Biodiversidade, Genética e Evolução

2014

Insights into the European rabbit immune system:
evolutionary history, diversity and expression of genes playing key roles in Myxoma virus infection and host-resistance

Ana Cristina Lemos de Matos
Programa Doutoral em Biodiversidade, Genética e Evolução
Departamento de Biologia
2014

Orientador
Pedro José Esteves, Ph.D., Investigador Principal
CIBIO/InBio Laboratório Associado, Universidade do Porto

Coorientador
Dennis Lanning, Ph.D., Research Assistant Professor
Department of Microbiology and Immunology
Stritch School of Medicine, Loyola University Chicago

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR
COMPETE
PROGRAMA OPERACIONAL FACTORES DE COMPETITIVIDADE
QREN QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL PORTUGAL 2007.2013
UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

Dissertação apresentada à Faculdade de Ciências da Universidade do Porto para a obtenção do grau de Doutor em Biodiversidade, Genética e Evolução.

À minha Mãe

Ao Mário

Declaração

Na elaboração desta dissertação, e nos termos do nº 2 do Artigo 8º do Decreto-lei nº 388/70 da República Portuguesa, os resultados de trabalhos já publicados foram totalmente aproveitados e fazem parte integrante de alguns Capítulos desta dissertação. Em todos estes trabalhos, o candidato participou na obtenção, interpretação, análise e discussão dos resultados e na elaboração das suas formas publicadas.

A Faculdade de Ciências da Universidade do Porto foi a instituição de origem do candidato, tendo o trabalho sido realizado sob orientação do Doutor Pedro José Esteves, Investigador do CIBIO - Centro de Investigação em Biodiversidade e Recursos Genéticos/InBio Laboratório Associado, e coorientação do Doutor Dennis Lanning, Investigador e Professor Assistente no Department of Microbiology and Immunology, Stritch School of Medicine, Loyola University Chicago, USA.

Este trabalho foi apoiado pela Fundação para a Ciência e a Tecnologia através da atribuição de uma bolsa de doutoramento de referência SFRH / BD / 48566 / 2008.

Statement

In this dissertation, and in compliance with no. 2 of article 8 of law no. 388/70 of the Portuguese Republic, the results of published works were totally used and included in some of the Chapters of this dissertation. In all these works, the candidate participated in obtaining, interpreting, analyzing and discussing the results and writing the published forms.

The *Faculdade de Ciências da Universidade do Porto* was the home institution of the candidate, and the work was directed by Dr. Pedro José Esteves, Researcher at *CIBIO - Centro de Investigação em Biodiversidade e Recursos Genéticos/InBio Laboratório Associado*, and co-directed by Dr. Dennis Lanning, Research Assistant Professor at the Department of Microbiology and Immunology, Stritch School of Medicine, Loyola University Chicago, USA.

This work was financially supported by the *Fundação para a Ciência e a Tecnologi*a PhD grant with reference SFRH / BD / 48566 / 2008.

Agradecimentos/Acknowledgements

Aos que, eventualmente, poder-me-ei ter esquecido de mencionar, as minhas mais sinceras desculpas./ To those that I may have forgot to mention, my sincere apologies.

Ao meu orientador, Doutor Pedro José Esteves começo por agradecer a oportunidade que me proporcionou em fazer parte integrante do seu grupo de investigação. Muito obrigada pela confiança depositada em mim e nas minhas capacidades para levar a cabo este projeto. O entusiasmo, a infinita paciência e a amizade, tal como os momentos de descontração, foram essenciais para conseguir ultrapassar as fases mais críticas desta dissertação. Agradeço os constantes e brilhantes ensinamentos científicos nas diferentes áreas que este trabalho abrange. Agradeço também os muitos ensinamentos pessoais. Este doutoramento fez-me crescer a vários níveis e devo-lhe essa oportunidade. Por fim, agradeço o apoio financeiro nesta fase final, pós bolsa de doutoramento.

To my co-supervisor, Dr. Dennis K. Lanning, I thank for sharing his scientific knowledge, time and bench space with me. The opportunity of learning profoundly different molecular biology techniques was essential for my completion as a biologist. Your vast knowledge in Immunology greatly increased my curiosity for this scientific field. Lastly, thank you so much for your invaluable help every time I cried for it.

I owe a very special acknowledgment to Dr. Grant McFadden for having me in his lab. Thank you so much for having replied kindly and patiently to all my inquiries when I first approached you. Your ability to teach and share knowledge is amazing, making me to want to learn more. Also, your input deeply improved my work and, consequently, this dissertation. Thank you for believing in my scientific capabilities.

Ao Professor Doutor Paulo Alexandrino e ao Professor Doutor Nuno Ferrand de Almeida, da Comissão Coordenadora e Científica do Programa Doutoral em Biodiversidade, Genética e Evolução (BIODIV), agradeço a disponibilidade para a resolução célere das questões burocráticas relacionadas com o programa.

À Fundação para a Ciência e a Tecnologia (FCT), agradeço a concessão de uma Bolsa de Doutoramento (SFRH/BD/48566/2008) e o subsídio de um Projeto de Investigação (PTDC/BIA-BEC/103158/2008).

Ao Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO) e ao Centro de Testagem Molecular (CTM), agradeço a ajuda técnica proporcionada por muitos dos seus elementos e as condições laboratoriais proporcionadas.

I would like to thank Jeff Wilcox and Dr. Michael Hamilton from the Blue Oak Ranch Reserve (University of California, Berkeley) for providing us brush rabbit tissue samples, which proved to be crucial to the development of this work.

I am very grateful to Dr. Katherine L. Knight for having me in her lab in the Department of Microbiology and Immunology, in the Stritch School of Medicine, Loyola University Chicago and for allowing me to integrate the lab and the department activities.

I would also like to thank all the help provided by the staff of the Department office, namely Sheila, Debbie and Jan.

A very special thanks to all the elements of the Knight lab that always made me feel cherished and that patiently helped me through my stay in the lab, specially Basile, Britte, Dom, Farenna, Greg, Karina, Mae, Pehga, Pi-Chen, Sara and Shi-Kang.

To Shi-Kang, the molecular biology 'master', and Greg I specially thank their help, and their precious tips and advices in PCR, molecular cloning and Southern Blot techniques. I also thank Greg for his friendship, his companionship, our great conversations and my funny 'lab name'.

To Pi-Chen and Farenna I thank their patience in teaching me the first steps in mammalian cell culture. I thank Farenna, whom I greatly admire, for all her complicity and her friendship.

I thank the people from the McFadden lab (Department of Molecular Genetics and Microbiology, College of Medicine, University of Florida) for all their help, support and guidance, especially Jia, Masmudur, Sherin and Winnie.

Sherin, I really appreciated your precious help when dealing with the 'insane' American bureaucracy. Also, I am very grateful for your immediate availability every time I needed something in the lab and for your patience in teaching me the Western blot steps.

To Jia, which was always present, I am very grateful for her personal support and incredible scientific guidance. Her ability to teach me patiently several molecular techniques applied to Virology was extremely enriching. Your vision of Science greatly improved me as a biologist and, consequently, enhanced the quality of this thesis.

A very special thanks to Nissin, Desyree and Baron, from the Condit lab (Department of Molecular Genetics and Microbiology, College of Medicine, University of Florida).

Aos amigos e colegas do CIBIO e do CTM, que desde o início foram um grande pilar para aceitar, realizar e concluir o desafio que foi esta dissertação.

À Joana agradeço a amizade, o incentivo e o grande impulso que deu para que eu aceitasse levar a cabo este projeto. A integração no CIBIO, no laboratório, na dinâmica e nas diferentes áreas científicas exploradas pelo grupo foram fundamentais.

À Sofia e à Diana um grande obrigada pela amizade, boa disposição, momentos de descontração e críticas construtivas ao meu trabalho.

À Margarida e à Fátima agradeço a amizade, o carinho e por acreditarem nas minhas capacidades para realizar esta tese (e também na minha aptidão para conduzir!).

Ao Zef e à Helena agradeço a preciosa e paciente ajuda na introdução ao mundo das análises filogenéticas e de seleção.

À Ana Pinheiro, à Cláudia Almendra, à Fabiana Neves, ao Fernando Lima, ao Miguel Carneiro, ao Nuno Queiroz, à Raquel Xavier, ao Ricardo Lopes, à Rita Campos, à Sandra Afonso, à Sara João, à Sónia Rosenbom, à Sílvia Carvalho, à Susana Lopes, à Teresa Silva e à Vânia Costa, um grande agradecimento, pois cada um à sua maneira fez-me sentir integrada e foi uma preciosa ajuda para o sucesso desta tese.

Às gestoras de projetos do CIBIO, Sara Ferreira e Sandra Rodrigues, agradeço a prontidão com que sempre me ajudaram na resolução das complicadas burocracias que acompanham a Ciência.

À Teresa “do bar” um grande obrigada por me receber tão bem sempre que apareço.

I am very grateful to Starrla, Sara and their cats, my roommates in Chicago, for making me feel at home and for their great support when everything in the US was new to me. Our cohabitation was not always easy, but our differences made me grow as a person. A very special thanks to Starrla for her friendship, for introducing me to Chicago, for our great conversations, for the Missouri trip and for other exceptional gestures.

A todos os meus amigos que sempre foram estando presente ao longo destes anos e que, de diferentes formas e mesmo à distância, sempre demonstraram a sua amizade, carinho, preocupação e apoio. À Albina, à Cátia, à Cláudia, à Estela, ao Hugo, à Ivone, à Li, ao Louis, ao Luizinho, ao Mário Luís, à Patrícia, à Raquel e à Vânia, um obrigada muito especial. Um enorme agradecimento também às amigas da família Elsa, Manuela, Aninhas e Zé que sempre se preocuparam comigo.

Às minhas meninas, Joana, Marlene e Marta um enorme obrigada por toda a amizade, companheirismo e preocupação que sempre demonstraram. O vosso enorme carinho e apoio têm sido fundamentais para a elaboração desta tese, tal como em outros momentos da minha vida. Um agradecimento especial também ao Filipe, ao Hélder, ao Jair e à pequena Beatriz.

À Cátia e ao Ginho, aquele agradecimento enorme e especial. Muito obrigada pela amizade, pelo muito carinho demonstrado e pela vossa constante presença. Muito obrigada pela possibilidade que tiveram de me visitar nos EUA e por todos os nossos momentos divertidos. Muito obrigada por me deixarem fazer parte das vossas vidas.

À minha família pelo apoio e pelo orgulho que sente em mim. Aos meus avós Celeste, Rolando e Nanda, e aos meus tios Nana e Zé, muito obrigada. Um agradecimento ainda mais especial à minha prima Zé pela sua loucura genuína e pelas gargalhadas que sempre me arrancou, mesmo à distância.

À Emerência. São muitas as vezes que penso que esta tese é tão tua quanto minha. Rejubilaste, quando rejubilei. Sofreste, quando sofri. Sorriste, quando sorri. Desesperaste, quando desesperei. Sempre comigo. É sem dúvida tão tua quanto minha. A cumplicidade, a amizade e o companheirismo, sem filtros, nem pudores e com toda a sinceridade, foram tão essenciais para a construção desta tese como para a minha construção pessoal ao longo destes anos. Muito obrigada!

À minha Mãe. A minha força. O meu porto de abrigo. Sem a tua coragem e sem o teu grande apoio, nada disto teria sido possível. Na realidade, nada na minha vida teria sido possível. O teu amor, a tua abnegação, a tua dedicação, a tua compreensão e a tua motivação são e serão sempre o meu suporte. Espero conseguir continuar a deixar-te orgulhosa das minhas atitudes e decisões. Esta tese é um “brinde a nós”, por tudo aquilo que passamos e por todos os obstáculos que tivemos que ultrapassar. “Perdoai-lhes Senhor, Porque eles sabem o que fazem”.

Ao Mário. Que aceitou entrar na minha vida quando a distância (geográfica) era a única grande certeza entre nós os dois. E no entanto, apesar de todos sacrifícios e contrariedades, aqui estamos, juntos, prontos para novos desafios. A ti, agradeço eternamente o amor, a amizade, a confiança, a dedicação, a partilha, o apoio, a tolerância, os sorrisos e as gargalhadas. Nos momentos mais críticos desta dissertação, as tuas certezas sobre as minhas capacidades foram fundamentais. Por tudo isto e por todas as outras pequenas coisas só nossas, o meu terno e eterno obrigada. “Amo-te por todas as razões e mais uma”.

Sumário

O género *Oryctolagus* (ordem Lagomorpha, família Leporidae) inclui uma única espécie, o coelho-bravo (*O. cuniculus*). Apesar de a espécie ter a sua origem na Península Ibérica, o coelho-bravo é um colonizador generalista, tendo sido introduzido em todo o mundo. Fora do seu habitat natural, o coelho-bravo é considerado uma espécie invasora, causadora de um elevado impacto sobre a biodiversidade local, tal como se tem verificado na Austrália desde a introdução da espécie, onde a erradicação do coelho-bravo se tornou uma prioridade para a conservação. Por outro lado, as populações existentes na Península Ibérica têm sofrido um declínio acentuado nos últimos anos. Como tal, o coelho-bravo está atualmente classificado na Lista Vermelha das Espécies Ameaçadas da União Internacional para a Conservação da Natureza (IUCN) como espécie Quase Ameaçada. A mixomatose, doença viral causada pela infeção com o vírus do mixoma (VMIX), é apontada como uma das principais causas responsáveis pelo declínio das populações de coelho-bravo. A elevada mortalidade originou, nos anos 50 do século XX, a introdução do VMIX na

Austrália como agente biológico, com o intuito de controlar as populações selvagens de coelho-bravo.

O VMIX é um poxvírus (género *Leporipoxvirus*) para o qual estão descritos dois hospedeiros naturais do género *Sylvilagus* (ordem Lagomorpha, família Leporidae): uma espécie sul-americana, o coelho-brasileiro (*Sylvilagus brasiliensis*), e uma espécie norte-americana, o coelho-do-chaparral (*S. bachmani*). Nos hospedeiros naturais, o vírus causa um fibroma cutâneo restrito ao local de inoculação. Foram definidos dois tipos de VMIX geograficamente distintos, o sul-americano e o norte-americano, correspondendo cada um dos tipos ao hospedeiro em que o vírus circula. Atualmente, os genomas da estirpe sul-americana Lausanne (Lu) e da estirpe norte-americana MSW estão sequenciados por completo, tendo-se verificado que o genoma da estirpe MSW (164,6 kb) é maior que o genoma da estirpe Lu (161,8 kb), devido principalmente à expansão das repetições terminais invertidas. A partir da sequenciação das estirpes, verificou-se que ambas as extremidades dos genomas do VMIX codificam proteínas imunomoduladoras responsáveis pela interação com o hospedeiro e que são altamente específicas dos leporipoxvírus, permitindo ao vírus contornar o sistema imunitário do hospedeiro. A estirpe sul-americana Lu tem sido a mais estudada e a mais frequentemente utilizada tanto em estudos *in vivo* como *in vitro*.

O VMIX é um exemplo clássico de um agente patogénico adaptado ao seu hospedeiro natural que pode infetar um novo hospedeiro, e adaptar-se a algumas populações desse novo hospedeiro. Este cenário foi observado na Austrália, onde algumas populações de coelho-bravo se tornaram resistentes à infeção pelo VMIX num processo de co-evolução entre hospedeiro e agente patogénico. No entanto, a base da resistência genética em coelhos-bravos selvagens australianos é ainda desconhecida. A explicação para a diferente patogenicidade da infeção por VMIX entre os hospedeiros naturais do género *Sylvilagus* e o novo hospedeiro, o coelho-bravo, é também ela desconhecida. Nesta dissertação, foram estudados os aspetos evolutivos de genes associados ao processo infecioso do VMIX, nomeadamente para o coelho-bravo e para a espécie norte-americana coelho-do-chaparral, com o intuito de determinar a existência de possíveis diferenças genéticas que clarifiquem a resistência e a suscetibilidade ao VMIX.

Anteriormente, um estudo evolutivo do recetor de quimiocinas *CCR5* em leporídeos mostrou a primeira diferença marcante entre os géneros *Oryctolagus* e *Sylvilagus*, conduzindo ao estudo neste trabalho de três quimiocinas inflamatórias que sinalizam através da ligação ao recetor CCR5, as quimiocinas CCL3, CCL4 e CCL5. No trabalho anterior, verificou-se que o recetor CCR5 do género *Oryctolagus*

experienciou um evento de conversão génica na segunda ansa extracelular com o recetor de quimiocinas CCR2, fenómeno não observado nos géneros *Sylvilagus* e *Lepus*, um terceiro género de leporídeos com algumas evidências de suscetibilidade ao VMIX. As análises de seleção realizadas no presente trabalho demonstraram que as três quimiocinas inflamatórias se encontram sob forte seleção negativa e, de facto, a divergência evolutiva entre as linhagens *Oryctolagus*-*Sylvilagus*-*Lepus* para cada proteína é semelhante. As diferenças genéticas diminutas entre as espécies de leporídeos sugerem restrições funcionais nas quimiocinas CCL3, CCL4 e CCL5. Porém, é fundamental avaliar se as substituições de aminoácidos observadas, específicas de certas espécies ou géneros, influenciarão os padrões de sinalização entre as quimiocinas e o recetor CCR5.

A proteína viral M062, fator determinante para o tropismo específico do VMIX, é essencial para a patogénese do vírus no coelho-bravo e para a replicação em vários tipos de células humanas. Esta proteína viral tem a capacidade de antagonizar as propriedades antivirais da proteína humana SAMD9, permitindo a replicação do VMIX em células humanas. Sendo SAMD9 um importante fator antiviral inato contra poxvírus, realizamos um estudo centrado na história evolutiva do gene *SAMD9* e do seu parálogo cromossomicamente adjacente SAMD9-like (*SAMD9L*) em mamíferos, tendo como objetivo a deteção de assinaturas de seleção positiva. Foram encontradas evidências de pressões seletivas ao longo da história evolutiva das duas proteínas (SAMD9 e SAMD9L), provavelmente exercidas por diferentes tipos de vírus, e foram identificados codões específicos sob seleção positiva. Num estudo realizado anteriormente, foi sugerido pelos autores que os genes *SAMD9* e *SAMD9L* nos mamíferos terão resultado de um evento de duplicação génica a partir de um ancestral comum. As análises realizadas neste trabalho permitiram suportar a hipótese avançada anteriormente. Concretamente, os nossos resultados sugerem que este evento ocorreu após a separação dos Marsupialia dos Placentalia. Com a perda aparente do gene *SAMD9* ou *SAMD9L* em algumas espécies, propusemos também que poderá existir alguma redundância funcional entre as duas proteínas.

A identificação a nível intracelular de ARN 'estranho' em células infetadas por vírus em fase replicativa é mediada por um grupo de três helicases de ARN: RIG-I, MDA5 e LGP2. Um estudo realizado anteriormente descreveu a proteína RIG-I como um sensor da infeção do VMIX em macrófagos humanos primários, tornando estas células não permissivas à replicação do vírus. No presente trabalho procuramos compreender a história evolutiva, para além de evidências de seleção positiva, nestas três helicases, tanto em mamíferos como mais especificamente em leporídeos dos géneros *Oryctolagus*, *Sylvilagus* e *Lepus*. Encontramos evidências claras de seleção

positiva a atuar nestas proteínas do sistema imunitário inato do hospedeiro nos dois conjuntos de dados analisados, tendo sido identificados vários codões sob seleção positiva. Uma vez que as helicases de ARN estão envolvidas na identificação de vírus patogénicos, estas proteínas deverão evoluir rapidamente num processo dinâmico com os agentes patogénicos e, portanto, estarão sujeitas a pressões seletivas ao longo da evolução dos mamíferos para evitar possíveis infeções. Estas pressões seletivas foram observadas de uma forma mais acentuada no domínio repressor da helicase RIG-I, região responsável pela identificação dependente de 5'-trifosfato (5'-ppp) e pela ligação aos substratos de ARN. Este resultado sugere que estas pressões terão sido impostas pelo grande número de vírus que são reconhecidos pela helicase RIG-I. Comparando especificamente a proteína RIG-I do coelho-bravo com a do coelho-do-chaparral, verificou-se que as diferenças de aminoácidos entre codões sob seleção das duas espécies também se encontram localizadas no domínio repressor. Estes resultados são importantes para a compreensão das respostas inatas específicas das espécies de hospedeiros a certos agentes patogénicos, e mais concretamente as respostas imunitárias à infeção por VMIX.

O papel da proteína TRIM5α na restrição de diversos retrovírus tem sido extensivamente descrito. Nos primatas, o domínio PRYSPRY desta proteína contém, numa região importante para a restrição retroviral específica de cada espécie, uma elevada concentração de resíduos positivamente selecionados. A compreensão da evolução do gene *TRIM5* nos lagomorfos tornou-se importante, uma vez que o lentivírus endógeno RELIK foi identificado em leporídeos do género *Oryctolagus*, *Bunolagus*, *Sylvilagus* e *Lepus*. A restrição de vários retrovírus pela proteína TRIM5α do coelho-bravo e da lebre-europeia (*Lepus europaeus*) foi descrita em diferentes estudos. No presente trabalho, o padrão específico de cada espécie, anteriormente observada na região v1 do domínio PRYSPRY entre *Oryctolagus* e *Lepus*, foi também identificado na proteína TRIM5α do género *Sylvilagus*. As análises de seleção realizadas para o gene *TRIM5* nos lagomorfos, e particularmente na região v1 do domínio PRYSPRY, sugerem que o gene evoluiu sob elevada pressão seletiva imposta por endovírus antigos.

Summary

The *Oryctolagus* genus (order Lagomorpha, family Leporidae) contains a single extant taxon, the European rabbit (*O. cuniculus*). The species was originated in the Iberian Peninsula; yet, the European rabbit is a widespread colonizer introduced by human-mediated action in several countries, including Australia and various European countries and islands. Outside of its natural range, the European rabbit is considered a pest species with high impact on the local biodiversity, such as in Australia where the eradication of the European rabbit became a priority for conservation. On the other hand, populations within the Iberian Peninsula have been suffering a massive decline in recent years. Indeed, the European rabbit is currently classified in the International Union for Conservation of Nature (IUCN) Red List of Threatened Species as Near Threatened. Amongst the primary causes of decreasing rabbit populations is a viral disease named myxomatosis, caused by myxoma virus (MYXV) infection. Nonetheless, the virus was introduced in Australia in the 1950s as a biological control agent in an attempt to manage the wild European rabbit populations.

MYXV is a poxvirus (genus *Leporipoxvirus*) naturally circulating in the South American tapeti (*Sylvilagus brasiliensis*) and in the North American brush rabbit (*S. bachmani*). In its native *Sylvilagus* (order Lagomorpha, family Leporidae) hosts, MYXV only causes a cutaneous fibroma restricted to the site of inoculation. Two types of geographically distinct MYXV, corresponding to each host in which it naturally circulates, have been identified. The complete genomes of the South American Lausanne (Lu) strain and the North American MSW strain have been sequenced. The MSW genome (164.6 kb) is larger than the Lu genome (161.8 kb), mainly due to an expansion of the terminal inverted repeats (TIRs). Both ends of the MYXV genomes encode host-interactive immunomodulatory proteins highly specific to leporipoxviruses, which allow the virus to circumvent host immune defenses. The South American Lu strain is the most studied and more often used strain both in *in vivo* and *in vitro* gene knockout studies.

MYXV is a classic example of a pathogen adapted to its native host that can infect a new susceptible host, yet adapt to some populations of the new species. This scenario was observed in Australia, where certain European rabbit populations became resistant to MYXV infection in a co-evolutionary process between host and pathogen. However, the genetic basis of resistance in Australian wild rabbits is still unknown. The origin of differential pathogenicity of MYXV infection in native *Sylvilagus* hosts and in the susceptible *Oryctolagus* host is also still uncertain. In this dissertation, we focused on studying evolutionary and genetic aspects of important host candidate genes, particularly between European rabbit and the North American brush rabbit, to shed some light on possible genomic differences that might explain MYXV susceptibility and resistance.

Previous studies of leporid C-C motif chemokine receptor 5 (*CCR5*) genes revealed the first striking difference between genera, and more importantly between *Oryctolagus* and *Sylvilagus*, which prompted the study of three of the receptor disease-related inflammatory C-C motif chemokine ligands CCL3, CCL4 and CCL5. Indeed, *Oryctolagus CCR5* experienced a gene conversion event with C-C motif chemokine receptor 2 (*CCR2*) in the second extracellular loop, what was not observed in *Sylvilagus* and *Lepus*, a third leporid genus with some evidences of susceptibility to MYXV. Selection analyses demonstrated that the three inflammatory CCR5 ligands are under strong purifying selection and, in fact, the evolutionary divergence between the *Oryctolagus-Sylvilagus-Lepus* lineages for each protein is similar. The genetic differences between species suggest functional binding constraints in CCL3, CCL4 and CCL5, yet it is fundamental to assess whether the species- or genus-specific amino

acid substitutions observed in signaling and receptor-binding regions of chemokines influence ligand-receptor binding.

The M062, a MYXV host range factor, is essential for pathogenesis of the virus in the European rabbit and for the replication of MYXV in several human cells. Importantly, M062 was also found to antagonize the anti-viral properties of human sterile alpha motif domain-containing protein 9 (SAMD9), allowing viral replication in cultured human cells. With the role of SAMD9 as an innate anti-viral factor against poxviruses, we were interested in studying the evolutionary history of *SAMD9* and its chromosomally adjacent paralogue SAMD9-like (*SAMD9L*) in mammals and to look for signatures of positive selection. We found evidence of long-term selective pressures acting on both mammalian SAMD9 and SAMD9L, probably exerted by viral pathogens, and identified specific codons under positive selection. Additionally, the previously suggested origin of *SAMD9* and *SAMD9L* from a mammalian ancestral duplication event was supported by our analyses. Specifically, our results suggest that this event occurred after the divergence of Marsupialia from Placentalia. With the apparent loss of *SAMD9* or *SAMD9L* in some species, we also anticipated that some overlapping functional redundancy might exist between the two proteins.

The intracellular recognition of 'nonself' RNA from actively replicating viruses in infected cells is mediated by a group of three DExD/H box RNA helicases: retinoic acid-inducible gene-I (RIG-I) protein, melanoma differentiation associated factor protein 5 (MDA5) and laboratory of genetics and physiology 2 (LGP2) protein. Strikingly, RIG-I was described as a sensor for MYXV infection in primary human macrophages, making these cells nonpermissive for virus replication. We searched for evidence of positive selection in the three members of the RIG-I-like receptor (RLR) family, both in mammals and in the leporid genera *Oryctolagus*, *Sylvilagus* and *Lepus*. We found clear evidence of positive selection operating in these host innate sentinel proteins in the two sets of data analyzed with several codons also under positive selection. Since RLRs are deeply involved in pathogen recognition, they must rapidly evolve in a dynamic arms race with pathogens, and thus are subjected to long-term positive selection pressures to avoid potential infections. The RIG-I repressor domain, the region responsible for 5'-triphosphate (5'-ppp)-dependent recognition and binding to its RNA substrates, exhibited the strongest evidence of selective pressures in mammals. This result suggests that pressure has been imposed by the great number of viruses that are recognized by RIG-I. When comparing specifically European rabbit and brush rabbit RIG-I proteins, the amino acid differences of positively-selected codons between the two species were also located in the repressor domain. These insights into the

evolution of the mammalian and leporid RIG-I genes might help illuminate the origins of the species-specific innate responses to pathogens, and more specifically to MYXV.

The tripartite motif-containing protein 5 alpha (TRIM5α) mediates the restriction of retroviruses. In Primates, the TRIM5α PRYSPRY domain contains a high concentration of positively selected residues in a 'patch' critical for species-specific retroviral restriction. We were interested in understanding the evolution of *TRIM5* gene in Lagomorpha, because an endogenous lentivirus, the rabbit endogenous lentivirus type K (RELIK), has been reported in the leporid genera *Oryctolagus*, *Bunolagus*, *Sylvilagus* and *Lepus*. The restriction of divergent retroviruses by European rabbit and European brown hare (*Lepus europaeus*) TRIM5α has also been reported in different studies. The divergent species-specific pattern previously observed between the *Oryctolagus* and *Lepus* PRYSPRY v1 regions was also present in *Sylvilagus* TRIM5α. Moreover, selection analysis of Lagomorpha *TRIM5*, and particularly the PRYSPRY v1 region, suggests that the gene has evolved under high selective pressure imposed by ancient endoviruses.

Palavras-chave/Keywords

- Coelho-bravo/European rabbit
- Coelho-do-chaparral/Brush rabbit
- Vírus do mixoma/Myxoma virus
- Genes do sistema imunitário inato/Innate immunity genes
- Interação hospedeiro-agente patogénico/Host-pathogen interaction
- História evolutiva/Evolutionary history
- Co-evolução/Co-evolution
- Seleção positiva/Positive selection

Index of contents

Paper 3

Paper 4

Index of Figures

Chapter 1

General introduction

Chapter 2

Innate anti-viral immunity – Genetic aspects of C-C motif chemokines in Leporidae genera

Index of Tables

Chapter 3

Innate anti-viral factors – Evolution and genetic characterization of sterile alpha motif domain-containing protein 9

Chapter 4

Innate anti-viral factors – Evolution and genetic characterization of RIG-I-like receptors

Abbreviations

AIC: Akaike information criterion

AML: acute myeloid leukemia

AST: average survival time

ATP: adenosine triphosphate

Bcl-2: B-cell lymphoma 2

BEB: Bayes empirical Bayes

CARD: caspase activation and recruitment domain

CCL: C-C motif chemokine

CCR: C-C motif chemokine receptor

CD: cluster of differentiation

cDNA: complementary DNA

CFR: case fatality rate

CLR: C-type lectin receptor

CTD: C-terminal domain

DAMBE: data analysis and molecular biology and evolution

dN: nonsynonymous substitution

DNA: deoxyribonucleic acid

dS: synonymous substitution

dsDNA: double-stranded DNA

dsRNA: double-stranded RNA

EGF: epithelial growth factor

eIF2α: eukaryotic translation initiation factor 2 alpha

ER: endoplasmic reticulum

FEL: fixed effect likelihood

FIBV: hare fibroma virus

FUBAR: fast unbiased Bayesian approximation

GARD: genetic algorithm for recombination detection

GARLI: genetic algorithm for rapid likelihood inference

GTR: general time reversible

HCMV: human cytomegalovirus

HIV: human immunodeficiency virus

ICTV: International Committee on Taxonomy of Viruses

IFITM: interferon-induced transmembrane protein

IFN: interferon

Ig: immunoglobulin

IL: interleukin

IRF: interferon regulatory factor

ISG: interferon-stimulated gene

IUCN: International Union for Conservation of Nature

JMML: juvenile myelomonocytic leukemia

JSRV: Jaagsiekte sheep retrovirus

JTT: Jones-Taylor-Thornton model

KH: Kishino-Hasegawa test

KSHV: Kaposis's sarcoma-associated herpesvirus

LGP2: laboratory of genetics and physiology 2 protein

LRT: likelihood ratio test

Lu: Lausanne strain

MCMC: Markov chain Monte Carlo

MDA5: melanoma differentiation-associated factor protein 5

MDS: myelodysplastic syndrome

MEME: mixed effects model of evolution

MGF: myxoma growth factor

MHC: major histocompatibility complex

ML: maximum likelihood

MLV: murine leukemia virus

MNF: myxoma nuclear factor

Mya: million years ago

MYXV: myxoma virus

NCBI: National Center for Biotechnology Information

NFκB: nuclear factor kappa B

NFTC: normophosphatemic familiar tumoral calcinosis

NJ: neighbor joining

NLR: NOD-like receptor

ORF: open reading frame

PAML: phylogenetic analysis by maximum likelihood

PAMP: pathogen-associated molecular pattern

PARRIS: partitioning approach for robust inference of selection

PCR: polymerase chain reaction

pi: post-infection

pI: isoelectric point

PKR: interferon-induced, double-stranded RNA-activated protein kinase R

PRR: pattern-recognition receptor

RD: repressor domain

REL: random effect likelihood

RELIK: rabbit endogenous lentivirus type K

RFLP: restriction fragment length polymorphism

RFV: rabbit fibroma virus

RHA: RNA helicase A

RHDV: rabbit haemorrhagic disease virus

RIG-I: retinoic acid-inducible gene-I protein

RING: really interesting new gene

RLR: RIG-I-like receptor

RNA: ribonucleic acid

RT-PCR: reverse transcription-polymerase chain reaction

SAM: sterile alpha motif

SAMD9: sterile alpha motif domain-containing protein 9

SAMD9L: sterile alpha motif domain-containing protein 9-like

SAMHD1: sterile alpha motif domain- and HD domain-containing protein 1

SFV: Shope fibroma virus

SIV: simian immunodeficiency virus

SLAC: single likelihood ancestor counting

SLS: standard laboratory strain

SQFV: squirrel fibroma virus

ssRNA: single-stranded RNA

TIM: transition model

TIR: terminal inverted repeat

TLR: Toll-like receptor

TNF: tumor necrosis factor

TPM: three-parameter model

TRIM: tripartite motif-containing protein

TRIM5α: tripartite motif-containing protein 5 alpha

Ub: ubiquitin

Ur: Uriarra strain

VV: vaccinia virus

ZAP: zinc-finger anti-viral protein

Thesis framework

The work presented in this dissertation was integrated in a project supported by the *Fundação para a Ciência e a Tecnologia* entitled 'Why is Myxoma virus infection highly lethal for the European rabbit (*Oryctolagus cuniculus*), while it is benign in other leporids?' (PTDC/BIA-BEC/103158/2008).

The wild populations of European rabbit (*Oryctolagus cuniculus*), both in their native and non-native ranges, have been suffering an alarming decline in the last decades. Although different causes for the populations' decline have been identified, one of the greatest forces behind such decrease has been the highly lethal disease myxomatosis, caused by the rabbit-specific myxoma virus (MYXV). On the other hand, MYXV is naturally found in two American leporid species, the tapeti (*Sylvilagus brasiliensis*) and the brush rabbit (*Sylvilagus bachmani*), in which it causes an innocuous localized cutaneous fibroma.

Understanding the processes behind MYXV's strict host-tropism requires both the study of viral proteins responsible for the manipulation of host defense mechanisms, and the study of host evasion strategies to circumvent the virus infectious machinery. The work developed in this thesis was designed to advance understanding of the host evasion strategies. Therefore, the evolution and genetic aspects of host candidate genes that might be relevant for the variable susceptibilities to MYXV in leporid species are the focus of the studies presented here.

The thesis is organized into six chapters and a total of five scientific manuscripts published in journals indexed in the Science Citation Index (SCI) are include.

A review of the current knowledge on MYXV is presented in the first chapter, entitled *Chapter 1. General introduction*. Aspects on MYXV natural history, evolution, pathogenicity and host genetic resistance are comprehensively reviewed. Furthermore, the evolutionary aspects of host restriction factors and their role in viral evasion are also reviewed in this chapter.

The genetic aspects of the C-C chemokine receptor 5 (CCR5) in leporid genera with different susceptibilities to MYXV (*Oryctolagus*, *Sylvilagus* and *Lepus*) was previously investigated. The most striking observation was the replacement, in *Oryctolagus*, but not in *Sylvilagus* and *Lepus*, of a CCR5 motif in the second extracellular loop by a motif characteristic of C-C chemokine receptor 2 (CCR2). This resulted from an event of gene conversion between *CCR5* and *CCR2* in the European rabbit. Although the possible role of this gene conversion in the susceptibility to MYXV was not explored, we studied the evolution in leporid of three of the inflammatory C-C chemokines that signal through CCR5: CCL3, CCL4 and CCL5. The resulting scientific manuscript corresponds to *Chapter 2. Innate anti-viral immunity – Genetic aspects of C-C motif chemokines in Leporidae genera*:

Paper 1. A. Lemos de Matos, D. K. Lanning, P. J. Esteves (2013). Genetic characterization of CCL3, CCL4 and CCL5 in leporid genera *Oryctolagus*, *Sylvilagus* and *Lepus*. International Journal of Immunogenetics, doi: 10.1111/iji.12095 *[Epub ahead of print]*

MYXV, like all poxviruses, encodes an extensive repertoire of virulence genes, which express an array of proteins responsible for the subversion and modulation of the host anti-viral responses. Additionally, a group of these genes, designated as host range genes, are essential for MYXV-specific tropism. It is likely that the action of these

MYXV species-specific gene products is responsible for the differing leporid susceptibilities to the virus. Simultaneously, the host proteins inhibited by these viral factors probably evolved to escape from virus action in the natural long-term host (*Sylvilagus* sp.), but not in the European rabbit. Therefore, in *Chapter 3. Innate anti-viral factors – Evolution and genetic characterization of sterile alpha motif domain-containing protein 9* we looked for evidence of selective pressures in a protein antagonized by MYXV host range factor M062, the sterile alpha motif domain-containing protein 9 (SAMD9) and in its chromosomally adjacent paralogue, SAMD9-like (SAMD9L). The results are presented in one scientific manuscript:

Paper 2. A. Lemos de Matos, J. Liu, G. McFadden, P. J. Esteves (2013). Evolution and divergence of the mammalian *SAMD9*/*SAMD9L* gene family. BMC Evolutionary Biology, 13: 121

The pattern-recognition receptor (PRR) retinoic acid-inducible gene-I (RIG-I) was described as a cytoplasmic sensor for MYXV infection in primary human macrophages, which are nonpermissive cells for MYXV replication. RIG-I sensing of MYXV in these cells leads to the co-induction of type I interferon (IFN) and tumor necrosis factor (TNF), resulting in the rapid termination of the virus replication cycle. This result in human cells supports the assumption that MYXV possesses the capacity to subvert RIG-I sensing in European rabbit cells, allowing completion of the virus replication cycle and consequent host susceptibility. By contrast, for MYXV natural long-term *Sylvilagus* host cells, we expect a similar result to that observed in primary human macrophages. Nevertheless, the mechanisms behind this species-specific subversion are still unknown. In *Chapter 4. Innate anti-viral factors – Evolution and genetic characterization of RIG-I-like receptors* the evolution of mammalian RIG-I protein and the two other members of the RIG-I-like receptor (RLR) family, the melanoma differentiation-associated factor protein 5 (MDA5) and the laboratory of genetics and physiology 2 (LGP2) protein, were investigated. The same studies were performed specifically for leporid genera where the host-virus interaction dynamics of MYXV have been previously reported. The results of these studies were reported in two manuscripts:

Paper 3. A. Lemos de Matos, G. McFadden, P. J. Esteves (2013). Positive evolutionary selection on the RIG-I-like receptor genes in mammals. PLoS One, 8 (11): e81864

Paper 4. A. Lemos de Matos, G. McFadden, P. J. Esteves (2014). Evolution of viral sensing RIG-I-like receptor genes in Leporidae genera *Oryctolagus*, *Sylvilagus*, and *Lepus*. Immunogenetics, 66 (1): 43-52

Host restriction factors are one of the most important elements of the innate immune system for fighting viral infection and replication. Besides their anti-viral activity, one of their distinguishable characteristics is the long-term evolutionary conflict with viral proteins, resulting in strong signatures of positive selection. One of these anti-viral restriction factors, the tripartite motif-containing protein 5 alpha (TRIM5α), was previously shown to restrict different retroviruses in the leporid species European rabbit and European brown hare (*Lepus europaeus*). Evolutionary analyses demonstrated that, as in primates, the domain responsible for species-specific restriction activity (PRYSPRY domain) showed evidence of selection in lagomorphs. This observation suggests that retroviruses may have subjected these lagomorph proteins to selective pressures. Thus, in *Chapter 5. Innate anti-viral factors – Genetic characterization of host restriction factor TRIM5α in Leporidae genera* we extended these observations to other leporid species, especially because the rabbit endogenous lentivirus type K (RELIK) was identified in the genomes of *Oryctolagus*, *Lepus* and *Sylvilagus*. One manuscript resulted from this work:

Paper 5. A. Lemos de Matos, W. van der Loo, H. Areal, D. K. Lanning, P. J. Esteves (2011). Study of *Sylvilagus* rabbit TRIM5α species-specific domain: how ancient endoviruses could have shaped the antiviral repertoire in Lagomorpha. BMC Evolutionary Biology, 11: 294

Finally, in *Chapter 6. Final considerations*, the major conclusions of this dissertation and the implications of the results for future research are presented.

Chapter 1

General introduction

1. Myxoma virus

Myxoma virus (MYXV) is a poxvirus (family *Poxviridae*; subfamily *Chordopoxvirinae*) belonging to the genus *Leporipoxvirus* [1]. MYXV shares with other poxviruses a characteristic large brick-shaped virion, exhibiting dimensions of 286 x 230 x 75 nm [2]. Moreover, MYXV possesses other poxvirus characteristic features, such as a large, linear double-stranded DNA (dsDNA) genome flanked by terminal inverted repeats (TIRs) and covalently closed hairpin loops at its extremities [3]. The complete MYXV replication cycle occurs exclusively and autonomously in the cytoplasm of infected cells [4].

The complete genomic DNA of the MYXV Lausanne (Lu) strain, the most studied and most often used strain both in *in vivo* and *in vitro* studies, has been sequenced [5]. In its 161.8 kb genome, a total of 171 open reading frames (ORFs) were identified, of which 12 are duplicated in each TIR [5]. Approximately 100 genes associated with the viral replicative machinery and structure are located within the central 120 kb region of the genome, while genes encoding host-interactive immunomodulatory proteins and host range proteins are primarily located within the 15-25 kb at both ends of the genome [5]. Although the replicative and structural genes are strongly conserved among poxviruses, the virulence-related genes are more specific to leporipoxviruses, and most likely evolved in close association with the natural hosts to allow the virus to endure in this non-favorable environment [4, 6]. Recently, the complete genomic DNA sequencing of a second MYXV strain, the Californian MSW strain, has been published [7]. A comparative analysis between the two virus strains is included in section 1.4.

1.1. Leporipoxviruses and host tropism

According to the International Committee on Taxonomy of Viruses (ICTV) current release (2012), the *Leporipoxvirus* genus includes myxoma virus (MYXV) and three other members: the hare fibroma virus (FIBV), the rabbit fibroma virus (RFV; or Shope fibroma virus, SFV) and the squirrel fibroma virus (SQFV). SQFV is endemic in eastern grey squirrel (*Sciurus carolinensis*) populations within the species natural range in eastern North America [8-10]. The remaining three leporipoxviruses naturally occur in leporid genera, namely *Lepus* (FIBV) and *Sylvilagus* (MYXV and RVF) [11-14].

1.1.1. *Leporipoxvirus* hosts: taxonomy and evolutionary relationships

Leporids are members of the taxonomic family Leporidae, integrating the order Lagomorpha together with the monotypic family Ochotonidae (a single genus,

Ochotona). Leporidae includes jackrabbits and hares (genus *Lepus*), and rabbits (genera *Brachylagus*, *Bunolagus*, *Caprolagus*, *Nesolagus*, *Oryctolagus*, *Pentalagus*, *Poelagus*, *Pronolagus*, *Romerolagus* and *Sylvilagus*) [15-17].

Of the two most widely distributed genera, *Lepus* is the most cosmopolitan, with representative species in America, Asia, Africa and Europe, while *Sylvilagus* species are extensively distributed throughout North and Central America and the northern half of South America [15, 17]. On the other hand, due to human action, *Oryctolagus* is the most widely dispersed genus beyond its natural range. As a monotypic genus, it includes the Iberian native species European rabbit (*Oryctolagus cuniculus*) as the only representative taxon. Multiple molecular markers allowed the identification of two subspecies, *Oryctolagus cuniculus cuniculus* and *Oryctolagus cuniculus algirus*, [e.g. 18-23] with an estimated separation time of ~1.8 million years ago (Mya) [24]. The lineage *cuniculus* was utilized for the human-mediated worldwide distribution and includes all domestic breeds, while the *algirus* lineage is restricted to the southwestern Iberian Peninsula and a few Atlantic islands [23]. European rabbit is the susceptible host of MYXV, since the virus causes an acute disseminated disease termed myxomatosis with mortality rates up to 100% [13].

A combination of several molecular markers allowed the establishment of a robust phylogeny for the Leporidae family, where the divergence time between *Lepus* and the common ancestor of *Sylvilagus* and *Oryctolagus* genera was estimated around 12 Mya [17]. Additionally, the separation time between the closely related *Sylvilagus* genus and the *Oryctolagus* ancestor was estimated around 10 Mya [17].

1.1.2. Myxoma virus in *Sylvilagus* natural hosts

Two closely related, but geographically distinct MYXV types occur in native *Sylvilagus* hosts: the South American (or Brazilian) type and the North American (also called Californian) type. The first circulates in jungle rabbit or tapeti (*S. brasiliensis*) populations in a broad region of South America, whereas the Californian type circulates in brush rabbit (*S. bachmani*) populations on the west coast of the United States of America, in Oregon and California, as well as in the Baja Peninsula of Mexico [12, 13, 25-29].

In both natural hosts, MYXV causes a relatively benign and localized cutaneous fibroma from which virus can be transmitted on the mouthparts of mosquitoes or other biting arthropods, such as fleas [13]. The transmission is passive as MYXV does not replicate in the vector [26, 30]. To ensure that the virus is able to persist long enough to be transmitted, it must encode diverse proteins that effectively suppress local

inflammatory and immune responses in both native *Sylvilagus* hosts [6]. Additionally, specific adaptations of each of these viruses to its respective natural host have occurred, since mosquitoes fed on brush rabbits experimentally inoculated with a South American MYXV strain did not transmit the virus [31] and mosquitoes carrying a Californian MYXV strain could not transmit the virus to tapeti rabbits [32] (Figure 1). All these characteristics taken together can be seen as the result of long-term host-virus co-evolution [13].

Strains of both MYXV types have been isolated and tested experimentally, especially in European rabbit, exhibiting differing levels of virulence and causing different pathogenesis. Both the highly virulent Lu strain isolated in Campinas, Brazil in 1949, and the highly virulent Brazilian standard laboratory strain (SLS), recovered from a naturally infected European rabbit, are examples of South American-derived strains [33, 34]. The Californian MSW strain (USA/San Francisco/1950/1) has been described as the most virulent of all MYXV strains [34], while the California-derived MSD (USA/San Diego/1949/1) is an attenuated strain [35].

1.1.3. Myxoma virus host tropism

MYXV natural host range is restricted to the two *Sylvilagus* species, tapeti and brush rabbit, but the susceptibility and epidemic potential of MYXV strains were experimentally tested in different leporid species [32, 36, 37].

A Californian strain recovered from a naturally infected brush rabbit was transmitted to individuals of the North American species desert cottontail (*S. audubonii*), eastern cottontail (*S. floridanus*), mountain cottontail (*S. nuttallii*) and pygmy rabbit (*Brachylagus idahoensis*). Although tumors developed in these species, the viral titres in tumor tissue were reduced and did not allow further transmission via mosquitoes (Figure 1) [32, 37]. No tumors developed in a North American *Lepus* species, the black-tailed jackrabbit (*L. californicus*), infected with the same Californian MYXV strain [32]. Interestingly, two of these North American rabbit species, the desert cottontail and the mountain cottontail, were described as susceptible to infection with South American MYXV strains and potentially able to transmit the virus (Figure 1) [36, 37]. The pathogenesis of South American and North American MYXV strains in these two *Sylvilagus* species is described in detail in Silver *et al.* (2010) [37].

The development of generalized disease, although very rare, has been documented in European brown hare (*L. europaeus*) individuals [13]. The pathogenesis of MYXV and the mechanisms of resistance/susceptibility in European rabbit are comprehensively described in section 1.2.

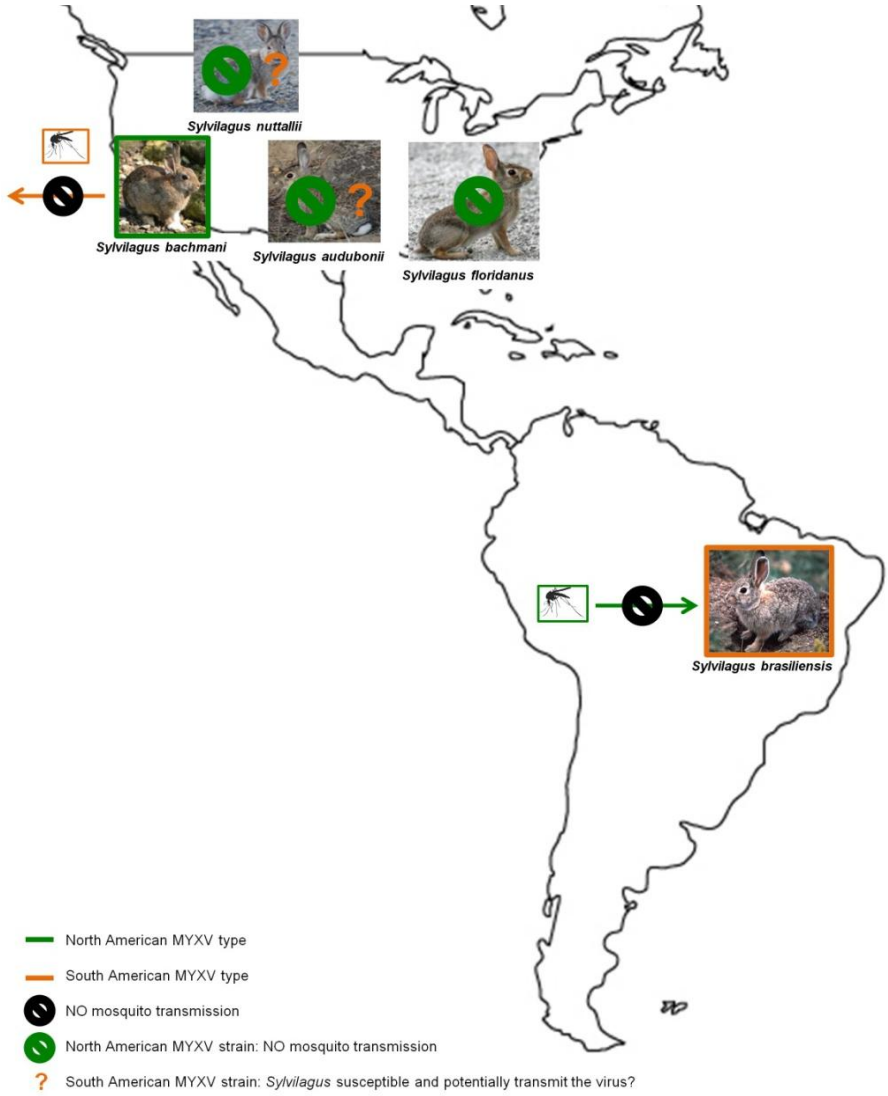


Figure 1 - MYXV and *Sylvilagus* hosts.
In the two American MYXV native hosts, *S. bachmani* and *S. brasiliensis*, the circulating strains belong to two specific MYXV types: the North American MYXV type (green) and the South American MYXV type (orange), respectively. Mosquitoes fed on *S. bachmani* specimens experimentally inoculated with a South American MYXV strain did not transmit the virus. Also, mosquitoes carrying a Californian MYXV strain could not transmit the virus to *S. brasiliensis* rabbits. The other three *Sylvilagus* species represented developed tumors when infected with a Californian strain, but mosquitoes did not transmit the virus. On the other hand, S. *audubonii* and *S. nuttallii* were described as susceptible to infection with South American MYXV strains and potentially able to transmit the virus. This figure presents a simplified illustration of *Sylvilagus* species locations and not an extensive representation of their natural range.

1.1.4. Rabbit fibroma virus

Rabbit fibroma virus (RFV) circulates in the natural reservoir eastern cottontail (*Sylvilagus floridanus*) in the East and central USA [13], causing cutaneous fibromas usually on the feet, legs, or muzzle [11]. Similarly to MYXV, the virus is spread

passively by biting arthropods, predominantly mosquitoes [38, 39]. RFV is genetically and antigenically closely related to MYXV, but in European rabbit no systemic signs of disease are observed, only a skin fibroma at the inoculation site [5, 13, 40]. The use of RFV as a live virus heterologous vaccine against myxomatosis has been documented [41, 42].

1.2. Host infection and disease: myxomatosis in European rabbit

1.2.1. Historic context

With the natural occurrence of MYXV in *Sylvilagus* hosts, the virus evolution and first reports of myxomatosis are closely related to the American continent. In 1895, the bacteriologist Giuseppe Sanarelli went to Uruguay to establish an Institute of Experimental Hygiene in Montevideo. While setting up the institute, domestic European rabbits acquired in Brazil were imported for the production of antibodies. Later, in 1896, the institute's rabbit colony suffered an outbreak of a highly contagious and lethal disease characterized by numerous lesions and tumors in the skin and conjunctiva of infected rabbits. Sanarelli recovered a non-cultivable and invisible infectious agent which he named 'myxomatogene virus' responsible for the disease he called 'infectious myxomatosis of rabbits' [13, 43, 44]. When MYXV began to be used as a biological control agent against the introduced European rabbit in Australia, leading to the first appearance of myxomatosis outside America, efforts to clarify the virus began, and it was only in 1953 that MYXV was integrated into the pox group of viruses (family *Poxviridae*) [45].

During the colonization of Australia in the middle of the 19th century, the early settlers introduced European rabbit as a source of meat and for sport hunting. Nevertheless, it was only in 1860, when Thomas Austin, a Victorian landowner, brought wild European rabbits that success in spreading the species was achieved, and rabbits rapidly became the most important animal pest in Australia with devastating effects on the continent's biodiversity [13, 46, 47]. Since myxomatosis is almost always fatal in European rabbit, the use of MYXV as a biological control agent for the feral rabbit population in Australia was suggested as early as 1919 [13, 48, 49]. Between 1933 and 1936, Sir Charles Martin carried out field experiments that lead him to state that MYXV should be suitable for the control of European rabbits in a confined area and that it appeared to be highly species-specific. Consequently, the Australian government approved several field releases [13, 44, 46, 47, 50]. The first epizootic of myxomatosis in Australia occurred in the summer of 1950/1951, after the apparent disappearance of the virus released during an experimental trial, which re-emerged in epizootic form [51].

The MYXV strain released was the highly virulent SLS, a strain derived, after numerous passages in rabbits, from another isolated in Brazil in 1911 [33, 50]. During the next 3 years, MYXV continued to re-emerge in epizootic forms spreading throughout the European rabbit-infested areas, with estimates of case fatality rates (CFRs) ranging from 40% to 99.8% [49, 52].

Myxomatosis also has a well-documented course in Europe, after the MYXV introduction in France and subsequent spread in the Old Continent [13]. In 1952, Dr. P. F. Armand Delille illegally inoculated wild European rabbits present in his property with MYXV Lu, the highly lethal strain isolated in Brazil that had undergone fewer than five rabbit passages [13, 53]. During the next 10 years, the virus spread uncontrollably into all areas of Europe where wild populations of European rabbit were present, reaching Britain in 1953 and becoming enzootic. The estimates of wild European rabbit population reduction reached values of 90-98% in France and greater than 99% in Britain [13, 54]. In the European rabbit native range, the Iberian Peninsula, the loss of rabbits due to myxomatosis, habitat destruction, hunting and subsequently introduced rabbit haemorrhagic disease virus (RHDV) created serious ecological problems by threatening the survival of top predators such as the imperial eagle and the Iberian lynx [55]. The wild European rabbit populations within the native range have declined an estimated 95% since 1950, 80% in Spain since 1975 and 24% in Portugal between 1995 and 2002 [56, 57]. The European rabbit is currently classified in the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (http:// www.iucnredlist.org/) as Near Threatened.

1.2.2. Myxoma virus and European rabbit co-evolution: the Australian and European examples

In contrast to the American continent, in Australia there was no natural reservoir host for MYXV. Consequently, the virus had to persist in the same host species that it was eradicating with high effectiveness. This led to the emergence of less virulent (attenuated) MYXV strains, prolonging the survival time of infected European rabbits with the virus in the epidermis, but ultimately causing the host death, and therefore achieving a more effective transmission of the virus by mosquitoes in the field [13, 58].

At the time of the first release in the Australian field, tests estimated the death of 99.8% of infected rabbits with the SLS strain, which was confirmed in the early epizootics [50]. In one of the sites where the field trials with the MYXV SLS strain was undertaken, the Lake Urana region in southern New South Wales, a mortality rate of 99% and a CFR estimated at 99.8% after the first epizootic in the summer of 1951/1952

were registered. In the second season (summer of 1952/1953), the estimated CFR had dropped to 90%, supporting the selection of attenuated MYXV strains [59]. To standardize virulence testing and based on assays for 92 MYXV strains, Fenner and Marshall (1957) [34] grouped viruses into five virulence grades defined by parameters such as CFR and average survival time (AST) (Table 1). The virulence classification raised subsequent criticism regarding the simplicity of the methods and statistical approaches [60-62], but it also allowed the standardized categorization of multiple field samples and demonstrated the ongoing evolution of MYXV [49].

Table 1 - Virulence grades and prototypes of MYXV (adapted from [34, 49])

Virulance grade	1 (very high)	2	3	4	5 (very low)
CFR (%)	99.5	95-99	70-95	50-70	<50
AST (days)	≤ 13	14-16	17-28	29-50	Not defined
Strains (examples)	SLS[a] Lu[a] MSW[b] MSD[c]	Not defined	KM13[d]	Uriarra[d]	Neuromyxoma[d] Nottingham[e]

[a] South American strain.
[b] North American strain.
[c] North American strain originally described as Grade 1 [34], but used as an attenuated strain by Silvers and colleagues (2006) [35].
[d] Derived from SLS.
[e] Derived from Lu.
CFR – case fatality rate; AST – average survival time.

Since the release of MYXV in Australia in 1950 as a biological control for European rabbit, a complex ongoing co-evolution between virus and host has been occurring. Despite the release of the highly virulent SLS during the first epizootic, some European rabbits survived. Also, the emergence of attenuated strains of MYXV was registered in the next season, increasing the probability that rabbits with a degree of genetic resistance would survive infection. At Lake Urana, in only 7 years, the CFR on challenge with KM13 strain (grade 3) had dropped from 90% to 26% [63, 64]. During the first 30 years of myxomatosis in Australia, grade 3 viruses became the main field strains [65]. At the same time virus attenuation occurred, the wild European rabbits were genetically selected for resistance to MYXV, becoming less susceptible to the virus infection. Australian wild rabbits are now strongly resistant to SLS [66, 67].

In general, the co-evolution between MYXV and wild European rabbit populations in continental Europe and Britain was quite similar to the Australian scenario [54]. Attenuation of MYXV field strains was observed a short period after the initial outbreaks of myxomatosis in both France and Britain [34, 68]. Grade 3 strains were isolated in Britain only 12 months after the introduction of MYXV [34]. Genetic resistance of wild

European rabbits to myxomatosis was identified in Britain in the mid-70s, enabling wild populations to gradually recover [69, 70].

1.2.3. Myxomatosis clinical signs

The course of infection with the highly virulent South American MYXV strains SLS and Lu in disease-susceptible laboratory European rabbits induces a number of classical diagnostic symptoms (Figure 2 A) that culminate with death, normally between days 9 and 12 post-infection (pi) [66, 71]. In the early stage of infection, between days 2 and 4 pi, a pink to red color large primary lesion and edema develops at the inoculation site [66]. Swellings (secondary lesions also called myxomas) occurring on the eyelids, nose, ears, anogenital region and subsequently over the entire body, conjunctival inflammation (blepharoconjuctivitis) and mucopurulent discharge from the nose and eyes are characteristic in the intermediate stage of the disease, between days 4 and 8 pi [66]. The late stage of infection is characterized by increased severity of symptoms, for example, obstruction of the nostrils and closure of the eyelids by the mucopurulent discharge, and acute respiratory distress [66]. Finally, European rabbits may develop secondary infections, such as severe bacterial infections of the conjunctivae and upper respiratory tract, which have been postulated as the probable cause of death [66, 72]. The necrosis of primary skin lesions observed in Lu strain infection has not been described for the SLS strain, making it the only registered clinical sign difference between the two grade 1 virulence strains [71].

Yet, myxomatosis clinical signs and consequent infection outcome depend not only on the virulence of the MYXV strain, but also on the resistance of the European rabbit (Figure 2) [35, 66, 67]. Naturally selected Australian wild rabbits, i.e. myxomatosis-resistant rabbits, inoculated with SLS also develop classical clinical signs of myxomatosis, although with less severity and a delayed onset (Figure 2 A) [66]. Moreover, by day 20 pi these rabbits are virtually recovered, with scabbing of the primary lesion and absence of mucopurulent discharge from nose or eyes [66]. The attenuated Australian Uriarra (Ur) MYXV field strain, when inoculated in susceptible laboratory rabbits, causes clinical myxomatosis, but less severe than that of infection with SLS and with similar symptoms to those seen in wild rabbits infected with SLS (Figure 2 B) [66]. Wild rabbits infected with Ur develop a very mild disease mainly limited to the primary lesion at the inoculation site and very discrete secondary lesions (Figure 2 B) [66].

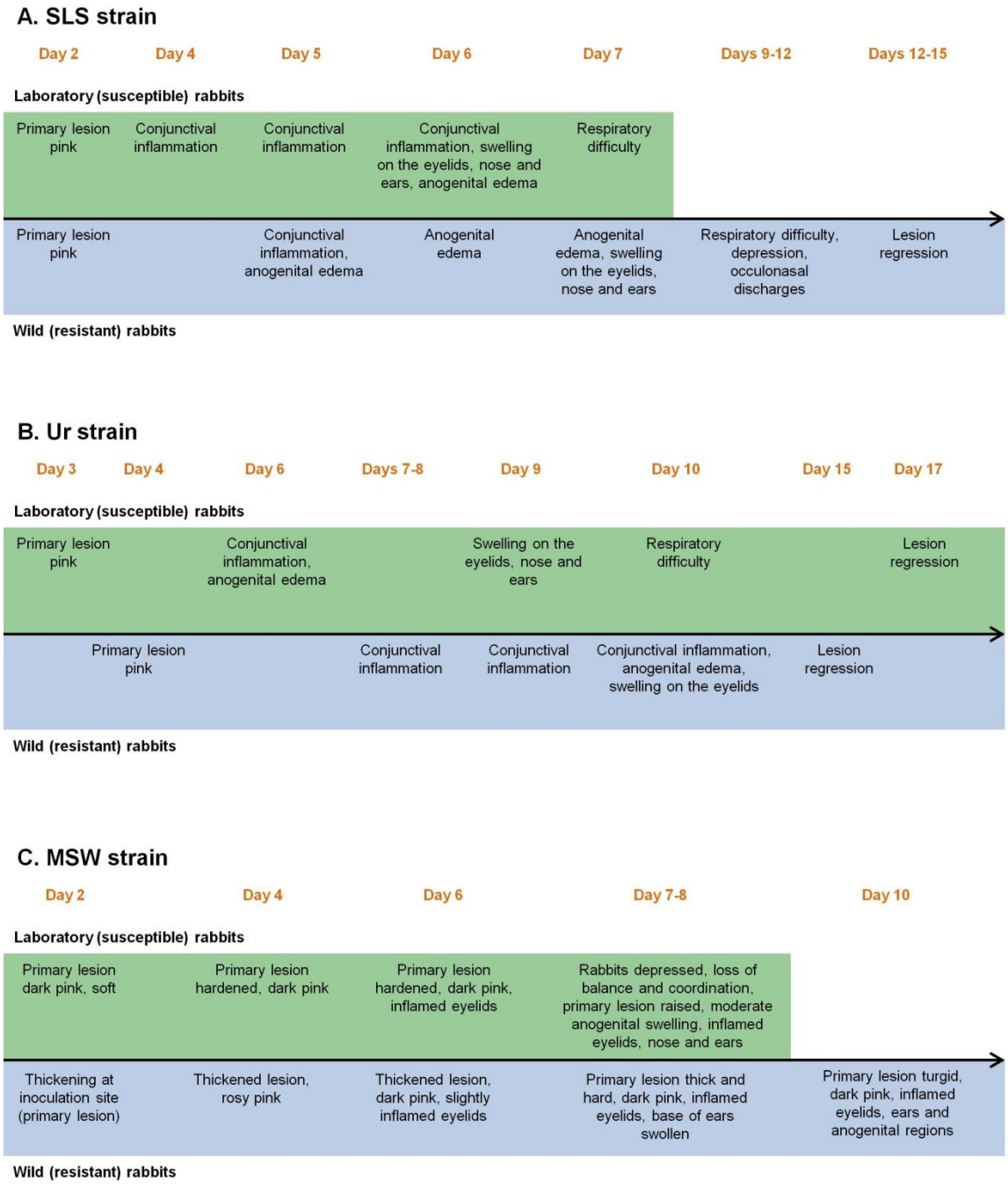


Figure 2 - Myxomatosis clinical signs and progression in laboratory and wild European rabbits infected with different MYXV strains. **A.** SLS strain, **B.** Ur strain and **C.** MSW strain (adapted from [35, 66]).

Infection with the highly virulent North American MSW strain in laboratory and Australian wild rabbits causes 100% mortality [35]. In fact, MSW completely overcomes the resistance in wild rabbits that has been developed against SLS strain, emphasizing once more the importance of both the virulence of the virus and the level of resistance of the rabbit [66, 67]. Severe classical myxomatosis clinical signs are absent from

rabbits infected with MSW, but signs of central nervous system dysfunction, such as muscle twitching, hypersensitivity to stimulation, depression and coma, particularly in laboratory rabbits, are present (Figure 2 C) [35]. In contrast, laboratory rabbits infected with the attenuated North American MSD strain develop classical signs of myxomatosis and no central nervous system dysfunction [35]. North American viruses were considered to be neurotropic, due to the high titres of virus found in the brain and to the described neurological signs [13]. However, high titres of virus were not found in the brain in subsequent studies with MSW strain [35].

Although secondary bacterial infection is often claimed to be the cause of death in acute myxomatosis, the infection is limited to the conjunctivae and upper respiratory tract with minimal pathology in the lungs [4, 49]. Asphyxiation due to nasal closure has also been proposed [72]. Nevertheless, the presented causes have never been satisfactory to fully explain death [4, 72]. For instance, rabbits infected with the highly virulent MSW strain or some Australian field strains die without exhibiting signs of secondary infection [35, 49]. Recently, it has been suggested that massive destruction of lymphoid cells and widespread tissue damage in the skin trigger an overwhelming response by inflammatory mediators, such as cytokines and chemokines, and lethal septic shock occurs [6, 35].

1.2.4. Pathogenesis in laboratory European rabbits

The pathogenesis of MYXV has been extensively described for susceptible laboratory European rabbits after infection with SLS [52, 66, 73, 74] or Lu strains [71], both with similar virus dissemination and replication. Following intradermal inoculation, MYXV initially replicates in the skin at the inoculation site, particularly in MHC-II positive denditric-like cells at the epidermal/dermal junction and deeper in the dermis at 24 hours (Figure 3) [74]. A characteristic feature of MYXV infection is also the presence of virus within endothelial cells of small blood vessels and within large stellate or polygonal cells (termed 'myxoma cells') which appear to bud through or from the endothelium of the blood vessels [73, 74]. Within 24 hours of infection, virus can be detected in the lymph node draining the inoculation site (Figure 3). Here, MYXV replicates to high titres within the lymphoid tissue of the paracortex and cortex, resulting in the massive loss of lymphocytes in the draining lymph node [74]. From the lymph node, MYXV then spreads, probably in lymphocytes and monocytes, to distal tissues such as lungs, testis, spleen and other lymphoid tissues, skin and mucocutaneous sites, such as nose, eyelids and the anogenital region (Figure 3) [52, 66, 74]. After several days of infection, the epithelial cells of the epidermis at the inoculation site and in secondary cutaneous

lesions, such as eyelid and ear myxomas, become packed with virus and are critical for transmission by mosquitoes and fleas (Figure 3) [52].

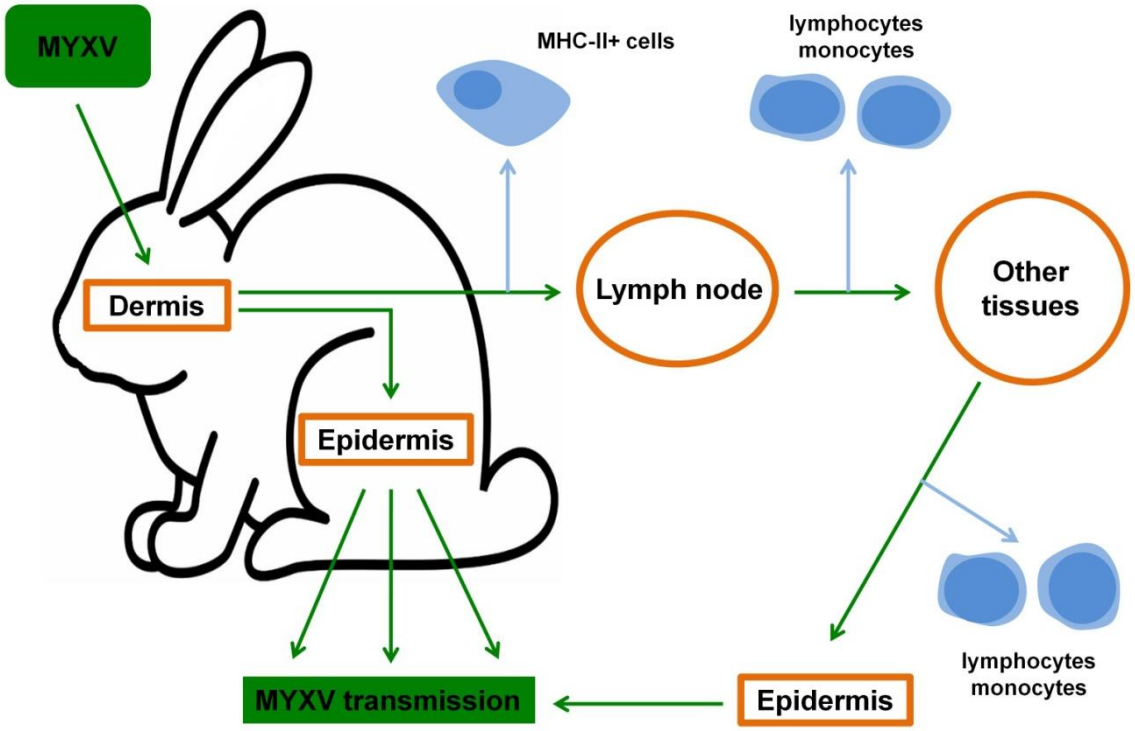


Figure 3 - Pathogenesis of MYXV in laboratory European rabbits.
MYXV is inoculated intradermally by arthropods probing for a blood meal. Initially, the virus replicates in MHC-II positive denditric-like cells in the dermis. Within 24 hours the virus is found in the lymph node draining the inoculation site and replicates within the lymphoid tissue of the paracortex and cortex. From the lymph node, MYXV spreads to other tissues such as lungs, testis, spleen and other lymphoid tissues, skin and mucocutaneous sites, probably within lymphocytes and monocytes. At the same time, MYXV replication at the site of inoculation shifts to the epidermis. Virus in the epidermis in secondary skin lesions or at the inoculation site can then be picked up and transmitted by biting arthropods (adapted from [4]).

The most important sites of pathology are the lymphoid tissue and skin, since dramatic alterations result from the combination of cellular proliferation and destruction [73]. In the lymphoid tissue, there is often complete loss of lymphocytes from both the T cell zones and the follicles, mostly as a result of apoptosis, and all T cell subpopulations (CD4+, CD8+, and CD4+CD8+) decrease [74]. After infection with Lu strain, T cells in lymphoid tissue are more affected than B cells, and within T lymphocytes, the virus shows a more pronounced affinity to the CD4+ lymphocyte subpopulation [71]. MYXV is not found as free virions in the blood, but is detected in the white cell fraction [52]. These observations are consistent with virus replication in MHC-II positive cells and

lymphocytes, which are believed to be critical in the development of innate and adaptive immune responses to MYXV, and also in the dissemination and development of systemic disease and secondary lesions [4, 74]. Besides destroying lymphocytes, MYXV also suppresses lymphocyte and macrophage activation and inflammatory responses [4, 75, 76]. Despite the immunosuppression induced by the virus, after infection with highly virulent SLS or Lu, rabbits still develop IgM and IgG antibodies to MYXV [66, 71]; however, the neutralizing antibodies do not protect rabbits from death [4, 66, 71].

Laboratory European rabbits inoculated with the attenuated Ur strain exhibit initial viral titres similar to those of SLS in the primary lesion site, and tend to be lower in the draining lymph node being cleared between days 10 and 15 pi. However, virus titres become 10-100-fold lower in distal organs, indicating that the rabbit is able to better control attenuated MYXV replication at distal sites [66]. Lymphocyte depletion is not a consistent characteristic of infection with the attenuated strain, despite widespread apoptosis of lymphocytes [66, 74].

1.2.5. Pathogenesis in Australian wild European rabbits: mechanisms of resistance

In Australia, resistance to myxomatosis resulted from the action of natural selection on wild European rabbit populations after the introduction of MYXV in the 1950s. Nevertheless, wild rabbits are not fully resistant to infection (described in section 1.2.3.) since they exhibit a similar initial pathogenesis to laboratory rabbits; however, selection for an enhanced immune response to MYXV leading to the control of infection and recovery could explain resistance. Indeed, at the primary lesion site, although the virus is present in high titres, a pronounced mononuclear cell infiltrate has been observed, as well as a prominent inflammatory response with strong expression of nitric oxide synthase, characteristic of macrophage activation [49, 74-76].

It has been suggested that the draining lymph node is critical for the development of a T cell response for controlling virus replication in distal tissues [66]. In fact, when infected with SLS, the viral replication in wild rabbits is constrained in draining and distal lymph nodes, spleen and lung, with titres 10-100-fold lower in wild rabbits than in laboratory rabbits [66], a probable consequence of both innate nonspecific and cell-mediated responses specific for MYXV. Also, despite the depletion of lymphocytes from the lymphoid tissue, lymph nodes recover by day 15 pi [74].

1.3. Molecular basis of myxoma virus pathogenesis

Poxvirus genomes are divided in two basic classes of genes: the highly conserved and centrally located genes associated with housekeeping functions, such as virus structure and assembly, gene transcription, DNA replication, cell entry and exit; and terminally located genes encoding proteins important for suppression and evasion of host innate and adaptive anti-viral responses [4, 49, 77]. For MYXV, these virulence-related genes, often referred to as immunomodulatory genes, have co-evolved with the *Sylvilagus* natural host, while in *Oryctolagus*, the subversion of the immune response by these proteins is purely accidental [4, 49]. Twenty-six of these MYXV genes (Lu strain), encoding proteins with known roles in modulating the host immune response, are described in Table 2.

The function of these modulators has been assessed *in vivo* and/or *in vitro*, by infecting laboratory European rabbits and specific cell lines, respectively, with mutant viruses. Immunomodulatory proteins can be grouped according to their targets and mode of action, and a classification into viroreceptors, virokines, anti-apoptotic factors, immune modulators and host range factors is presented here [reviewed in 4, 6, 47, 49, 78]; however, for several of these proteins, more than one function has been discovered [79]. Some of these viral proteins are rabbit-specific, while others target the immune response from different hosts, including mouse and human [47]. A brief description of the most important immunomodulatory proteins is included next.

1.3.1. Viroreceptors

When cells are infected with MYXV, viroreceptors are often secreted or expressed on the surface, binding and inhibiting extracellular host ligands that are intended to induce an inflammatory or anti-viral response following virus infection. M002, a tumor necrosis factor (TNF) receptor homologue [80], and M007, an interferon-gamma (IFN-γ) receptor homologue [81, 82], are two European rabbit-specific MYXV viroreceptors fundamental for the regulation of the external environment of infected cells. Another MYXV protein, M001, exhibits binding properties that sequester soluble host chemokine from binding their cognate cell-surface receptors in a species-independent manner [83].

Table 2 - Role and impact on virulence of MYXV proteins that modulate the host response to infection (adapted from [49])

Gene[a]	Protein function ([b] [c])	Effect on virulence of knockout	Reference
M001L/R	Chemokine binding (260; E)	Generalized myxomatosis; 1/6 survived	[85]
M002L/R	TNF binding; anti-apoptosis (326; E)	Moderate to severe myxomatosis; 5/8 animals survived	[80, 86-88]
M004L/R	RDEL motif; anti-apoptosis (237; E)	Small rapidly resolved primary lesions; 1/8 rabbits had a secondary; all animals recovered	[89, 90]
M005L/R	Host-range; anti-apoptosis; E3 Ub ligase (483; E)	Primary lesion only; rapid resolution; no signs of clinical myxomatosis	[91]
M007L/R	Secreted IFN-γ binding protein; chemokine binding (263; E)	12/13 rabbits mild to moderate disease; lymphocyte infiltration	[81, 82, 92, 93]
M008.1L/R	Serp 1; secreted serine proteinase inhibitor (369; L)	Moderate to severe generalized myxomatosis; 5/8 rabbits recovered from infection; enhanced inflammatory response	[94]
M010L	MGF; epidermal growth factor homologue (85; E)	Moderately severe generalized myxomatosis; 25% of animals became moribund; 75% recovered	[95]
M011L	Anti-apoptotic factor (166; E)	All rabbits survived; large protuberant demarcated primary; large secondaries; mild conjunctivitis/rhinitis	[95, 96]
M013L	Pyrin domain inflammasome; NFκB inhibition (126; E)	Mild clinical signs rapidly resolved; small secondaries; no mortality; rapid inflammatory response	[97-100]
M029L	Type I interferon resistance/PKR inhibition; conscripts pro-viral RHA/DHX9 (115; E)	No signs of primary; 1/4 rabbits exhibited a small red, swollen area and were protected from disease after re-challenge	[101, 102]
M062R	Host range (158;E/L)	Abortive infection in rabbits and rabbit cells	[103]
M063R	Host range (215; E)	No virus replication in rabbits and rabbit cells	[104]
M064R	Poxvirus C7L family member, but lacks host range functions; virion component (203; E/L)	Delayed development of disease; no survival	[105]
M128L	CD47 homologue; macrophage inhibition (281; L)	Mild generalized disease; rapid resolution; no mortality	[76]
M130R	Unknown function; localized to ER/ Golgi (122; L)	Clinical generalized myxomatosis but no deaths	[106]
M131R	Superoxide dismutase inhibition (163; L)	All animals euthanized days 10–11; RFV is attenuated	[107-109]
M135R	Immunomodulatory (178; E)	All animals survived; mild disease with little generalization	[110]
M138L	Sialyltransferase (290; E)	Severe fatal myxomatosis; survival time prolonged	[111]
M141R	OX-2 (CD200) homologue (218; E)	Mild generalized disease, rapid resolution, all survived; increased macrophage and T cell activation	[75]
M148R	Ankyrin repeat; putative E3 Ub ligase (675; L)	Moderate generalized myxomatosis; 2/5 rabbits euthanized at 21 days; mononuclear inflammatory response	[112, 113]
M149R	Ankyrin repeat; putative E3 Ub ligase (490; E/L?)	Moderate generalized myxomatosis with delayed secondaries; 5/5 rabbits survived	[112, 113]
M150R	NFκB inhibition; E3 Ub ligase (494; E)	Rapid inflammatory response at primary site; few small secondaries; 12/12 recovered by day 21	[113-115]
M151R	Serp 2 (333; E)	Local primary but few or no secondary lesions; 7/10 infected rabbits recovered; 3/10 euthanized – respiratory disease	[116]
M152R	Serp 3 (266; L)	4/10 infected rabbits recovered; 6/10 euthanized because of respiratory disease; no secondary lesions	[117]
M153R	MHC downregulation; E3 Ub ligase (206; E)	Generalized myxomatosis; 4/12 rabbits euthanized day 14, the remainder recovered	[118-120]
M156R	Interferon resistance; eIF2α homologue (102; L)	Not tested	[121]

[a] Direction of gene transcription indicated by L or R, while genes duplicated in TIR are identified by L/R.
[b] No. of amino acids.
[c] Transcription time: early (E), late (L).

The chemokine viroreceptor M001 is secreted from infected cells both early and late in infection. This viral protein interacts with a large spectrum of CXC- and CC-chemokine subfamilies, but shows a higher avidity for CC-chemokines [84]. By binding and sequestering CC-chemokines, M001 can interrupt the chemoattractant gradient and disrupt the trafficking of effector cells toward the site of viral infection [83].

M002 specifically binds to European rabbit TNF with high affinity and inhibits receptor binding of this cytokine [80, 122]. The M002 protein is secreted as both a monomer and a dimer, although the dimeric form is more effective at inhibiting cell lysis by rabbit TNF [123]. This protein is essential for virus virulence. European rabbits infected with a mutant virus in which both copies of *M002L/R* gene had been deleted had a lower mortality rate (40%) compared to European rabbits infected with the parental MYXV (100%). European rabbits infected with the mutant virus also had a dramatic reduction in clinical signs of myxomatosis [80].

M007 is a protein that directly modulates the European rabbit IFN-γ response, a key mediator of the anti-viral response, by competing with the IFN-γ receptor to bind IFN-γ with high affinity [81, 82]. M007 is the most abundantly secreted protein from MYXV-infected cells at $>10^7$ molecules per hour and is present at both early and late times after infection [82]. With the deletion of both copies of *M007L/R*, the mutant virus was highly attenuated and clinical signs were significantly reduced in the infected European rabbits, when compared to those infected with the wild-type virus, demonstrating that M007 is a critical virulence factor for MYXV pathogenesis [92]. M007 also interacts with CC-, CXC-, and C- chemokines via heparin binding domains, thus potentially disrupting the establishment of chemokine gradients by inhibition of chemokine-glycosaminoglycan binding [93].

1.3.2. Virokines

Virokines are viral proteins secreted from infected cells that have the ability to mimic host immune system ligands or growth factors, although usually the viral versions are smaller and exhibit additional biological properties [124].

Serp 1 is encoded by the MYXV *M008.1L/R* gene and functions as an irreversible inhibitor of serine proteases [125]. European rabbits infected with MYXV deficient in *M008.1L/R* exhibited moderate clinical signs of myxomatosis when compared to European rabbits infected with the wild-type virus. Besides being a crucial virulence factor, Serp 1 is also important in inhibiting inflammation at the site of virus replication by preventing the infiltration of monocytes into the primary lesion in infected European rabbits [94].

The *M010L* gene encodes myxoma growth factor (MGF), a secreted glycoprotein with high sequence homology to members of the epithelial growth factor (EGF) family of proteins. M010 is also an essential virulence factor. Viral replication was severely attenuated in European rabbits infected with MYXV deficient in the *M010L* gene and the majority of the rabbits never became seriously ill [95].

1.3.3. Anti-apoptotic factors

The interaction between MYXV anti-apoptotic immunomodulatory proteins and the host apoptotic pathway is fundamental for successful viral replication in European rabbit. These viral proteins manipulate and inhibit host-cell apoptotic responses during early stages of viral infection [126, 127]. Several MYXV anti-apoptotic factors have been identified, including M002, M004, M005, M011 and M151.

In addition to its function in binding and inhibiting European rabbit TNF, the intracellular form of M002 also plays a key role in blocking the induction of apoptosis in lymphocytes infected with MYXV [86, 87, 128].

The immunomodulatory protein M004 specifically localizes in the endoplasmic reticulum (ER) of infected cells. Its anti-apoptotic role was determined when cultured European rabbit lymphocytes infected with MYXV deficient in *M004L/R* underwent extensive cellular apoptotic response [89]. *In vivo*, infection with the mutant MYXV resulted in disease attenuation, probably due to the low number of infected T lymphocytes available to carry the virus to distal lymph nodes in European rabbit [89].

M005 is a critical host range factor that inhibits apoptosis of infected lymphocytes. *In vitro*, T lymphocytes infected with MYXV deficient in *M005L/R* underwent extensive apoptosis together with a rapid inhibition of both host and viral genes [91]. In European rabbits infected with the same mutant virus, infection did not progress beyond the primary site of inoculation, and a rapid and effective inflammatory response resulted, demonstrating that M005 is an essential virulence factor for disease progression of myxomatosis [91].

M011 is also a critical virulence factor for the development of the disease in European rabbit, since the deletion of the *M011L* gene totally abrogated the ability of the virus to cause the classical clinical signs of myxomatosis [95]. In European rabbit T lymphocytes and primary monocytes infected with *M011L* knockout virus, cells underwent apoptosis [86, 129]. This viral anti-apoptotic protein is a structural homologue of Bcl-2, which prevents the loss of mitochondrial membrane potential and interacts with the cell death host proteins Bak and Bax [130-133].

The *M151R* gene encodes the intracellular viral serpin Serp 2, an essential MYXV virulence factor, since European rabbits infected with *M151R* knockout virus only presented moderate clinical signs of myxomatosis [116]. In these rabbits, a rapid apoptosis was observed in the lymph node lymphocytes, supporting the critical role of Serp 2 in the inhibition of apoptosis in lymphocytes and allowing their consequent spread to secondary sites of infection [116].

1.3.4. Immune modulators

The subversion and manipulation of the host anti-viral response by MYXV strategically operates on multiple fronts. Besides inhibiting cell death, TNF and IFN-γ signaling, and the formation of chemokine gradients and inflammatory cascades, MYXV immune modulators also inhibit the activation of macrophages and T cells, pattern-recognition receptors (PRRs) pathways and type I IFN signaling.

Viral protein M153 is responsible for MYXV downregulation of MHC-I expression on the surface of infected cells, as MHC-I presentation of viral proteins is critical for clearance of virus by the adaptive immune system [118]. M153 also promotes the downregulation of both surface Fas/CD95 and the T cell co-receptor molecule CD4, as well as the downregulation of CD166, all potentially interfering with T cell responses [118, 119, 134]. Deletion of the *M153R* gene significantly attenuates MYXV infection in European rabbit [118]. Other viral proteins are involved in inhibiting T cell activation, important to produce crucial mediators at the site of infection and to directly kill MYXV infected cells. M141, a homologue of CD200 expressed on the cell surface, decreases T cell activation since it acts as a negative regulator of macrophage activation; the recruitment and activation of monocytes/macrophages and lymphocytes greatly increased in lymphoid organs infected with the *M141R* mutant MYXV [75]. Also, infected cell surface expression of the viral CD47 homologue, MYXV protein M128, appears to prevent macrophage activation *in vivo* [76].

The host germline-encoded PRRs, such as Toll-like receptors (TLRs) and RIG-I-like receptors (RLRs), recognize and react with pathogen-associated molecular patterns (PAMPs), which include viral nucleic acids in infected cells [e.g. 135-137]. PRR signaling leads to the stimulation of gene transcription and to the expression of interferons (IFNs), and triggers inflammasome activation as a mediator of proinflammatory responses. Moreover, the NFκB pathway is critically involved in the control of responses to viral infection recognized by PPRs [138]. MYXV M013 binds host ASC-1 protein and inhibits caspase 1 activation in the host inflammasome pathway, preventing caspase 1-mediated cleavage of pro-interleukin-1β (pro-IL-1β) and

pro-interleukin-18 (pro-IL-18) to release the mature forms of these proinflammatory cytokines [97]. Also, the expression of these cytokines is prevented by M013 binding to NFκB1/p105 in the NFκB pathway [99]. *M013L* mutant MYXV-infected European rabbits did not develop myxomatosis, demonstrating that M013 is a critical virulence factor [97]. M150 co-localizes with NFκB in the nucleus of cells infected with MYXV and treated with TNF, and therefore the protein was named myxoma nuclear factor (MNF). This factor is also critical for MYXV virulence as demonstrated in *in vivo* studies [114].

Besides the rabbit-specific IFN-γ receptor homologue M0 0 7, MYXV expresses other proteins that modulate IFNs, which are key mediators of anti-viral responses. The action of type 1 IFNs (IFN α/β) is inhibited by MYXV-encoded protein M156, a structural mimic of the cellular eukaryotic translation initiation factor 2 alpha (eIF2α) and a viral pseudosubstrate for type 1 IFN-induced, double-stranded RNA-activated protein kinase R (PKR) [121]. Binding of PKR to double-stranded RNA (dsRNA) produced during a viral infection leads to the dimerization, trans-autophosphorylation, and consequent activation of the protein kinase domain; activated PKR promotes phosphorylation of eIF2α, resulting in the inhibition of both cellular and viral protein synthesis [139]. On the other hand, M029 antagonizes PKR-mediated anti-viral responses by binding dsRNA, but additionally, this viral protein also binds and conscripts RNA helicase A (RHA)/DHX9 as a pro-viral effector to promote MYXV replication in a cell-specific manner [101, 102]. MYXV deficient in *M029L* failed to cause any symptoms of myxomatosis in susceptible European rabbits, showing that M029 is a critical virulence factor [102].

1.3.5. Host range factors

The virus-encoded proteins essential for the biologic tropism and host range of a specific virus are called host range factors [140]. The MYXV repertoire of host range factors target specific intracellular pathways to create a favorable environment within the infected cells for viral replication and to block the induction of anti-viral responses in rabbit lymphocytes, particularly through anti-apoptotic proteins [140].

The anti-apoptotic factor M005 is an ankyrin-repeat-containing protein related to other poxvirus host range factors [91]. Although MYXV is a rabbit-specific virus, it has been demonstrated that M005 is crucial in MYXV tropism in human cancer cells [141].

Recently, it has been demonstrated that double-function M029 viral protein is a critical host range factor for MYXV replication in rabbit cells and in a large variety of different non-lagomorph mammalian cells, such as humans, non-human primates and mouse [102].

M062R and *M063R* are members of the C7L family of host range genes from orthopoxviruses. European rabbits infected with MYXV deficient in *M062R* failed to develop any classical clinical signs of myxomatosis and all rabbit cell lines infected with the mutant virus underwent abortive infection [103]. Similarly, *M063R* MYXV knockout did not replicate in either infected European rabbits or tested rabbit cell lines [104]. Nevertheless, both mutant viruses still productively infected certain human cancer cells [103, 104]. Therefore, M062 and M063 proteins are host range factors crucial to control productive MYXV replication in rabbit cells and in various human cells. It is also known that both form a heteromeric complex during viral infection that binds and inhibits the cellular anti-viral factor sterile alpha motif domain-containing protein 9 (SAMD9) in human cells [103].

1.4. Evolution and genetic diversity of myxoma virus

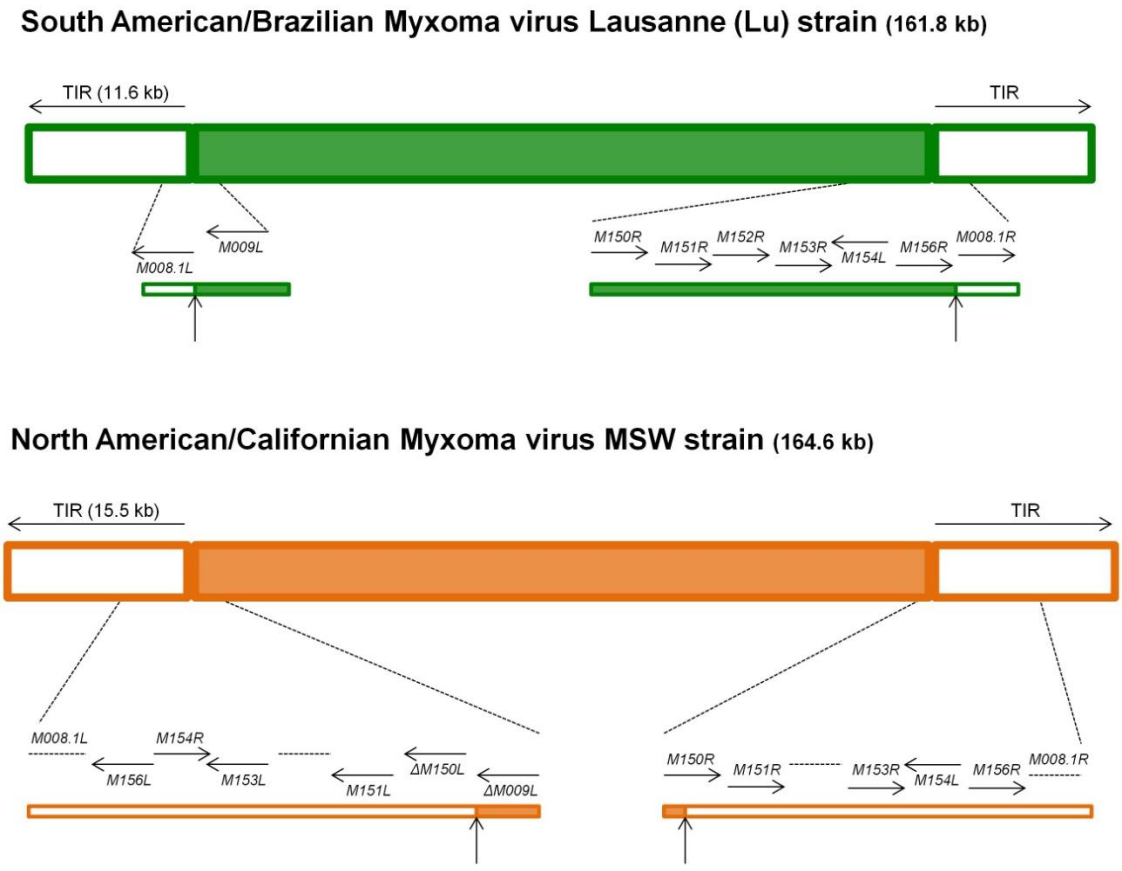
Only recently a genome-scale study shed some light on the evolutionarily complex dynamics of MYXV, covering in real-time the parallel Australian and European epidemics [142]. With a unique dataset, almost 50 years of viral evolution and the full range of virulence grades were studied. The first striking observation was the strong evidence for the rapid evolution of MYXV in both continents, supported by a rate of nucleotide substitution of ~$1x10^{-5}$ substitutions/site/year, one of the highest ever reported for a dsDNA virus or other poxviruses [143-145], with the exception of variola virus, for which a similar substitution rate has been reported [144]. By genetically characterizing strains with defined phenotypes, (i.e. the grade of virulence), some notable observations were achieved [142]. Firstly, changes in virulence involved multiple genes, likely losses of gene function due to insertion-deletion events [142]. For example, for the highly attenuated Ur strain, a C nucleotide insertion in *M005L/R* causing a disruption of the reading frame was identified; since M005 is critical for MYXV virulence by manipulating cell cycle progression and inhibiting cell death [91, 146], this indel is likely the main mutation responsible for the strain attenuation. Secondly, there are multiple genetic routes to attain either highly virulent or attenuated phenotypes (phenotypic convergence), and the pattern of virulence evolution is never associated with genotypic convergence [142]. There are no coding mutations common to specific virulence grades. No mutations, for example, are uniquely shared by three grade 1 viruses isolated in the 1990s [142].

Studies of genetic variation of different MYXV strains using the restriction fragment length polymorphism (RFLP) technique revealed very few differences between SLS derived strains, including Ur, and the Lu strain [147]. The degree of

genetic alteration in Australian field isolates was also very limited [148]; however, the limitations of the RFLP technique were noted [148]. Yet, in the recent genome-wide study on MYXV evolution, phylogenetic analysis revealed a division between the viruses released in Australia (SLS progenitor) and Europe (Lu progenitor) with a strong temporal structure [142].

The genetic characterization of field strains isolated in Europe, particularly in the Iberian Peninsula, has also been performed. In Spain, a specific field isolate, the MYXV strain 6918, was selected for further studies due to its pathogenicity and immunogenic potential for use as a vaccine against myxomatosis [149], and interestingly, although isolated 43 years after the introduction of Lu strain in Europe, the genomes were 99.95% identical [150]. These data were supported by other molecular studies on recent field strains isolated both in Portugal and Spain, where low levels of genetic variability were found [151, 152].

All the evolutionary and genetic characterizations of MYXV have been performed in South American/Brazilian-derived strains. Only recently, the complete genome of the North American/Californian MSW strain has been sequenced [7], which allowed for the first time a comparative analysis between the two geographically distinct types of MYXV. The MSW genome (164.6 kb) is larger than the Lu (161.8 kb), the first strain with complete genome sequencing [5] and which has been extensively used in *in vivo* and *in vitro* gene knockout studies. In the MSW genome, an expansion of the TIRs was observed, with duplication of the *M156R*, *M154L*, *M153R*, *M152R*, *M151R* and part of the *M150R* genes from the right-hand end of the genome at the left-hand TIR, resulting in the size difference between strains (Figure 4) [7]. The gene order in the MSW strain is identical to that of Lu, with the exception that the *M000.5L/R* open reading frames are absent at the extreme ends of MSW TIRs. Also, five genes were disrupted by multiple indels or mutations to the ATG start codon: *M008.1L/R*, *M009L*, *M023R*, *M131R* and *M152R* genes [7]. Despite all these differences, no novel genes were identified in MSW genome [7].



.

Figure 4 - Comparison between Lausanne (Lu) and MSW terminal inverted repeats (TIR) regions.

The gene order around the TIR boundary at the left-hand and right-hand of the genome for Lu and MSW are represented. The vertical arrows indicate the TIR junctions. In MSW, a dashed representation instead of an arrow corresponds to a sequence that does not encode an ORF (*M008.1L/R* and *M152L/R*); *M008.1L/R* is labeled in the diagram for clarity and the dashed line between *M151L/R* and *M153L/R* matches to *M152L/R* sequence. ΔM150L and ΔM009L represent the truncated forms of these ORFs in MSW. The figure is not drawn to scale (adapted from [7]).

2. Viral evasion by host restriction factors

2.1 General characteristics of host restriction factors

The components of the host innate immune system and the viral factors that have evolved to evade or destroy these defense barriers determine, in part, the host susceptibility to viral infection. These potent host-encoded gene products include the so-called restriction factors, which are naturally selected for the ability to inhibit the replication of viruses during their life cycle in host cells [153, 154]. The majority of the host restriction factors share essential, but not universal, characteristics, such as anti-viral activity as their major biological function, induction by IFNs or by virus infection, antagonism by a viral protein and the hallmarks of evolutionary selective pressures [153, 155, 156]. Table 3 contains the main characteristics of some of the best-studied examples of restriction factors.

There are restriction factors with the ability to target a wide range of unrelated viruses, such as tetherin or MxA/Mx1 (Table 3). On the other hand, there are factors that appear to be viral genus- or species-specific, such as Fv4, which restricts murine leukemia virus (MLV) [157], and the sheep-specific enJSRV, which neutralizes Jaagsiekte sheep retrovirus (JSRV) [158, 159]. Nevertheless, several other restriction factors target numerous viruses within a particular family, like the tripartite motif-containing protein 5 alpha (TRIM5α) proteins, which restrict diverse retroviruses [160-163].

Viruses have also evolved strategies to circumvent the action of restriction factors by producing viral antagonists, which are often encoded by 'accessory genes' [reviewed in 156, 164-172]. Restriction factors that target several virus families, like tetherin, can be neutralized by diverse viral proteins (Table 3). However, no known viral antagonists have been identified for some restriction factors. In these cases, viral escape occurs through mutation of the viral protein restricted by the host factor. The most well-known case is lentiviral evasion of TRIM5α-mediated restriction through viral capsid amino acid changes [173-175].

2.2 Host restriction factors as the paradigm of virus-host co-evolution

2.2.1 Positive selection

Of all the defining characteristics, the one that consistently applies to most host restriction factors is evidence of natural selection signatures, and more specifically, of strong positive selection (Table 3) [153, 155, 156]. Since host restriction factors naturally

conflict with viral proteins, an excess of fixed non-synonymous substitutions (d_N) compared with synonymous substitutions (d_S) is observed, even in a background of negative (purifying) selection essential to constraint the protein function. Thus, a ratio of the rate of non-synonymous substitutions to the rate of synonymous substitutions (d_N/d_S) with a value greater than one is accepted as presumptive evidence for positive selection [176, 177].

Table 3 - Characteristics of some well-studied examples of restriction factors (adapted from [153, 154])

Restriction factor	Viral targets	Viral antagonists	Under positive selection?
Fv4	MLV [157]	Unknown	Not determined
enJSRV	JSRV [158, 159]	Unknown	Yes [178]
Fv1	Retroviruses [179]	Unknown	Yes [180]
TRIM5α and TRIM-CYP	Retroviruses [160-163]	Unknown (escape through capsid mutations)	Yes [181]
APOBEC3 family	Retroviruses [182], retrotransposons [183], hepadnaviruses [184]	Vif (lentiviruses), Bet (spumaviruses), Gag (gammaretroviruses)	APOBEC3DE [185], APOBEC3G [186, 187], APOBEC3H [188]
SAMHD1	Retroviruses [189, 190]	Vpx (some SIVs), Vpr (some SIVs)	Yes [191, 192]
ZAP	Retroviruses [193, 194], filoviruses [195], alphaviruses [196]	Unknown	Yes [197]
Tetherin	Retroviruses [198, 199], flaviviruses [200], herpesviruses [201], rhabdoviruses [202], paramyxoviruses [203], arenaviruses [204]	Nef (some SIVs), Vpu (HIV-1), Env (HIV-2), glycoprotein (Ebola virus), K5 (KSHV)	Yes [205-207]
Viperin	Orthomyxoviruses [208], flaviriruses [209], herpesviruses [210], togaviruses [211], paramyxoviruses [212]	Unknown	Yes [213]
MxA/Mx1	Orthomyxoviruses [214], paramyxoviruses [215], hepadnaviruses [216], rhabdoviruses [214], bunyaviruses [217], togaviruses [218]	Unknown	Not determined
IFITM1, IFITM2 and IFITM3	Orthomyxoviruses, flaviviruses, coronaviruses [219, 220]	Unknown	Not determined
PKR	Poxviruses [221]	K3L and E3L (VV), TRS1 and IRS1 (HCMV), and many others	Yes [222, 223]

HCMV, human cytomegalovirus; HIV, human immunodeficiency virus; IFITM, interferon-induced transmembrane protein; JSRV, Jaagsiekte sheep retrovirus; KSHV, Kaposis's sarcoma-associated herpesvirus; PKR, interferon-induced, double-stranded RNA-activated protein kinase R; SAMHD1, SAM domain- and HD domain-containing protein 1; SIV, simian immunodeficiency virus; TRIM, tripartite motif-containing protein; VV, vaccinia virus ; ZAP, zinc-finger anti-viral protein.

In traditional methods, the estimation of d_N and d_S between two protein-coding DNA sequences consists in counting non-synonymous and synonymous sites in the two sequences, counting the number of non-synonymous and synonymous substitutions between the two sequences and correcting for multiple substitutions [177]. Although these methods are intuitive and robust, they are also over-simplistic and ignore the effect of both the transition/transversion rate and codon-usage biases, despite the effort of some later methods to integrate these [224-229]. Much more sensitive maximum

likelihood-based methods were developed for the estimation of d_N and d_S between two sequences, allowing also the identification of individual positively-selected codons [177, 230, 231]. To perform these codon-based methods, an alignment of several orthologous gene sequences and a phylogenetic tree describing the evolutionary relationship between the analyzed species are required. Also, statistical tests, such as the likelihood ratio test (LRT), can be performed to determine whether d_N is significantly higher than d_S [177].

2.2.2 Arms race: virus-host co-evolution

Most of the genes encoding cellular host restriction factors are engaged in an evolutionary arms race dynamics as a result of the genetic conflict between the host and the virus (Figure 5). The host becomes susceptible to infection when a restriction factor is antagonized by a viral protein. This will exert a selective pressure on the host, where mutations that allow the restriction factor to escape a viral antagonist will provide the host a fitness advantage. In a cyclical process, the host restriction factor now exerts selective pressure on the viral antagonist, creating a disadvantageous condition for the virus. In response, new mutations that allow the virus to evade host restriction factor action are selected in the pathogen [153, 154]. This back-and-forth process between the interacting virus and host, leading to the rapid evolution of both, is an example of the 'Red Queen hypothesis', the evolutionary law proposed by Leigh Van Valen in 1973 [232].

The known existence of viruses throughout the course of vertebrate evolution and the rapid rate at which virus populations change [e.g. 233-235] are revealing of the selective pressure that host restriction factors have been subjected to for many millions of years. Therefore, this long-term arms race imposes selection on host restriction factors, not by a single virus, but by the many different viruses they have encountered over evolutionary periods [153, 154].

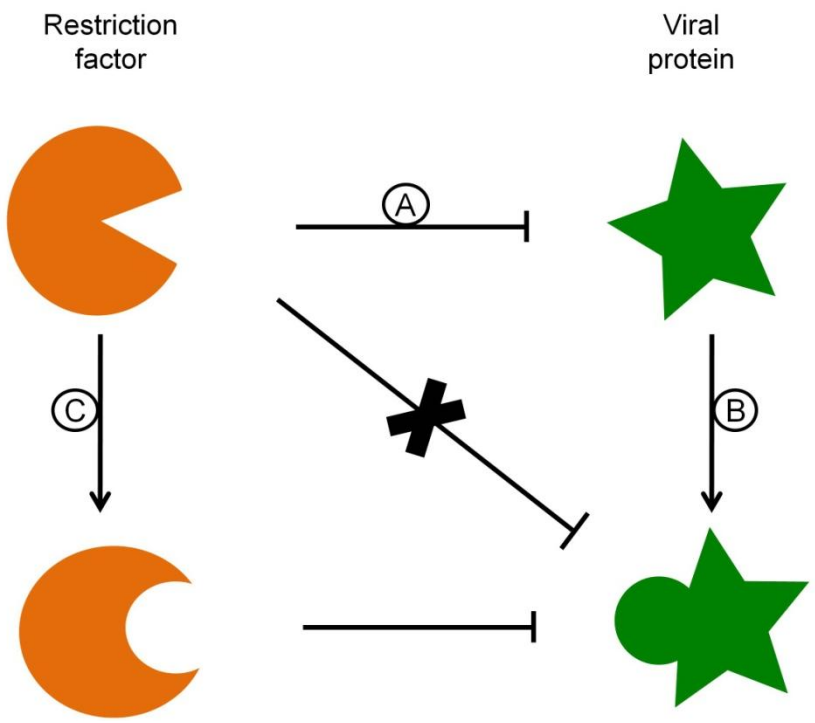


Figure 5 - Genetic conflict between a host restriction factor and its viral target.

A. The host restriction factor recognizes a specific feature of the viral protein, exerting selective pressure and causing a disadvantageous condition for the virus. **B.** Consequently, new mutations that allow the viral protein to evade the host restriction factor action are selected. **C.** In the host restriction factor new mutations that re-establish the interaction with the viral protein will be favored.

3. References

1. Fenner F: Portraits of viruses: the poxviruses. *Intervirology* 1979, 11:137-157.
2. Farrant JL, Fenner F: A comparison of the morphology of vaccinia and myxoma viruses. *Australian Journal of Experimental Biology & Medical Science* 1953, 31:121-125.
3. Moss B: Poxviridae: The viruses and their replication. In *Fundamental Biology.* 3rd edition. Edited by Fields BN, Knipe DM, Howley PM. Philadelphia: Lippincott-Raven Publishers; 1996.
4. Kerr P, McFadden G: Immune responses to myxoma virus. *Viral Immunology* 2002, 15:229-246.
5. Cameron C, Hota-Mitchell S, Chen L, Barrett J, Cao JX, Macaulay C, Willer D, Evans D, McFadden G: The complete DNA sequence of myxoma virus. *Virology* 1999, 264:298-318.
6. Stanford MM, Werden SJ, McFadden G: Myxoma virus in the European rabbit: interactions between the virus and its susceptible host. *Veterinary Research* 2007, 38:299-318.
7. Kerr PJ, Rogers MB, Fitch A, Depasse JV, Cattadori IM, Hudson PJ, Tscharke DC, Holmes EC, Ghedin E: Comparative analysis of the complete genome sequence of the California MSW strain of myxoma virus reveals potential host adaptations. *Journal of Virology* 2013, 87:12080-12089.
8. Kilham L, Herman CM, Fisher ER: Naturally occurring fibromas of grey squirrels related to Shope's rabbit fibroma. *Proceedings of the Society for Experimental Biology and Medicine* 1953, 82:298-301.
9. O'Connor DJ, Diters RW, Nielsen SW: Poxvirus and multiple tumors in an eastern gray squirrel. *Journal of the American Veterinary Medical Association* 1980, 177:792-795.
10. Terrell S, Forrester D, Mederer H, Regan T: An epizootic of fibromatosis in gray squirrels (*Sciurus carolinensis*) in Florida. *Journal of Wildlife Diseases* 2002, 38:305-312.
11. Shope RE: A transmissible tumor-like condition in rabbits. *The Journal of Experimental Medicine* 1932, 56:793-802.
12. Marshall ID, Regnery DC: Myxomatosis in a California brush rabbit (*Sylvilagus bachmani*). *Nature* 1960, 188:73-74.
13. Fenner F, Ratcliffe F: *Myxomatosis.* Cambridge, UK: Cambridge University Press; 1965.

14. Karstad L, Thorsen J, Davies G, Kaminjolo J: Poxvirus fibromas on African hares. *Journal of Wildlife Diseases* 1977, 13:245-247.
15. Chapman JA, Flux JEC: Chapter 1: Introduction and Overview of the Lagomorphs. In *Rabbits, Hares and Pikas - Status survey and conservation action plan.* Edited by Chapman JA, Flux JEC. Gland, Switzerland: IUCN; 1990: 1-6.
16. Angermann R, Flux JEC, Chapman JA, Smith AT: Chapter 2: Lagomorph Classification. In *Rabbits, Hares and Pikas - Status survey and conservation action plan.* Edited by Chapman JA, Flux JEC. Gland, Switzerland: IUCN; 1990: 7-13.
17. Matthee CA, van Vuuren BJ, Bell D, Robinson TJ: A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. *Systematic Biology* 2004, 53:433-447.
18. Branco M, Ferrand N, Monnerot M: Phylogeography of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome b gene. *Heredity (Edinb)* 2000, 85 Pt 4:307-317.
19. Branco M, Monnerot M, Ferrand N, Templeton AR: Postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA genetic variation. *Evolution* 2002, 56:792-803.
20. Geraldes A, Rogel-Gaillard C, Ferrand N: High levels of nucleotide diversity in the European rabbit (*Oryctolagus cuniculus*) SRY gene. *Animal Genetics* 2005, 36:349-351.
21. Geraldes A, Ferrand N, Nachman MW: Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics* 2006, 173:919-933.
22. Geraldes A, Ferrand N: A 7-bp insertion in the 3' untranslated region suggests the duplication and concerted evolution of the rabbit SRY gene. *Genetics Selection Evolution* 2006, 38:313-320.
23. Ferrand N, Branco M: The evolutionary history of the European rabbit (*Oryctolagus cuniculus*): major patterns of population differentiation and geographic expansion inferred from protein polymorphism. In *Phylogeography of Southern European Refugia.* Edited by Weiss S, Ferrand N: Springer Netherlands; 2007: 207-235.
24. Carneiro M, Ferrand N, Nachman MW: Recombination and speciation: loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics* 2009, 181:593-606.

25. Aragão HDB: Myxoma dos coelhos. *Memórias do Instituto Oswaldo Cruz* 1927, 20:225-247.
26. Aragão HDB: O virus do mixoma no coelho do mato (*Sylvilagus minenses*), sua transmissão pelos *Aedes scapularis* e *aegypti*. *Memórias do Instituto Oswaldo Cruz* 1943, 38:93-99.
27. Marshall ID, Regnery DC, Grodhaus G: Studies in the epidemiology of myxomatosis in California: observations on two outbreaks of myxomatosis in coastal California and the recovery of myxoma virus from a brush rabbit (*Sylvilagus bachmani*). *American Journal of Epidemiology* 1963, 77:195-204.
28. Regnery DC, Miller JH: A myxoma virus epizootic in a brush rabbit population. *Journal of Wildlife Diseases* 1972, 8:327-331.
29. Licón Luna RM: First report of myxomatosis in Mexico. *Journal of Wildlife Diseases* 2000, 36:580-583.
30. Day MF, Fenner F, Woodroofe GM, McIntyre GA: Further studies on the mechanism of mosquito transmission of myxomatosis in the European rabbit. *The Journal of Hygiene (Camb)* 1956, 54:258-283.
31. Marshall ID, Regnery DC: Studies in the epidemiology of myxomatosis in California: the response of brush rabbits (*Sylvilagus bachmani*) to infection with exotic and enzootic strains of myxoma virus, and the relative infectivity of the tumors for mosquitoes. *American Journal of Epidemiology* 1963, 77:213-219.
32. Regnery DC, Marshall ID: Studies in the epidemiology of myxomatosis in California. IV. The susceptibility of six leporid species to Californian myxoma virus and the relative infectivity of their tumors for mosquitoes. *American Journal of Epidemiology* 1971, 94:508-513.
33. Moses A: O virus do mixoma dos coelhos. *Memórias do Instituto Oswaldo Cruz* 1911, 3:46-53.
34. Fenner F, Marshall ID: A comparison of the virulence for European rabbits (*Oryctolagus cuniculus*) of strains of myxoma virus recovered in the field in Australia, Europe and America. *The Journal of Hygiene (Camb)* 1957, 55:149-191.
35. Silvers L, Inglis B, Labudovic A, Janssens PA, van Leeuwen BH, Kerr PJ: Virulence and pathogenesis of the MSW and MSD strains of Californian myxoma virus in European rabbits with genetic resistance to myxomatosis compared to rabbits with no genetic resistance. *Virology* 2006, 348:72-83.
36. Regnery DC: The epidemic potential of Brazilian myxoma virus (Lausanne strain) for three species of North American cottontails. *American Journal of Epidemiology* 1971, 94:514-519.

37. Silvers L, Barnard D, Knowlton F, Inglis B, Labudovic A, Holland MK, Janssens PA, van Leeuwen BH, Kerr PJ: Host-specificity of myxoma virus: Pathogenesis of South American and North American strains of myxoma virus in two North American lagomorph species. *Veterinary Microbiology* 2010, 141:289-300.
38. Kilham L, Woke PA: Laboratory transmission of fibromas (Shope) in cottontail rabbits by means of fleas and mosquitoes. *Proceedings of the Society for Experimental Biology and Medicine* 1953, 83:296-301.
39. Kilham L, Dalmat HT: Host-virus-mosquito relations of Shope fibromas in cottontail rabbits. *American Journal of Hygiene* 1955, 61:45-54.
40. Willer DO, McFadden G, Evans DH: The complete genome sequence of shope (rabbit) fibroma virus. *Virology* 1999, 264:319-343.
41. Fenner F, Woodroofe GM: Protection of laboratory rabbits against myxomatosis by vaccination with fibroma virus. *Australian Journal of Experimental Biology & Medical Science* 1954, 32:653-668.
42. Marlier D: Vaccination strategies against myxomavirus infections: are we really doing the best? *Tijdschr Diergeneeskd* 2010, 135:194-198.
43. Fenner F: Adventures with poxviruses of vertebrates. *FEMS Microbiology Reviews* 2000, 24:123-133.
44. Zuniga MC: A pox on thee! Manipulation of the host immune system by myxoma virus and implications for viral-host co-adaptation. *Virus Research* 2002, 88:17-33.
45. Fenner F: Classification of Myxoma and Fibroma Viruses. *Nature* 1953, 171:562-563.
46. Fenner F: Deliberate introduction of the European rabbit, *Oryctolagus cuniculus*, into Australia. *Revue Scientifique et Technique* 2010, 29:103-111.
47. Spiesschaert B, McFadden G, Hermans K, Nauwynck H, Van de Walle GR: The current status and future directions of myxoma virus, a master in immune evasion. *Veterinary Research* 2011, 42:76.
48. Aragão HDB: Letter and Report on the virus of rabbit myxoma 30.7.1919, to the Commonwealth of Australia In *Book Letter and Report on the virus of rabbit myxoma 30.7.1919, to the Commonwealth of Australia*. Institute of Science and Industry, Copy in CSIRO Division of Animal Health, File A31; 1919.
49. Kerr PJ: Myxomatosis in Australia and Europe: a model for emerging infectious diseases. *Antiviral Research* 2012, 93:387-415.
50. Martin CJ: Observations on myxomatosis cuniculi (Sanarelli) made with a view to the use of the virus in the control of rabbit plagues. *CSIR Australia Research Bulletin* 1936, 96:1-28.

51. Ratcliffe FN, Myers K, Fennessy BV, Calaby JH: Myxomatosis in Australia; a step towards the biological control of the rabbit. *Nature* 1952, 170:7-11.
52. Fenner F, Woodroofe GM: The pathogenesis of infectious myxomatosis: the mechanism of infection and the immunological response in the European rabbit (*Oryctolagus cuniculus*). *British Journal of Experimental Pathology* 1953, 34:400-410.
53. Fenner F, Fantini B: *Biological Control of Vertebrate Pests: The History of Myxomatosis, An Experiment in Evolution.* New York: CAB International Publ.; 1999.
54. Fenner F, Ross J: Myxomatosis. In *The European rabbit: the history and biology of a successful colonizer.* Edited by Thompson HV, King CM. Oxford: Oxford University Press; 1994: 205-240.
55. Rogers PM, Arthur CP, Soriguer RC: The rabbit in continental Europe. In *The European rabbit: the history and biology of a successful colonizer.* Edited by Thompson HV, King CM. Oxford: Oxford University Press; 1994: 22-63.
56. Delibes M, Rodrigues A, Ferreras P: Action Plan for the Conservation of the Iberian Lynx (*Lynx pardinus*) in Europe. In *Convention on the conservation of European wildlife and natural habitats.* Oslo, Norway; 2000: 1-40.
57. Alves PC, Ferreira C Determinação da abundância relativa das populações de coelho-bravo (*Oryctolagus cuniculus algirus*) em Portugal Continental. In *Protocolo de Colaboração no âmbito do projecto "Revisão do Livro Vermelho dos Vertebrados de Portugal".* Portugal: ICETA, Universidade do Porto; 2004.
58. Fenner F, Day MF, Woodroofe GM: Epidemiological consequences of the mechanical transmission of myxomatosis by mosquitoes. *The Journal of Hygiene (Camb)* 1956, 54:284-303.
59. Myers K, Marshall ID, Fenner F: Studies in the epidemiology of infectious myxomatosis of rabbits: III. Observations on two succeeding epizootics in australian wild rabbits on the riverine plain of south-eastern Australia 1951–1953. *The Journal of Hygiene (Camb)* 1954, 52:337-360.
60. Ross J, Sanders MF: Changes in the virulence of myxoma virus strains in Britain. *Epidemiology & Infection* 1987, 98:113-117.
61. Parer I, Sobey W, Conolly D, Morton R: Virulence of Strains of Myxoma Virus and the Resistance of Wild Rabbits, *Oryctolagus cuniculus* (L), From Different Locations in Australasia. *Australian Journal of Zoology* 1994, 42:347-362.
62. Parer I: Relationship Between Survival Rate and Survival-Time of Rabbits, *Oryctolagus cuniculus* (L), Challenged With Myxoma Virus. *Australian Journal of Zoology* 1995, 43:303-311.

63. Marshall ID, Fenner F: Studies in the epidemiology of infectious myxomatosis of rabbits: V. Changes in the innate resistance of Australian Wild rabbits exposed to myxomatosis. *The Journal of Hygiene (Camb)* 1958, 56:288-302.
64. Marshall ID, Douglas GW: Studies in the epidemiology of infectious myxomatosis of rabbits: VIII. Further observations on changes in the innate resistance of Australian wild rabbits exposed to myxomatosis. *The Journal of Hygiene (Camb)* 1961, 59:117-122.
65. Fenner F: The Florey lecture, 1983. Biological control, as exemplified by smallpox eradication and myxomatosis. *Proceedings of the Royal Society B: Biological Sciences* 1983, 218:259-285.
66. Best SM, Kerr PJ: Coevolution of host and virus: the pathogenesis of virulent and attenuated strains of myxoma virus in resistant and susceptible European rabbits. *Virology* 2000, 267:36-48.
67. Kerr PJ, Perkins HD, Inglis B, Stagg R, McLaughlin E, Collins SV, Van Leeuwen BH: Expression of rabbit IL-4 by recombinant myxoma viruses enhances virulence and overcomes genetic resistance to myxomatosis. *Virology* 2004, 324:117-128.
68. Fenner F, Marshall ID: Occurrence of Attenuated Strains of Myxoma Virus in Europe. *Nature* 1955, 176:782-783.
69. Ross J, Sanders MF: Innate resistance to myxomatosis in wild rabbits in England. *The Journal of Hygiene (Camb)* 1977, 79:411-416.
70. Ross J, Sanders MF: The development of genetic resistance to myxomatosis in wild rabbits in Britain. *The Journal of Hygiene (Camb)* 1984, 92:255-261.
71. Jeklova E, Leva L, Matiasovic J, Kovarcik K, Kudlackova H, Nevorankova Z, Psikal I, Faldyna M: Characterisation of immunosuppression in rabbits after infection with myxoma virus. *Veterinary Microbiology* 2008, 129:117-130.
72. Hobbs JR: Studies of the nature of infectious myxoma of rabbits. *American Journal of Hygiene* 1928, 8:800-839.
73. Hurst EW: Myxoma and the Shope fibroma. I. The histology of myxoma. *British Journal of Experimental Pathology* 1937, 18:1-15.
74. Best SM, Collins SV, Kerr PJ: Coevolution of host and virus: cellular localization of virus in myxoma virus infection of resistant and susceptible European rabbits. *Virology* 2000, 277:76-91.
75. Cameron CM, Barrett JW, Liu L, Lucas AR, McFadden G: Myxoma virus M141R expresses a viral CD200 (vOX-2) that is responsible for down-regulation of macrophage and T-cell activation in vivo. *Journal of Virology* 2005, 79:6052-6067.

76. Cameron CM, Barrett JW, Mann M, Lucas A, McFadden G: Myxoma virus M128L is expressed as a cell surface CD47-like virulence factor that contributes to the downregulation of macrophage activation in vivo. *Virology* 2005, 337:55-67.
77. Seet BT, Johnston JB, Brunetti CR, Barrett JW, Everett H, Cameron C, Sypula J, Nazarian SH, Lucas A, McFadden G: Poxviruses and immune evasion. *Annual Review of Immunology* 2003, 21:377-423.
78. Liu J, Wennier S, McFadden G: The immunoregulatory properties of oncolytic myxoma virus and their implications in therapeutics. *Microbes and Infection* 2010, 12:1144-1152.
79. Nash P, Barrett J, Cao JX, Hota-Mitchell S, Lalani AS, Everett H, Xu XM, Robichaud J, Hnatiuk S, Ainslie C, Seet BT, McFadden G: Immunomodulation by viruses: the myxoma virus story. *Immunological Reviews* 1999, 168:103-120.
80. Upton C, Macen JL, Schreiber M, McFadden G: Myxoma virus expresses a secreted protein with homology to the tumor necrosis factor receptor gene family that contributes to viral virulence. *Virology* 1991, 184:370-382.
81. Upton C, Mossman K, McFadden G: Encoding of a homolog of the IFN-gamma receptor by myxoma virus. *Science* 1992, 258:1369-1372.
82. Mossman K, Upton C, McFadden G: The myxoma virus-soluble interferon-gamma receptor homolog, M-T7, inhibits interferon-gamma in a species-specific manner. *The Journal of Biological Chemistry* 1995, 270:3031-3038.
83. Lalani AS, Ness TL, Singh R, Harrison JK, Seet BT, Kelvin DJ, McFadden G, Moyer RW: Functional comparisons among members of the poxvirus T1/35kDa family of soluble CC-chemokine inhibitor glycoproteins. *Virology* 1998, 250:173-184.
84. Graham KA, Lalani AS, Macen JL, Ness TL, Barry M, Liu LY, Lucas A, Clark-Lewis I, Moyer RW, McFadden G: The T1/35kDa family of poxvirus-secreted proteins bind chemokines and modulate leukocyte influx into virus-infected tissues. *Virology* 1997, 229:12-24.
85. Lalani AS, Masters J, Graham K, Liu L, Lucas A, McFadden G: Role of the myxoma virus soluble CC-chemokine inhibitor glycoprotein, M-T1, during myxoma virus pathogenesis. *Virology* 1999, 256:233-245.
86. Macen JL, Graham KA, Lee SF, Schreiber M, Boshkov LK, McFadden G: Expression of the myxoma virus tumor necrosis factor receptor homologue and M11L genes is required to prevent virus-induced apoptosis in infected rabbit T lymphocytes. *Virology* 1996, 218:232-237.

87. Schreiber M, Sedger L, McFadden G: Distinct domains of M-T2, the myxoma virus tumor necrosis factor (TNF) receptor homolog, mediate extracellular TNF binding and intracellular apoptosis inhibition. *Journal of Virology* 1997, 71:2171-2181.
88. Sedger LM, Osvath SR, Xu XM, Li G, Chan FK, Barrett JW, McFadden G: Poxvirus tumor necrosis factor receptor (TNFR)-like T2 proteins contain a conserved preligand assembly domain that inhibits cellular TNFR1-induced cell death. *Journal of Virology* 2006, 80:9300-9309.
89. Barry M, Hnatiuk S, Mossman K, Lee SF, Boshkov L, McFadden G: The myxoma virus M-T4 gene encodes a novel RDEL-containing protein that is retained within the endoplasmic reticulum and is important for the productive infection of lymphocytes. *Virology* 1997, 239:360-377.
90. Hnatiuk S, Barry M, Zeng W, Liu L, Lucas A, Percy D, McFadden G: Role of the C-terminal RDEL motif of the myxoma virus M-T4 protein in terms of apoptosis regulation and viral pathogenesis. *Virology* 1999, 263:290-306.
91. Mossman K, Lee SF, Barry M, Boshkov L, McFadden G: Disruption of M-T5, a novel myxoma virus gene member of poxvirus host range superfamily, results in dramatic attenuation of myxomatosis in infected European rabbits. *Journal of Virology* 1996, 70:4394-4410.
92. Mossman K, Nation P, Macen J, Garbutt M, Lucas A, McFadden G: Myxoma virus M-T7, a secreted homolog of the interferon-gamma receptor, is a critical virulence factor for the development of myxomatosis in European rabbits. *Virology* 1996, 215:17-30.
93. Lalani AS, Graham K, Mossman K, Rajarathnam K, Clark-Lewis I, Kelvin D, McFadden G: The purified myxoma virus gamma interferon receptor homolog M-T7 interacts with the heparin-binding domains of chemokines. *Journal of Virology* 1997, 71:4356-4363.
94. Macen JL, Upton C, Nation N, McFadden G: SERP1, a serine proteinase inhibitor encoded by myxoma virus, is a secreted glycoprotein that interferes with inflammation. *Virology* 1993, 195:348-363.
95. Opgenorth A, Graham K, Nation N, Strayer D, McFadden G: Deletion analysis of two tandemly arranged virulence genes in myxoma virus, M11L and myxoma growth factor. *Journal of Virology* 1992, 66:4720-4731.
96. Graham KA, Opgenorth A, Upton C, McFadden G: Myxoma virus M11L ORF encodes a protein for which cell surface localization is critical in manifestation of viral virulence. *Virology* 1992, 191:112-124.

97. Johnston JB, Barrett JW, Nazarian SH, Goodwin M, Ricciuto D, Wang G, McFadden G: A poxvirus-encoded pyrin domain protein interacts with ASC-1 to inhibit host inflammatory and apoptotic responses to infection. *Immunity* 2005, 23:587-598.
98. Dorfleutner A, Talbott SJ, Bryan NB, Funya KN, Rellick SL, Reed JC, Shi X, Rojanasakul Y, Flynn DC, Stehlik C: A Shope Fibroma virus PYRIN-only protein modulates the host immune response. *Virus Genes* 2007, 35:685-694.
99. Rahman MM, Mohamed MR, Kim M, Smallwood S, McFadden G: Co-regulation of NF-kappaB and inflammasome-mediated inflammatory responses by myxoma virus pyrin domain-containing protein M013. *PLoS Pathogens* 2009, 5:e1000635.
100. Rahman MM, McFadden G: Myxoma virus lacking the pyrin-like protein M013 is sensed in human myeloid cells by both NLRP3 and multiple Toll-like receptors, which independently activate the inflammasome and NF-kappaB innate response pathways. *Journal of Virology* 2011, 85:12505-12517.
101. Myskiw C, Arsenio J, Hammett C, van Bruggen R, Deschambault Y, Beausoleil N, Babiuk S, Cao J: Comparative analysis of poxvirus orthologues of the vaccinia virus E3 protein: modulation of protein kinase R activity, cytokine responses, and virus pathogenicity. *Journal of Virology* 2011, 85:12280-12291.
102. Rahman MM, Liu J, Chan WM, Rothenburg S, McFadden G: Myxoma virus protein M029 is a dual function immunomodulator that inhibits PKR and also conscripts RHA/DHX9 to promote expanded host tropism and viral replication. *PLoS Pathogens* 2013, 9:e1003465.
103. Liu J, Wennier S, Zhang L, McFadden G: M062 is a host range factor essential for myxoma virus pathogenesis and functions as an antagonist of host SAMD9 in human cells. *Journal of Virology* 2011, 85:3270-3282.
104. Barrett JW, Shun Chang C, Wang G, Werden SJ, Shao Z, Barrett C, Gao X, Belsito TA, Villenevue D, McFadden G: Myxoma virus M063R is a host range gene essential for virus replication in rabbit cells. *Virology* 2007, 361:123-132.
105. Liu J, Wennier S, Moussatche N, Reinhard M, Condit R, McFadden G: Myxoma virus M064 is a novel member of the poxvirus C7L superfamily of host range factors that controls the kinetics of myxomatosis in European rabbits. *Journal of Virology* 2012, 86:5371-5375.
106. Barrett JW, Werden SJ, Wang F, McKillop WM, Jimenez J, Villeneuve D, McFadden G, Dekaban GA: Myxoma virus M130R is a novel virulence factor required for lethal myxomatosis in rabbits. *Virus Research* 2009, 144:258-265.

107. Cao JX, Teoh ML, Moon M, McFadden G, Evans DH: Leporipoxvirus Cu-Zn superoxide dismutase homologs inhibit cellular superoxide dismutase, but are not essential for virus replication or virulence. *Virology* 2002, 296:125-135.
108. Teoh ML, Walasek PJ, Evans DH: Leporipoxvirus Cu,Zn-superoxide dismutase (SOD) homologs are catalytically inert decoy proteins that bind copper chaperone for SOD. *The Journal of Biological Chemistry* 2003, 278:33175-33184.
109. Teoh ML, Turner PV, Evans DH: Tumorigenic poxviruses up-regulate intracellular superoxide to inhibit apoptosis and promote cell proliferation. *Journal of Virology* 2005, 79:5799-5811.
110. Barrett JW, Sypula J, Wang F, Alston LR, Shao Z, Gao X, Irvine TS, McFadden G: M135R is a novel cell surface virulence factor of myxoma virus. *Journal of Virology* 2007, 81:106-114.
111. Jackson RJ, Hall DF, Kerr PJ: Myxoma virus encodes an alpha2,3-sialyltransferase that enhances virulence. *Journal of Virology* 1999, 73:2376-2384.
112. Blanié S, Mortier J, Delverdier M, Bertagnoli S, Camus-Bouclainville C: M148R and M149R are two virulence factors for myxoma virus pathogenesis in the European rabbit. *Veterinary Research* 2009, 40:11.
113. Zhang L, Villa NY, McFadden G: Interplay between poxviruses and the cellular ubiquitin/ubiquitin-like pathways. *FEBS Letters* 2009, 583:607-614.
114. Camus-Bouclainville C, Fiette L, Bouchiha S, Pignolet B, Counor D, Filipe C, Gelfi J, Messud-Petit F: A virulence factor of myxoma virus colocalizes with NF-kappaB in the nucleus and interferes with inflammation. *Journal of Virology* 2004, 78:2510-2516.
115. Blanié S, Gelfi J, Bertagnoli S, Camus-Bouclainville C: MNF, an ankyrin repeat protein of myxoma virus, is part of a native cellular SCF complex during viral infection. *Virology Journal* 2010, 7:56.
116. Messud-Petit F, Gelfi J, Delverdier M, Amardeilh MF, Py R, Sutter G, Bertagnoli S: Serp2, an inhibitor of the interleukin-1beta-converting enzyme, is critical in the pathobiology of myxoma virus. *Journal of Virology* 1998, 72:7830-7839.
117. Guerin JL, Gelfi J, Camus C, Delverdier M, Whisstock JC, Amardeihl MF, Py R, Bertagnoli S, Messud-Petit F: Characterization and functional analysis of Serp3: a novel myxoma virus-encoded serpin involved in virulence. *Journal of General Virology* 2001, 82:1407-1417.

118. Guerin JL, Gelfi J, Boullier S, Delverdier M, Bellanger FA, Bertagnoli S, Drexler I, Sutter G, Messud-Petit F: Myxoma virus leukemia-associated protein is responsible for major histocompatibility complex class I and Fas-CD95 down-regulation and defines scrapins, a new group of surface cellular receptor abductor proteins. *Journal of Virology* 2002, 76:2912-2923.
119. Mansouri M, Bartee E, Gouveia K, Hovey Nerenberg BT, Barrett J, Thomas L, Thomas G, McFadden G, Fruh K: The PHD/LAP-domain protein M153R of myxomavirus is a ubiquitin ligase that induces the rapid internalization and lysosomal destruction of CD4. *Journal of Virology* 2003, 77:1427-1440.
120. Collin N, Guerin JL, Drexler I, Blanie S, Gelfi J, Boullier S, Foucras G, Sutter G, Messud-Petit F: The poxviral scrapin MV-LAP requires a myxoma viral infection context to efficiently downregulate MHC-I molecules. *Virology* 2005, 343:171-178.
121. Ramelot TA, Cort JR, Yee AA, Liu F, Goshe MB, Edwards AM, Smith RD, Arrowsmith CH, Dever TE, Kennedy MA: Myxoma virus immunomodulatory protein M156R is a structural mimic of eukaryotic translation initiation factor eIF2alpha. *Journal of Molecular Biology* 2002, 322:943-954.
122. Schreiber M, McFadden G: The myxoma virus TNF-receptor homologue (T2) inhibits tumor necrosis factor-alpha in a species-specific fashion. *Virology* 1994, 204:692-705.
123. Schreiber M, Rajarathnam K, McFadden G: Myxoma virus T2 protein, a tumor necrosis factor (TNF) receptor homolog, is secreted as a monomer and dimer that each bind rabbit TNFalpha, but the dimer is a more potent TNF inhibitor. *The Journal of Biological Chemistry* 1996, 271:13333-13341.
124. Smith SA, Kotwal GJ: Virokines: novel immunomodulatory agents. *Expert Opinion on Biological Therapy* 2001, 1:343-357.
125. Upton C, Macen JL, Wishart DS, McFadden G: Myxoma virus and malignant rabbit fibroma virus encode a serpin-like protein important for virus virulence. *Virology* 1990, 179:618-631.
126. Everett H, McFadden G: Viral proteins and the mitochondrial apoptotic checkpoint. *Cytokine & Growth Factor Reviews* 2001, 12:181-188.
127. Everett H, McFadden G: Viruses and apoptosis: meddling with mitochondria. *Virology* 2001, 288:1-7.
128. Xu X, Nash P, McFadden G: Myxoma virus expresses a TNF receptor homolog with two distinct functions. *Virus Genes* 2000, 21:97-109.

129. Everett H, Barry M, Lee SF, Sun X, Graham K, Stone J, Bleackley RC, McFadden G: M11L: a novel mitochondria-localized protein of myxoma virus that blocks apoptosis of infected leukocytes. *The Journal of Experimental Medicine* 2000, 191:1487-1498.
130. Everett H, Barry M, Sun X, Lee SF, Frantz C, Berthiaume LG, McFadden G, Bleackley RC: The myxoma poxvirus protein, M11L, prevents apoptosis by direct interaction with the mitochondrial permeability transition pore. *The Journal of Experimental Medicine* 2002, 196:1127-1139.
131. Wang G, Barrett JW, Nazarian SH, Everett H, Gao X, Bleackley C, Colwill K, Moran MF, McFadden G: Myxoma virus M11L prevents apoptosis through constitutive interaction with Bak. *Journal of Virology* 2004, 78:7097-7111.
132. Su J, Wang G, Barrett JW, Irvine TS, Gao X, McFadden G: Myxoma virus M11L blocks apoptosis through inhibition of conformational activation of Bax at the mitochondria. *Journal of Virology* 2006, 80:1140-1151.
133. Douglas AE, Corbett KD, Berger JM, McFadden G, Handel TM: Structure of M11L: A myxoma virus structural homolog of the apoptosis inhibitor, Bcl-2. *Protein Science* 2007, 16:695-703.
134. Bartee E, McCormack A, Fruh K: Quantitative membrane proteomics reveals new cellular targets of viral immune modulators. *PLoS Pathogens* 2006, 2:e107.
135. Medzhitov R, Janeway C, Jr.: Innate immunity. *The New England Journal of Medicine* 2000, 343:338-344.
136. Akira S, Uematsu S, Takeuchi O: Pathogen recognition and innate immunity. *Cell* 2006, 124:783-801.
137. Kawai T, Akira S: The roles of TLRs, RLRs and NLRs in pathogen recognition. *International Immunology* 2009, 21:317-337.
138. Wilkins C, Gale Jr M: Recognition of viruses by cytoplasmic sensors. *Current Opinion in Immunology* 2010, 22:41-47.
139. Garcia MA, Meurs EF, Esteban M: The dsRNA protein kinase PKR: virus and cell control. *Biochimie* 2007, 89:799-811.
140. Werden SJ, Rahman MM, McFadden G: Poxvirus host range genes. *Advances in Virus Research* 2008, 71:135-171.
141. Wang G, Barrett JW, Stanford M, Werden SJ, Johnston JB, Gao X, Sun M, Cheng JQ, McFadden G: Infection of human cancer cells with myxoma virus requires Akt activation via interaction with a viral ankyrin-repeat host range factor. *Proceedings of the National Academy of Sciences of the United States of America* 2006, 103:4640-4645.

142. Kerr PJ, Ghedin E, DePasse JV, Fitch A, Cattadori IM, Hudson PJ, Tscharke DC, Read AF, Holmes EC: Evolutionary history and attenuation of myxoma virus on two continents. *PLoS Pathogens* 2012, 8:e1002950.
143. Babkin IV, Shchelkunov SN: Time scale of Poxvirus evolution. *Molecular Biology* 2006, 40:16-19.
144. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A: Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Molecular Biology and Evolution* 2010, 27:2038-2051.
145. Babkin IV, Babkina IN: Molecular dating in the evolution of vertebrate poxviruses. *Intervirology* 2011, 54:253-260.
146. Werden SJ, McFadden G: The role of cell signaling in poxvirus tropism: the case of the M-T5 host range protein of myxoma virus. *Biochimica et Biophysica Acta* 2008, 1784:228-237.
147. Russell RJ, Robbins SJ: Cloning and molecular characterization of the myxoma virus genome. *Virology* 1989, 170:147-159.
148. Saint KM, French N, Kerr P: Genetic variation in Australian isolates of myxoma virus: an evolutionary and epidemiological study. *Archives of Virology* 2001, 146:1105-1123.
149. Bárcena J, Pagès-Manté A, March R, Morales M, Ramírez MA, Sánchez-Vizcaíno JM, Torres JM: Isolation of an attenuated myxoma virus field strain that can confer protection against myxomatosis on contacts of vaccinates. *Archives of Virology* 2000, 145:759-771.
150. Morales M, Ramírez MA, Cano MJ, Párraga M, Castilla J, Pérez-Ordoyo LI, Torres JM, Bárcena J: Genome comparison of a nonpathogenic myxoma virus field strain with its ancestor, the virulent Lausanne strain. *Journal of Virology* 2009, 83:2397-2403.
151. Alda F, Gaitero T, Suárez M, Doadrio I: Molecular characterisation and recent evolution of myxoma virus in Spain. *Archives of Virology* 2009, 154:1659-1670.
152. Muller A, Silva E, Abrantes J, Esteves PJ, Ferreira PG, Carvalheira JC, Nowotny N, Thompson G: Partial sequencing of recent Portuguese myxoma virus field isolates exhibits a high degree of genetic stability. *Veterinary Microbiology* 2010, 140:161-166.
153. Duggal NK, Emerman M: Evolutionary conflicts between viruses and restriction factors shape immunity. *Nature Reviews Immunology* 2012, 12:687-695.
154. Johnson WE: Rapid adversarial co-evolution of viruses and cellular restriction factors. *Current Topics in Microbiology and Immunology* 2013, 371:123-151.

155. Hatziioannou T, Bieniasz PD: Antiretroviral restriction factors. *Current Opinion in Virology* 2011, 1:526-532.
156. Zheng YH, Jeang KT, Tokunaga K: Host restriction factors in retroviral infection: promises in virus-host interaction. *Retrovirology* 2012, 9:112.
157. Takeda A, Matano T: Inhibition of infectious murine leukemia virus production by Fv-4 env gene products exerting dominant negative effect on viral envelope glycoprotein. *Microbes and Infection* 2007, 9:1590-1596.
158. Spencer TE, Mura M, Gray CA, Griebel PJ, Palmarini M: Receptor usage and fetal expression of ovine endogenous betaretroviruses: implications for coevolution of endogenous and exogenous retroviruses. *Journal of Virology* 2003, 77:749-753.
159. Mura M, Murcia P, Caporale M, Spencer TE, Nagashima K, Rein A, Palmarini M: Late viral interference induced by transdominant Gag of an endogenous retrovirus. *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101:11117-11122.
160. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J: The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* 2004, 427:848-853.
161. Yap MW, Nisole S, Lynch C, Stoye JP: Trim5alpha protein restricts both HIV-1 and murine leukemia virus. *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101:10786-10791.
162. Perez-Caballero D, Hatziioannou T, Yang A, Cowan S, Bieniasz PD: Human tripartite motif 5alpha domains responsible for retrovirus restriction activity and specificity. *Journal of Virology* 2005, 79:8969-8978.
163. Song B, Javanbakht H, Perron M, Park DH, Stremlau M, Sodroski J: Retrovirus restriction by TRIM5alpha variants from Old World and New World primates. *Journal of Virology* 2005, 79:3930-3937.
164. Trono D: HIV accessory proteins: leading roles for the supporting cast. *Cell* 1995, 82:189-192.
165. Seelamgari A, Maddukuri A, Berro R, de la Fuente C, Kehn K, Deng L, Dadgar S, Bottazzi ME, Ghedin E, Pumfery A, Kashanchi F: Role of viral regulatory and accessory proteins in HIV-1 replication. *Frontiers in Bioscience* 2004, 9:2388-2413.
166. Perdiguero B, Esteban M: The interferon system and vaccinia virus evasion mechanisms. *Journal of Interferon & Cytokine Research* 2009, 29:581-598.

167. Tokarev A, Skasko M, Fitzpatrick K, Guatelli J: Antiviral activity of the interferon-induced cellular protein BST-2/tetherin. *AIDS Research and Human Retroviruses* 2009, 25:1197-1210.
168. Malim MH, Emerman M: HIV-1 accessory proteins-ensuring viral survival in a hostile environment. *Cell Host & Microbe* 2008, 3:388-398.
169. Kirchhoff F: Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. *Cell Host & Microbe* 2010, 8:55-67.
170. Kuhl BD, Cheng V, Wainberg MA, Liang C: Tetherin and its viral antagonists. *Journal of Neuroimmune Pharmacology* 2011, 6:188-201.
171. Malim MH, Bieniasz PD: HIV restriction factors and mechanisms of evasion. *Cold Spring Harbor Perspectives in Medicine* 2012, 2.
172. Nomaguchi M, Doi N, Matsumoto Y, Sakai Y, Fujiwara S, Adachi A: Species tropism of HIV-1 modulated by viral accessory proteins. *Frontiers in Microbiology* 2012, 3:267.
173. Kuroishi A, Bozek K, Shioda T, Nakayama E: A single amino acid substitution of the human immunodeficiency virus type 1 capsid protein affects viral sensitivity to TRIM5alpha. *Retrovirology* 2010, 7:58.
174. Pacheco B, Finzi A, Stremlau M, Sodroski J: Adaptation of HIV-1 to cells expressing rhesus monkey TRIM5α. *Virology* 2010, 408:204-212.
175. Ohkura S, Goldstone DC, Yap MW, Holden-Dye K, Taylor IA, Stoye JP: Novel escape mutants suggest an extensive TRIM5α binding site spanning the entire outer surface of the murine leukemia virus capsid protein. *PLoS Pathogens* 2011, 7:e1002011.
176. Kimura M: Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 1977, 267:275-276.
177. Yang Z, Bielawski JP: Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution* 2000, 15:496-503.
178. Arnaud F, Caporale M, Varela M, Biek R, Chessa B, Alberti A, Golder M, Mura M, Zhang Y-p, Yu L, Pereira F, DeMartini JC, Leymaster K, Spencer TE, Palmarini M: A paradigm for virus–host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses. *PLoS Pathogens* 2007, 3:e170.
179. Pincus T, Rowe WP, Lilly F: A major genetic locus affecting resistance to infection with murine leukemia viruses. II. Apparent identity to a major locus described for resistance to friend murine leukemia virus. *The Journal of Experimental Medicine* 1971, 133:1234-1241.

180. Yan Y, Buckler-White A, Wollenberg K, Kozak CA: Origin, antiviral function and evidence for positive selection of the gammaretrovirus restriction gene Fv1 in the genus Mus. *Proceedings of the National Academy of Sciences of the United States of America* 2009, 106:3259-3263.
181. Sawyer SL, Wu LI, Emerman M, Malik HS: Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102:2832- 2837.
182. Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH: Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Current Biology* 2004, 14:1392-1396.
183. Kinomoto M, Kanno T, Shimura M, Ishizaka Y, Kojima A, Kurata T, Sata T, Tokunaga K: All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Research* 2007, 35:2955-2964.
184. Suspene R, Guetard D, Henry M, Sommer P, Wain-Hobson S, Vartanian JP: Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102:8321-8326.
185. Duggal NK, Malik HS, Emerman M: The breadth of antiviral activity of Apobec3DE in chimpanzees has been driven by positive selection. *Journal of Virology* 2011, 85:11361-11371.
186. Sawyer SL, Emerman M, Malik HS: Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biology* 2004, 2:e275.
187. Zhang J, Webb DM: Rapid evolution of primate antiviral enzyme APOBEC3G. *Human Molecular Genetics* 2004, 13:1785-1791.
188. OhAinle M, Kerns JA, Malik HS, Emerman M: Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. *Journal of Virology* 2006, 80:3853-3862.
189. Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Segeral E, Yatim A, Emiliani S, Schwartz O, Benkirane M: SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* 2011, 474:654-657.
190. Gramberg T, Kahle T, Bloch N, Wittmann S, Mullers E, Daddacha W, Hofmann H, Kim B, Lindemann D, Landau NR: Restriction of diverse retroviruses by SAMHD1. *Retrovirology* 2013, 10:26.

191. Laguette N, Rahm N, Sobhian B, Chable-Bessia C, Munch J, Snoeck J, Sauter D, Switzer WM, Heneine W, Kirchhoff F, Delsuc F, Telenti A, Benkirane M: Evolutionary and functional analyses of the interaction between the myeloid restriction factor SAMHD1 and the lentiviral Vpx protein. *Cell Host & Microbe* 2012, 11:205-217.
192. Lim ES, Fregoso OI, McCoy CO, Matsen FA, Malik HS, Emerman M: The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. *Cell Host & Microbe* 2012, 11:194-204.
193. Gao G, Guo X, Goff SP: Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science* 2002, 297:1703-1706.
194. Zhu Y, Chen G, Lv F, Wang X, Ji X, Xu Y, Sun J, Wu L, Zheng YT, Gao G: Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. *Proceedings of the National Academy of Sciences of the United States of America* 2011, 108:15834-15839.
195. Muller S, Moller P, Bick MJ, Wurr S, Becker S, Gunther S, Kummerer BM: Inhibition of filovirus replication by the zinc finger antiviral protein. *Journal of Virology* 2007, 81:2391-2400.
196. Bick MJ, Carroll JW, Gao G, Goff SP, Rice CM, MacDonald MR: Expression of the zinc-finger antiviral protein inhibits alphavirus replication. *Journal of Virology* 2003, 77:11555-11562.
197. Kerns JA, Emerman M, Malik HS: Positive selection and increased antiviral activity associated with the PARP - containing isoform of human zinc-finger antiviral protein. *PLoS Genetics* 2008, 4:e21.
198. Neil SJ, Sandrin V, Sundquist WI, Bieniasz PD: An interferon-alpha-induced tethering mechanism inhibits HIV-1 and Ebola virus particle release but is counteracted by the HIV-1 Vpu protein. *Cell Host & Microbe* 2007, 2:193-203.
199. Neil SJD, Zang T, Bieniasz PD: Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* 2008, 451:425-430.
200. Pan X-B, Han J-C, Cong X, Wei L: BST2 / Tetherin inhibits dengue virus release from human hepatoma cells. *PLoS One* 2012, 7:e51033.
201. Mansouri M, Viswanathan K, Douglas JL, Hines J, Gustin J, Moses AV, Fruh K: Molecular mechanism of BST2/tetherin downregulation by K5/MIR2 of Kaposi's sarcoma-associated herpesvirus. *Journal of Virology* 2009, 83:9672-9681.
202. Weidner JM, Jiang D, Pan X-B, Chang J, Block TM, Guo J-T: Interferon-induced cell membrane proteins, IFITM3 and tetherin, inhibit vesicular stomatitis virus infection via distinct mechanisms. *Journal of Virology* 2010, 84:12646-12657.

203. Kong WS, Irie T, Yoshida A, Kawabata R, Kadoi T, Sakaguchi T: Inhibition of virus-like particle release of Sendai virus and Nipah virus, but not that of mumps virus, by tetherin/CD317/BST-2. *Hiroshima Journal of Medical Sciences* 2012, 61:59-67.
204. Radoshitzky SR, Dong L, Chi X, Clester JC, Retterer C, Spurgers K, Kuhn JH, Sandwick S, Ruthel G, Kota K, Boltz D, Warren T, Kranzusch PJ, Whelan SP, Bavari S: Infectious Lassa virus, but not filoviruses, is restricted by BST-2/tetherin. *Journal of Virology* 2010, 84:10569-10580.
205. McNatt MW, Zang T, Hatziioannou T, Bartlett M, Fofana IB, Johnson WE, Neil SJ, Bieniasz PD: Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants. *PLoS Pathogens* 2009, 5:e1000300.
206. Lim ES, Malik HS, Emerman M: Ancient adaptive evolution of tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses. *Journal of Virology* 2010, 84:7124-7134.
207. Liu J, Chen K, Wang JH, Zhang C: Molecular evolution of the primate antiviral restriction factor tetherin. *PLoS One* 2010, 5:e11904.
208. Wang X, Hinson ER, Cresswell P: The interferon-inducible protein viperin inhibits influenza virus release by perturbing lipid rafts. *Cell Host & Microbe* 2007, 2:96-105.
209. Szretter KJ, Brien JD, Thackray LB, Virgin HW, Cresswell P, Diamond MS: The interferon-inducible gene viperin restricts West Nile virus pathogenesis. *Journal of Virology* 2011, 85:11557-11566.
210. Chin KC, Cresswell P: Viperin (cig5), an IFN-inducible antiviral protein directly induced by human cytomegalovirus. *Proceedings of the National Academy of Sciences of the United States of America* 2001, 98:15125-15130.
211. Teng TS, Foo SS, Simamarta D, Lum FM, Teo TH, Lulla A, Yeo NK, Koh EG, Chow A, Leo YS, Merits A, Chin KC, Ng LF: Viperin restricts chikungunya virus replication and pathology. *The Journal of Clinical Investigation* 2012, 122:4447-4460.
212. Severa M, Coccia EM, Fitzgerald KA: Toll-like receptor-dependent and -independent viperin gene expression and counter-regulation by PRDI-binding factor-1/BLIMP1. *Journal of Biological Chemistry* 2006, 281:26188-26195.
213. Lim E, Wu L, Malik H, Emerman M: The function and evolution of the restriction factor viperin in primates was not driven by lentiviruses. *Retrovirology* 2012, 9:55.
214. Pavlovic J, Zurcher T, Haller O, Staeheli P: Resistance to influenza virus and vesicular stomatitis virus conferred by expression of human MxA protein. *Journal of Virology* 1990, 64:3370-3375.

215. Schnorr JJ, Schneider-Schaulies S, Simon-Jödicke A, Pavlovic J, Horisberger MA, ter Meulen V: MxA-dependent inhibition of measles virus glycoprotein synthesis in a stably transfected human monocytic cell line. *Journal of Virology* 1993, 67:4760-4768.
216. Gordien E, Rosmorduc O, Peltekian C, Garreau F, Brechot C, Kremsdorf D: Inhibition of hepatitis B virus replication by the interferon-inducible MxA protein. *Journal of Virology* 2001, 75:2684-2691.
217. Frese M, Kochs G, Feldmann H, Hertkorn C, Haller O: Inhibition of bunyaviruses, phleboviruses, and hantaviruses by human MxA protein. *Journal of Virology* 1996, 70:915-923.
218. Landis H, Simon-Jodicke A, Kloti A, Di Paolo C, Schnorr JJ, Schneider-Schaulies S, Hefti HP, Pavlovic J: Human MxA protein confers resistance to Semliki Forest virus and inhibits the amplification of a Semliki Forest virus-based replicon in the absence of viral structural proteins. *Journal of Virology* 1998, 72:1516-1522.
219. Brass AL, Huang IC, Benita Y, John SP, Krishnan MN, Feeley EM, Ryan BJ, Weyer JL, van der Weyden L, Fikrig E, Adams DJ, Xavier RJ, Farzan M, Elledge SJ: The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell* 2009, 139:1243-1254.
220. Huang IC, Bailey CC, Weyer JL, Radoshitzky SR, Becker MM, Chiang JJ, Brass AL, Ahmed AA, Chi X, Dong L, Longobardi LE, Boltz D, Kuhn JH, Elledge SJ, Bavari S, Denison MR, Choe H, Farzan M: Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. *PLoS Pathogens* 2011, 7:e1001258.
221. Davies MV, Chang HW, Jacobs BL, Kaufman RJ: The E3L and K3L vaccinia virus gene products stimulate translation through inhibition of the double-stranded RNA-dependent protein kinase by different mechanisms. *Journal of Virology* 1993, 67:1688-1692.
222. Elde NC, Child SJ, Geballe AP, Malik HS: Protein kinase R reveals an evolutionary model for defeating viral mimicry. *Nature* 2009, 457:485-489.
223. Rothenburg S, Seo EJ, Gibbs JS, Dever TE, Dittmar K: Rapid evolution of protein kinase PKR alters sensitivity to viral inhibitors. *Nature Structural & Molecular Biology* 2009, 16:63-70.
224. Miyata T, Yasunaga T: Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *Journal of Molecular Evolution* 1980, 16:23-36.

225. Li WH, Wu CI, Luo CC: A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular Biology and Evolution* 1985, 2:150-174.
226. Nei M, Gojobori T: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* 1986, 3:418-426.
227. Li WH: Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *Journal of Molecular Evolution* 1993, 36:96-99.
228. Comeron JM: A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. *Journal of Molecular Evolution* 1995, 41:1152-1159.
229. Ina Y: New methods for estimating the numbers of synonymous and nonsynonymous substitutions. *Journal of Molecular Evolution* 1995, 40:190-226.
230. Goldman N, Yang Z: A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution* 1994, 11:725-736.
231. Muse SV, Gaut BS: A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution* 1994, 11:715-724.
232. Van Valen L: A new evolutionary law. *Evolutionary Theory* 1973, 1:1-30.
233. Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, VandePol S: Rapid evolution of RNA genomes. *Science* 1982, 215:1577-1585.
234. Jenkins GM, Rambaut A, Pybus OG, Holmes EC: Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *Journal of Molecular Evolution* 2002, 54:156-165.
235. Hanada K, Suzuki Y, Gojobori T: A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Molecular Biology and Evolution* 2004, 21:1074-1080.

Chapter 2

Innate anti-viral immunity – Genetic aspects of C-C motif chemokines in Leporidae genera

Paper 1

Genetic characterization of CCL3, CCL4 and CCL5 in leporid genera *Oryctolagus*, *Sylvilagus* and *Lepus*

A. Lemos de Matos, D. K. Lanning & P. J. Esteves

1. Summary

The genetic diversity of C-C motif chemokine receptor 5 (CCR5) ligands CCL3, CCL4 and CCL5 in the leporid genera *Oryctolagus*, *Sylvilagus* and *Lepus* was studied. Our results demonstrate that the three CCR5 chemokine ligands are under strong purifying selection as a result of possible functional binding constraints.

2. Short communication

The chemokine system consists of a family of low molecular weight chemotactic cytokines defined by four conserved cysteine residues that activate seven-transmembrane, specific G protein-coupled receptors [1, 2]. Functionally, some chemokines play critical roles in development and homeostasis when produced and secreted constitutively (homeostatic chemokines), while others are crucial in immune and inflammatory responses when produced by cells during infection or in the presence of a pro-inflammatory stimulus (inflammatory chemokines) [1-3].

One of the characteristics of the chemokine system is the so-called binding promiscuity, as a given single receptor frequently binds several chemokines and *vice versa* [4]. Understanding this phenomenon triggered several studies on the evolution of mammalian chemokine genes. A mixed process of birth-and-death and concerted evolution has been suggested, as the chemokine multigene family is apparently resultant from ancestral gene duplication events and, in a still ongoing process, consequent lineage-specific gene gains and losses [5-8]. Furthermore, gene conversion events between paralogous genes in one species, which is a characteristic feature of

concerted evolution, may have contributed to the binding promiscuity of the chemokine system [9-14].

The C-C motif chemokine receptor 5 (CCR5) and its ligands are an exceptional representative of the binding promiscuity typical of the chemokine system. A study on leporid genera *Oryctolagus*, *Sylvilagus* and *Lepus CCR5* sequences revealed that the second extracellular loop of *Oryctolagus CCR5* experienced a gene conversion event with the C-C motif chemokine receptor 2 (*CCR2*), where the CCR5 sequence motif QTLKMT was replaced by the CCR2 motif HTIMRN [10]. The functional significance of this gene conversion event on *Oryctolagus* CCR5 remains unknown, but it became imperative to study CCR5 ligands' genetic aspects. A first study on leporid C-C motif chemokine 8 (CCL8), a CCR5 ligand, revealed an *Oryctolagus-Bunolagus* lineage-specific feature, as *CCL8* gene was pseudogenized while it was intact in the *Lepus-Sylvilagus* lineage [15]. This result on leporid *CCL8* prompted us to investigate the genetic variability and to look for signatures of selection in CCR5 disease-related C-C motif chemokine ligands CCL3, CCL4 and CCL5 in the three well-studied Leporidae genera *Oryctolagus*, *Sylvilagus* and *Lepus*.

Liver and spleen tissues from specimens of European rabbit subspecies *Oryctolagus cuniculus algirus* (two individuals), brush rabbit (*Sylvilagus bachmani*; two individuals), European brown hare (*Lepus europaeus*; three individuals) and Iberian hare (*Lepus granatensis*; three individuals) were used for RNA extraction. It was not necessary to obtain approval from an ethics committee for tissues samples, once these samples were already described and used in previous publications [10, 15-18]. Total RNA preparation and cDNA synthesis protocols are described in de Matos *et al.* (2011) [18]. Amplification by RT-PCR was performed for leporid *CCL3* (F_5'-TCGTCACCTGCTCGGCAC-3'; R_5'-CTTGCAGCTTCCGGCCTC-'3), *CCL4* (F_5'-CAGCCTCTGCCCTGAGAAAG-3'; R_5'-CTGAAGACTTCCAGCCTGGAG-'3) and *CCL5* (F_5'-CCGCTTCTCCACAGCTCTG-3'; R_5'-GTGTGCAAGTTCAGGTTCAAGGC-3') using the Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Finnzymes, Waltham, MA, USA) and according to manufacturer's protocol. Primers were designed based on the available sequences for European rabbit in Ensembl (accession numbers are listed in Figure 1). Amplified PCR products were cloned into the pGEM-T Easy vector (Promega, Madison, WI, USA). At least five independent clones were sequenced per used specimen. Sequencing was performed with an ABI PRISM 3130 Genetic Analyser (Life Technologies, Applied Biosystems, Carlsbad, CA, USA), following the ABI PRISM BigDye Terminator Cycle sequencing protocol. Each described allele for the different species was defined by the existence of at least three clones. Nucleotide sequence data obtained in this study for *CCL3*, *CCL4* and *CCL5* have been submitted to

GenBank and have been assigned the following accession numbers: KF527417 to KF527437.

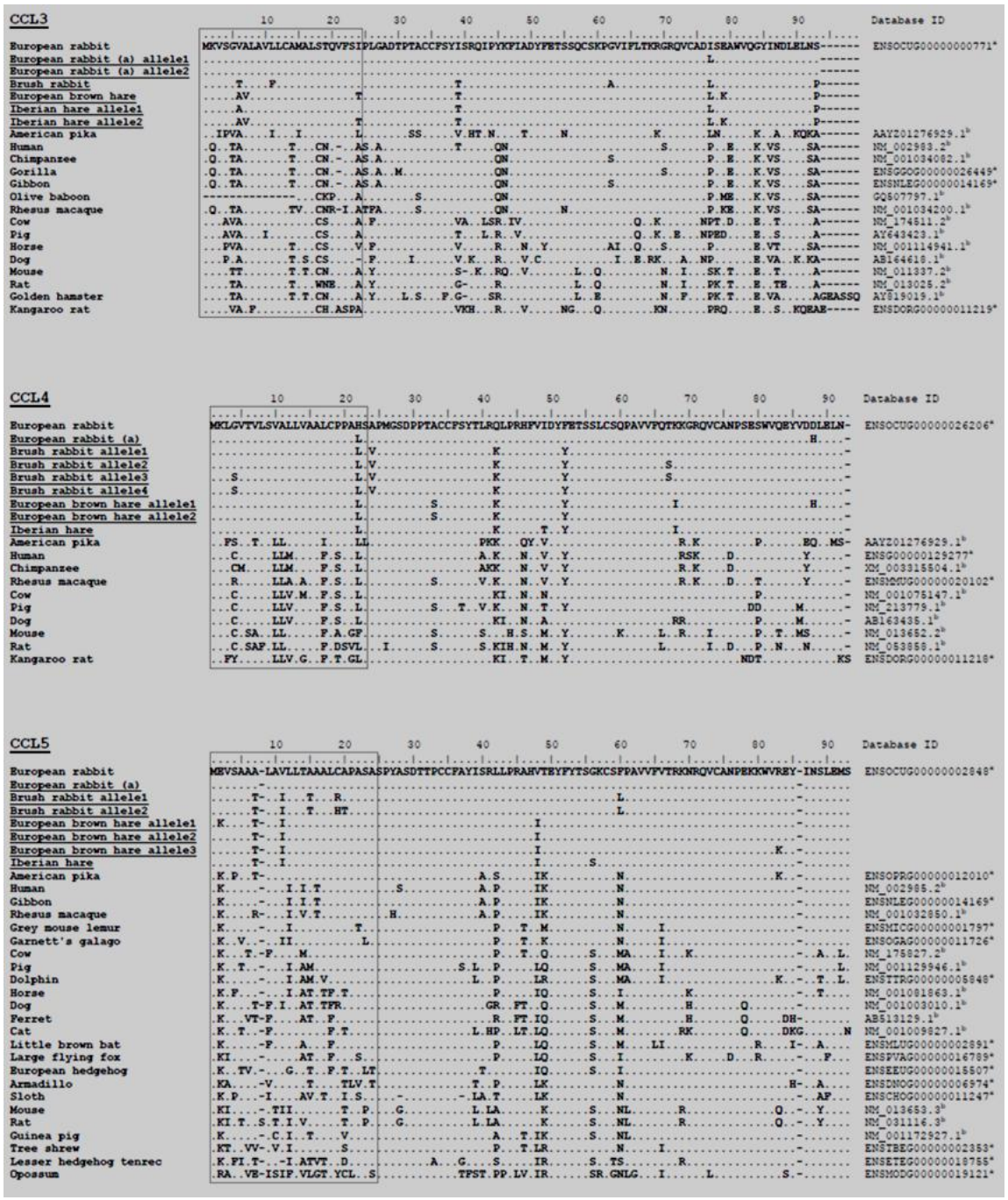


Figure 1 - Deduced protein sequences of mammalian CCR5 chemokine ligands CCL3, CCL4 and CCL5.

The inferred protein sequences for leporid samples are underlined, while the remaining mammalian sequences were collected from (a) Ensembl and (b) NCBI databases. European rabbit (a) corresponds to the subspecies *Oryctolagus cuniculus algirus*. The box is delimiting the signal peptide of each protein. Dashes (-) represent alignment gaps; dots (.) correspond to identical sites when compared to the master sequence (European rabbit, *Oryctolagus cuniculus cuniculus*).

The inferred protein sequences were aligned with the remaining mammalian sequences collected for CCL3, CCL4 and CCL5 (Figure 1). Several allelic forms were identified for the three CCR5 chemokine ligands and in different leporid species (Figure 1). The number of amino acid differences per site between leporid sequences for the three proteins was estimated (Table 1). Apparently, the evolutionary divergence between *Oryctolagus*-*Sylvilagus*-*Lepus* lineages for each protein was similar. Pairwise amino acid distances between the leporid genera and American pika, a representative of the other Lagomorpha family (Ochotonidae), were also determined. Differences between American pika and the three leporid genera was similar for CCL3 (0.237-0.258) and CCL4 (0.217-0.239) proteins, but much lower distance values were registered for CCL5 (0.077-0.132).

The majority of the differences between leporid chemokine inferred protein sequences are observed in the signal peptide, especially in CCL5 (Figure 1). This signal sequence is cleaved during protein maturation; thus, such differences are not expected to reflect any functional consequence. On the other hand, it has been described that the signaling and receptor-binding regions of chemokines are located in the flexible N-terminal segment and in the 20s loop (N-loop) of the mature protein, respectively [19]. In our study, the N-terminal region and the 20s loop (N-loop) match approximately amino acid 25-55 of the three protein precursors (Figure 1). The existence of species or genus-specific amino acid substitutions in leporid CCL3, CCL4 and CCL5 in these regions supports the performance of functional studies to perceive their influence on the chemokine ligand-receptor binding pattern.

Gene conversion, an event of recombination, between the chemokine receptors CCR2/CCR5 and CCR1/CCR3 has been identified in a number of mammals [9-14]. Therefore, recombination analysis was performed for the alignment of each CCR5 chemokine ligand, as it can affect in an adverse way the accuracy of positive selection inference [20]. No evidence of recombination was detected while running GARD, a genetic algorithm for recombination detection [21] (data not shown).

To look for signatures of natural selection operating in mammalian *CCL3*, *CCL4* and *CCL5* we used PAML [22] and compared site-based models (M1a vs. M2a and M7 vs. M8) to determine whether a model that allows positive selection (M2a and M8) is a better fit to the data than a neutral null model (M1a and M7). Through a likelihood ratio test (LRT), it was not possible to reject the null hypothesis of neutral selection for *CCL3* and *CCL4* genes, as none of the comparisons were statistically significant (Table 2). Comparison analysis for *CCL5* between models M1a and M2a retrieved an LRT of 2.20, supporting the non-rejection of the null hypothesis of neutral selection. On the other hand, *CCL5* comparison between models M7 and M8 yielded an LRT of 7.64 with a 5%

of significance and a positively selected site was identified (site 20, Pr=0.948). Nevertheless, this *CCL5* positively selected site is located in the signal peptide, which is cleaved during the protein maturation process.

Table 1 - Estimates of evolutionary divergence between Lagomorpha CCR5 chemokine ligands amino acid sequences

CCL3										
	1	2	3	4	5	6	7	8		
1. European rabbit (a)	-									
2. European rabbit (b) allele1	0.011	-								
3. European rabbit (b) allele2	0.000	0.011	-							
4. Brush rabbit	0.065	0.054	0.065	-						
5. European brown hare	0.075	0.065	0.075	0.065	-					
6. Iberian hare allele1	0.043	0.032	0.043	0.032	0.032	-				
7. Iberian hare allele2	0.075	0.065	0.075	0.065	0.000	0.032	-			
8. American pika	0.258	0.247	0.258	0.258	0.258	0.237	0.258	-		
CCL4										
	1	2	3	4	5	6	7	8	9	10
1. European rabbit (a)	-									
2. European rabbit (b)	0.022	-								
3. Brush rabbit allele1	0.043	0.043	-							
4. Brush rabbit allele2	0.054	0.054	0.011	-						
5. Brush rabbit allele3	0.065	0.065	0.022	0.011	-					
6. Brush rabbit allele4	0.054	0.054	0.011	0.022	0.011	-				
7. European brown hare allele1	0.065	0.043	0.043	0.054	0.065	0.054	-			
8. European brown hare allele2	0.043	0.043	0.022	0.033	0.043	0.033	0.022	-		
9. Iberian hare	0.054	0.054	0.033	0.043	0.054	0.043	0.033	0.033	-	
10. American pika	0.228	0.217	0.228	0.239	0.228	0.217	0.239	0.228	0.228	-
CCL5										
	1	2	3	4	5	6	7	8	9	
1. European rabbit (a)	-									
2. European rabbit (b)	0.000	-								
3. Brush rabbit allele1	0.055	0.055	-							
4. Brush rabbit allele2	0.066	0.066	0.022	-						
5. European brown hare allele1	0.044	0.044	0.055	0.066	-					
6. European brown hare allele2	0.033	0.033	0.044	0.055	0.011	-				
7. European brown hare allele3	0.044	0.044	0.055	0.066	0.022	0.011	-			
8. Iberian hare	0.044	0.044	0.055	0.066	0.022	0.011	0.022	-		
9. American pika	0.099	0.099	0.121	0.132	0.077	0.088	0.077	0.099	-	

Evolutionary analyses were conducted in MEGA5 [23]. All positions containing gaps and missing data were eliminated. Options used: amino acid distance, complete deletion and p-distance.
European rabbit (a): *Oryctolagus cuniculus* subspecies *cuniculus*; European rabbit (b): *Oryctolagus cuniculus* subspecies *algirus*.

Table 2 - *CCL3*, *CCL4* and *CCL5* model parameter estimates, d_N/d_S ratios, log likelihood values and test statistics for PAML site-based models

Gene	Model	Parameters	d_N/d_S	*p*	*l*	2Δ*l*
CCL3	M1a: Nearly Neutral	ω_0 = 0.13792 ω_1 = 1.00 p_0 = 0.73043 (p_1 = 0.26957)	0.3703	2	-1902.28	M1a vs. M2a: 0
	M2a: Selection	ω_0 = 0.13792 ω_1 = 1.00 ω_2 = 1.00 p_0 = 0.73043 p_1 = 0.19829 (p_2 = 0.07129)	0.3703	4	-1902.28	
	M7: Neutral, beta	p = 0.50224 q = 1.01699	0.3297	2	-1901.82	M7 vs. M8: 5.08
	M8: Selection, beta + ω	p_0 = 0.81494 p = 1.02901 q = 4.32831 (p_1 = 0.18506) ω = 1.15386	0.3678	4	-1899.28	
CCL4	M1a: Nearly Neutral	ω_0 = 0.09821 ω_1 = 1.00 p_0 = 0.75661 (p_1 = 0.24339)	0.3177	2	-1534.44	M1a vs. M2a: 0
	M2a: Selection	ω_0 = 0.09821 ω_1 = 1.00 ω_2 = 1.00 p_0 = 0.75661 p_1 = 0.19919 (p_2 = 0.04420)	0.3177	4	-1534.44	
	M7: Neutral, beta	p = 0.32056 q = 0.85888	0.2710	2	-1531.26	M7 vs. M8: 2.30
	M8: Selection, beta + ω	p_0 = 0.98270 p = 0.36723 q = 1.09572 (p_1 = 0.01730) ω = 2.40780	0.2869	4	-1530.11	
CCL5	M1a: Nearly Neutral	ω_0 = 0.11147 ω_1 = 1.00 p_0 = 0.74956 (p_1 = 0.25044)	0.3340	2	-2584.69	M1a vs. M2a: 2.20
	M2a: Selection	ω_0 = 0.11456 ω_1 = 1.00 ω_2 = 2.29256 p_0 = 0.74238 p_1 = 0.22901 (p_2 = 0.02861)	0.3796	4	-2583.59	
	M7: Neutral, beta	p = 0.33253 q = 0.93018	0.2623	2	-2559.86	M7 vs. M8: **7.64**
	M8: Selection, beta + ω	p_0 = 0.94502 p = 0.42600 q = 1.61116 (p_1 = 0.05498) ω = 1.76237	0.2920	4	-2556.04	

Significant LRT values are in bold.

Functional studies should be performed to further investigate all the observed differences in disease-related CCL3, CCL4 and CCL5 inferred protein sequences between leporid genera and possible consequences to pathogen susceptibility/resistance. Clearly, the three CCR5 chemokine ligands are under strong purifying selection as a result of possible functional binding constraints.

3. References

1. Zlotnik A, Yoshie O: Chemokines: a new classification system and their role in immunity. *Immunity* 2000, 12:121-127.
2. Fernandez EJ, Lolis E: Structure, function, and inhibition of chemokines. *Annual Review of Pharmacology and Toxicology* 2002, 42:469-499.
3. Zlotnik A, Yoshie O: The chemokine superfamily revisited. *Immunity* 2012, 36:705-716.
4. Mantovani A: The chemokine system: redundancy for robust outputs. *Immunology Today* 1999, 20:254-257.
5. Nei M, Rooney AP: Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics* 2005, 39:121-152.
6. DeVries ME, Kelvin AA, Xu L, Ran L, Robinson J, Kelvin DJ: Defining the origins and evolution of the chemokine/chemokine receptor system. *The Journal of Immunology* 2006, 176:401-415.
7. Zlotnik A, Yoshie O, Nomiyama H: The chemokine and chemokine receptor superfamilies and their molecular evolution. *Genome Biology* 2006, 7:243.
8. Nomiyama H, Osada N, Yoshie O: The evolution of mammalian chemokine genes. *Cytokine & Growth Factor Reviews* 2010, 21:253-262.
9. Shields DC: Gene conversion among chemokine receptors. *Gene* 2000, 246:239-245.
10. Carmo CR, Esteves PJ, Ferrand N, van der Loo W: Genetic variation at chemokine receptor CCR5 in leporids: alteration at the 2nd extracellular domain by gene conversion with CCR2 in *Oryctolagus*, but not in *Sylvilagus* and *Lepus* species. *Immunogenetics* 2006, 58:494-501.
11. Esteves PJ, Abrantes J, van der Loo W: Extensive gene conversion between CCR2 and CCR5 in domestic cat (*Felis catus*). *International Journal of Immunogenetics* 2007, 34:321-324.
12. Vazquez-Salat N, Yuhki N, Beck T, O'Brien SJ, Murphy WJ: Gene conversion between mammalian CCR2 and CCR5 chemokine receptor genes: a potential mechanism for receptor dimerization. *Genomics* 2007, 90:213-224.
13. Perelygin AA, Zharkikh AA, Astakhova NM, Lear TL, Brinton MA: Concerted evolution of vertebrate CCR2 and CCR5 genes and the origin of a recombinant equine CCR5/2 gene. *Journal of Heredity* 2008, 99:500-511.

14. Abrantes J, Carmo C, Matthee C, Yamada F, Loo W, Esteves P: A shared unusual genetic change at the chemokine receptor type 5 between *Oryctolagus*, *Bunolagus* and *Pentalagus*. *Conservation Genetics* 2011, 12:325-330.
15. van der Loo W, Afonso S, de Matos AL, Abrantes J, Esteves PJ: Pseudogenization of the MCP-2/CCL8 chemokine gene in European rabbit (genus *Oryctolagus*), but not in species of Cottontail rabbit (*Sylvilagus*) and Hare (*Lepus*). *BMC Genetics* 2012, 13:72.
16. Esteves PJ, Lanning D, Ferrand N, Knight KL, Zhai SK, Loo W: The evolution of the immunoglobulin heavy chain variable region (IgV H) in Leporids: an unusual case of transspecies polymorphism. *Immunogenetics* 2005, 57:874-882.
17. van der Loo W, Abrantes J, Esteves PJ: Sharing of endogenous lentiviral gene fragments among leporid lineages separated for more than 12 Million years. *Journal of Virology* 2009, 83:2386-2388.
18. de Matos AL, van der Loo W, Areal H, Lanning DK, Esteves PJ: Study of *Sylvilagus* rabbit TRIM5alpha species-specific domain: how ancient endoviruses could have shaped the antiviral repertoire in Lagomorpha. *BMC Evolutionary Biology* 2011, 11:294.
19. Laurence JS, Blanpain C, De Leener A, Parmentier M, LiWang PJ: Importance of basic residues and quaternary structure in the function of MIP-1 beta: CCR5 binding and cell surface sugar interactions. *Biochemistry* 2001, 40:4990-4999.
20. Shriner D, Nickle DC, Jensen MA, Mullins JI: Potential impact of recombination on sitewise approaches for detecting positive natural selection. *Genetics Research* 2003, 81:115-121.
21. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD: GARD: a genetic algorithm for recombination detection. *Bioinformatics* 2006, 22:3096-3098.
22. Yang Z: PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 2007, 24:1586-1591.
23. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 2011, 28:2731-2739.

Chapter 3

Innate anti-viral factors – Evolution and genetic characterization of sterile alpha motif domain-containing protein 9

Paper 2

Evolution and divergence of the mammalian *SAMD9*/*SAMD9L* gene family

A. Lemos de Matos, J. Liu , G. McFadden & P. J. Esteves

1. Summary

The physiological functions of the human sterile alpha motif domain-containing protein 9 (*SAMD9*) gene and its chromosomally adjacent paralogue, *SAMD9*-like (*SAMD9L*), currently remain unknown. However, the direct links between the deleterious mutations or deletions in these two genes and several human disorders, such as inherited inflammatory calcified tumors and acute myeloid leukemia, suggest their biological importance. SAMD9 and SAMD9L have also recently been shown to play key roles in the innate immune responses to stimuli such as viral infection. We were particularly interested in understanding the mammalian evolutionary history of these two genes. The phylogeny of *SAMD9* and *SAMD9L* genes was reconstructed using the Maximum Likelihood method. Furthermore, six different methods were applied to detect SAMD9 and SAMD9L codons under selective pressure: the site-specific model M8 implemented in the codeml program in the PAML software and five methods available on the Datamonkey web server, including the Single Likelihood Ancestor Counting method, the Fixed Effect Likelihood method, the Random Effect Likelihood method, the Mixed Effects Model of Evolution method and the Fast Unbiased Bayesian AppRoximation method. Additionally, the house mouse (*Mus musculus*) genome has lost the *SAMD9* gene, while keeping *SAMD9L* intact, prompting us to investigate whether this loss is a unique event during evolution.

Our evolutionary analyses suggest that *SAMD9* and *SAMD9L* arose through an ancestral gene duplication event after the divergence of Marsupialia from Placentalia. Additionally, selection analyses demonstrated that both genes have been subjected to positive evolutionary selection. The absence of either *SAMD9* or *SAMD9L* genes from

some mammalian species supports a partial functional redundancy between the two genes.

To the best of our knowledge, this work is the first study on the evolutionary history of mammalian *SAMD9* and *SAMD9L* genes. We conclude that evolutionary selective pressure has acted on both of these two genes since their divergence, suggesting their importance in multiple cellular processes, such as the immune responses to viral pathogens.

2. Introduction

The Sterile Alpha Motif Domain-containing 9 (*SAMD9*) gene is located in chromosome 7q21.2 of the human genome, and is adjacent to its close paralogue, *SAMD9*-like (*SAMD9L*), in a head-to-tail position [1, 2] and separated by approximately 12 kb. The physiological functions of both *SAMD9* and *SAMD9L* currently remain poorly understood, but the importance of human *SAMD9* has been recently emphasized during the discovery of the genetic cause of a rare life-threatening human disease, normophosphatemic familiar tumoral calcinosis (NFTC) [3, 4]. Patients with NFTC exhibited normal calcium and phosphate metabolism while developing calcified tumorous nodules at their extremities, accompanied by severe gingivitis. Two independent founder genetic events leading to the deleterious mutations in *SAMD9* are responsible for the autosomal recessive disease of NFTC [3, 4]. Interestingly, these patients and their kindred are from a culturally isolated ethical group, namely Jewish-Yemenite, suggesting a potential selection pressure associated with this population [3, 4]. In addition to NFTC, misregulated human *SAMD9* expression was also shown to be associated with aggressive fibromatosis, breast, and colon cancers [1].

Human *SAMD9* expression can be upregulated by tumor necrosis factor (*TNF*) [4] or by type I [5] and type II interferons (*IFN*s) [6], and it is classified as an interferon-stimulated gene (ISG). Recently, an interferon regulatory factor (*IRF-1*) binding element was identified in the promoter region of the *SAMD9* gene in humans [6], and overexpression of *IRF-1* can lead to elevated *SAMD9* gene expression [7]. All these observations suggest a key role of *SAMD9* as a signaling hub in response to innate immune stimulations. Most importantly, human *SAMD9* also has very recently been shown to possess anti-viral properties in cultured cells [8, 9] emphasizing its crucial role in host defense against viral pathogens.

On the other hand, the human *SAMD9L* gene was shown to exhibit lower expression levels in breast cancer tissue than in normal breast tissue from the same patient [1]. It was also identified to be an inducible gene for type I *IFN*s (*IFNα* and *β*),

and in activated human T cells the function of *SAMD9L* is correlated with its *IFN*-induced inhibitory effects on cell migration [10]. The murine *SAMD9L* gene expression was also found to be upregulated by calcitonin [11], suggesting a potential involvement in calcium homeostasis as well.

Lastly, the human *SAMD9* and *SAMD9L* genes were both classified as myeloid tumor suppressors, as they are localized within a microdeletion cluster associated with myeloid disorders, such as juvenile myelomonocytic leukemia (JMML), acute myeloid leukemia (AML), and myelodysplastic syndrome (MDS) [2]. In another study investigating altered immune responses in patients with metastatic melanoma, both *SAMD9* and *SAMD9L* expression were shown to be significantly reduced in T and B cell populations when compared with those from healthy control individuals [12]. It has been suggested that since these two proteins exhibit considerable sequence similarity, they may function redundantly or in related pathways, but it should be noted that patients with NFTC possess mutations only in *SAMD9* and thus it is likely that the two proteins perform non-identical tasks in humans.

Evolutionarily, the orthologous genes for both *SAMD9* and *SAMD9L* are highly conserved in many mammalian genomes, such as rat, primates and rabbit, but not in chicken, frog and fish species, or insects [1]. This suggests that the origin of these two related genes, possibly from an ancestral duplication event, occurred at some point after branching of the mammalian species. In addition, one intriguing fact is that the house mouse genome (*Mus musculus*, Mumu) has lost the *SAMD9* gene while maintaining *SAMD9L,* after an evolutionary chromosome breakage event [1].

The absence of *SAMD9* from the house mouse (Mumu) genome led us to question if it was a unique event restricted to this taxon and stimulated the study of *SAMD9* and *SAMD9L* evolution and divergence in different mammalian genomes. We have examined the evolutionary history and phylogeny of *SAMD9* and *SAMD9L,* using all the available and complete mammalian genomic sequences of both genes in NCBI and Ensembl databases, in order to obtain a broader understanding of the origin of these two genes. Our deduced phylogenetic tree suggests that *SAMD9* and *SAMD9L* indeed resulted from an ancestral gene duplication event that occurred after the divergence of Marsupialia from Placentalia. At the same time, we applied six different Maximum Likelihood (ML) methods to test for potential positive selective pressures exerted at the gene level, and we also looked for evidence of positive selection at the deduced protein level. The analyses revealed that SAMD9 and SAMD9L, at both the genome and deduced protein sequence levels, were under the effects of what appears to be sustained positive selective pressures. Our results suggest that these two proteins have been selected by long-term environmental pressures, such as those exerted by

pathogen responses that are under the control of innate immune regulators like the type I interferons.

3. Materials and Methods

3.1. SAMD9 and SAMD9L nucleotide and protein sequences

All the available mammalian *SAMD9* and *SAMD9L* genes coding sequences used in the phylogenetic and positive selection analyses were retrieved from NCBI (http://www.ncbi.nlm.nih.gov) and Ensembl (http://www.ensembl.org/index.html) databases. Next, sequences were aligned with Clustal W [13] implemented in BioEdit v7.0.9 [14], followed by visual inspection. Nucleotide sequences translation into protein sequences was performed using also BioEdit.

SAMD9 and *SAMD9L* genes coding sequences were collected for fifteen and nineteen species, respectively. Based on the Mammal Species of the World database classification (http://www.bucknell.edu/msw3/), representative species of mammalian infraclasses Metatheria (Order Didelphimorphia) and Eutheria (Order Artiodactyla, Carnivora, Chiroptera, Erinaceomorpha, Lagomorpha, Perissodactyla, Primates, Proboscidea, Rodentia and Soricomorpha) were included in this study. Table 1 summarizes the species collected for each gene and their respective accession numbers.

The isoeletric point (pI) of SAMD9 and SAMD9L deduced proteins for different species was estimated using DAMBE (Data Analysis and Molecular Biology and Evolution) [15].

3.2. Recombination and phylogenetic analyses

Recombination can mislead phylogenetic and positive selection analyses [16], and particularly for *SAMD9* and *SAMD9L*, the genes close location (~12 kb in human genome, for example) might increase the probability of recombination to occur. Therefore, we first performed recombination testing on placental *SAMD9* and *SAMD9L* nucleotide sequences alignments, and also on the alignment of both genes together (*SAMD9+SAMD9L*). The software GARD (Genetic Algorithm for Recombination Detection) [17, 18], implemented in the Datamonkey web server [19], was used to detect possible recombination breakpoints.

Table 1 - Mammalian *SAMD9* and *SAMD9L* genes accession numbers from species used in phylogenetic and selection analyses

SAMD9

Mammalian order	Common name	Species name	Database ID	Abbreviation
Artiodactyla	Cow	*Bos taurus*	Chromosome 4: 10,302,667-10,307,412[a]	SAMD9_Bota
	Pig	*Sus scrofa*	Chromosome 9: 79,679,836-79,684,587[a]	SAMD9_Susc
Chiroptera	Little brown myotis	*Myotis lucifugus*	Scaffold AAPE02063303: 7,766-12,520[a]	SAMD9_Mylu
Lagomorpha	European rabbit	*Oryctolagus cuniculus*	Chromosome 10: 35,728,133-35,732,926[a]	SAMD9_Orcu
Perissodactyla	Horse	*Equus caballus*	Chromosome 4: 36,749,161-36,753,927 [a]	SAMD9_Eqca
Primates	Common chimpanzee	*Pan troglodytes*	Chromosome 7: 92,731,148-92,735,917 [a]	SAMD9_Patr
	Human	*Homo sapiens*	Chromosome 7: 92,728,829-92,747,336 [a]	SAMD9_Hosa
	Northern white-cheeked gibbon	*Nomascus leucogenys*	SuperContig GL397261.1: 24,263,901-24,268,665 [a]	SAMD9_Nole
	Rhesus monkey	*Macaca mulatta*	Chromosome 3: 124,130,532-124,147,894 [a]	SAMD9_Mamu
	Sumatran orangutan	*Pongo abelii*	Chromosome 7: 83,034,053-83,038,819 [a]	SAMD9_Poab
	Western gorilla	*Gorilla gorilla*	Chromosome 7: 90,353,240-90,358,009 [a]	SAMD9_Gogo
Rodentia	Brown rat	*Rattus norvegicus*	XM_575365.2 [b]	SAMD9_Rano
	Chinese hamster	*Cricetulus griseus*	AFTD01024384.1 [b]	SAMD9_Crgr
	Domestic Guinea pig	*Cavia porcellus*	AAKN02016823.1 [b]	SAMD9_Capo
Soricomorpha	Common shrew	*Sorex araneus*	Scaffold_257382: 52,686-57,449 [a]	SAMD9_Soar

SAMD9L

Mammalian order	Common name	Species name	Database ID	Abbreviation
Carnivora	Domestic dog	*Canis lupus familiaris*	XM_539422.3[b]	SAMD9L_Calu
	Giant panda	*Ailuropoda melanoleuca*	Scaffold GL192585.1: 1,477,672-1,482,429[a]	SAMD9L_Aime
Didelphimorphia (Marsupialia)	Gray short-tailed opossum	*Monodelphis domestica*	XM_001378475.1[b]	SAMD9L_Modo
Erinaceomorpha	West European hedgehog	*Erinaceus europaeus*	GeneScaffold_8766: 48,007-52,945 [a]	SAMD9L_Ereu
Lagomorpha	European rabbit	*Oryctolagus cuniculus*	Chromosome 10: 35,699,236-35,703,990 [a]	SAMD9L_Orcu
Perissodactyla	Horse	*Equus caballus*	Chromosome 4: 36,788,011-36,792,765 [a]	SAMD9L_Eqca
Primates	Common chimpanzee	*Pan troglodytes*	Chromosome 7: 92,759,911-92,778,202 [a]	SAMD9L_Patr
	Common marmoset	*Callithrix jacchus*	Chromosome 8: 54,405,622-54,420,907 [a]	SAMD9L_Caja
	Human	*Homo sapiens*	Chromosome 7: 92,759,368-92,777,682 [a]	SAMD9L_Hosa
	Northern white-cheeked gibbon	*Nomascus leucogenys*	SuperContig GL397261.1: 24,263,209-24,320,238 [a]	SAMD9L_Nole
	Rhesus monkey	*Macaca mulatta*	Chromosome 3: 124,099,607-124,117,554 [a]	SAMD9L_Mamu
	Sumatran orangutan	*Pongo abelii*	Chromosome 7: 83,003,315-83,008,287 [a]	SAMD9L_Poab
	Western gorilla	*Gorilla gorilla*	Chromosome 7: 90,382,062-90,397,829 [a]	SAMD9L_Gogo
Proboscidea	African bush elephant	*Loxodonta africana*	XM_003407146.1[b]	SAMD9L_Loaf
Rodentia	Brown rat	*Rattus norvegicus*	Chromosome 4: 28,180,812-28,185,536 [a]	SAMD9L_Rano
	Chinese hamster	*Cricetulus griseus*	XM_003496952.1 [b]	SAMD9L_Crgr
	Domestic Guinea pig	*Cavia porcellus*	scaffold_11: 24,689,192-24,742,963 [a]	SAMD9L_Capo
	House mouse	*Mus musculus*	Chromosome 6: 3,322,257-3,349,571[a]	SAMD9L_Mumu
Soricomorpha	Common shrew	*Sorex araneus*	scaffold 194773: 6,206-10,964 [a]	SAMD9L_Soar

Database ID: [a]Ensembl; [b]NCBI GenBank.

For *SAMD9* and *SAMD9L* genes alignments no significant breakpoints were detected while using GARD, thus the complete alignments were used to establish each gene phylogeny. As indicated by the Akaike information criterion (AIC) implemented in jModelTest v0.1.1 [20], the nucleotide substitution model TVM+G was used for *SAMD9* tree estimation, while the GTR+G model was the consensus model selected for *SAMD9L* phylogenetic tree construction. On the other hand, a significant breakpoint was detected when running GARD for the *SAMD9*+*SAMD9L* alignment and a phylogenetic tree was estimated for each segment. For the left segment, the AIC in jModelTest indicated GTR+I+G as the best-fit nucleotide substitution model, whereas for the right segment the TPM2uf+G model was indicated as the best for the tree estimation. Also, for the *SAMD9*+*SAMD9L* alignment, a phylogenetic tree was estimated without testing recombination. In this case, the jModelTest AIC estimated GTR+I+G model as the best-fit nucleotide substitution model.

To establish mammalian phylogeny for *SAMD9*, *SAMD9L* and *SAMD9*+*SAMD9L*, based on nucleotide sequences, the Maximum Likelihood (ML) method implemented on GARLI v2.0 (Genetic Algorithm for Rapid Likelihood Inference) was used [21]. The analyses were performed with 1,000,000 generations and 1,000 bootstrap searches. ML trees were displayed using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/).

3.3. Codon-based analyses of positive selection

A useful measurement for identifying adaptive protein evolution is the nonsynonymous (d_N)/synonymous substitution (d_S) rate ($\omega = d_N/d_S$), where values of $\omega = 1$, < 1, and > 1 indicate neutral selection, negative selection, and positive selection, respectively [22, 23]. Naturally, and due to protein structural and functional constraints, ω is expected to be close to 0 and full protein analysis rarely detects positive selection [24]. As a result, several methods, based on models of codon substitution, have been developed to detect adaptive evolution (positive selection) at individual sites in a background of negative selection [25, 26]. We employed six different methods to detect sites under selection, and based on the methodology adopted by several authors [27, 28] only codons identified by at least three of the six used methods were considered to be under positive selection.

To detect selection based on the ratio ω and at the gene-level, for both *SAMD9* and *SAMD9L*, PAML v4.4 (Phylogenetic Analysis by Maximum Likelihood) [29, 30] was used and the codon frequency model F3x4 was fitted to both alignments. In the site-specific models that allow the ratio ω to vary among codons, we performed Likelihood Ratio Tests (LRTs) with 2 degrees of freedom to compare the following models (*NS*

sites): M1 (nearly neutral) with M2 (selection) and M7 (neutral, β distribution of ω<1) with M8 (selection, β distribution of ω>1). A significant LRT demonstrates that the selection model fits better than the neutral model [25, 26]. For model M8, a Bayes empirical Bayes (BEB) approach was employed to detect codons with a posterior probability >90% of being under selection [31]. Also the branch-site model A was performed for testing positive selection on individual sites along a specific lineage, called foreground branch, where the other lineages are background branches. In branch-site model A, three ω ratios are assumed for foreground ($0 < \omega_0 < 1$, $\omega_1 = 1$, $\omega_2 > 1$) and two ω ratios for background ($0 < \omega_0 < 1$, $\omega_1 = 1$). The null model is the same as model A, but $\omega_2 = 1$ is fixed. We also used BEB approach to calculate the posterior probability of a specific codon site and to identify those most likely to be under positive selection (posterior probability >90%) [31].

Both *SAMD9* and *SAMD9L* genes were also analyzed using HyPhy software implemented in the Datamonkey web server [19]. Datamonkey includes three classic ML methods to detect sites under selection: the Single Likelihood Ancestor Counting (SLAC) model, the Fixed Effect Likelihood (FEL) model and the Random Effect Likelihood (REL) model [32]. Besides these three methods, two other recently developed and implemented in the Datamonkey web server were applied to our dataset: the Mixed Effects Model of Evolution (MEME) that allows the distribution of ω to vary from site to site and also from branch to branch at a site, being capable of identifying both episodic and pervasive positive selection [33], and the Fast Unbiased Bayesian AppRoximation (FUBAR) method that can detect positive selection under a model faster than the existing fixed effects likelihood models through the introduction of an ultra-fast Markov chain Monte Carlo (MCMC) routine and that allows to visualize Bayesian inference for each site [34]. All these methods were run using the best model chosen by AIC on a defined Neighbor-Joining (NJ) phylogenetic tree after running GARD to detect recombination. To avoid a high false-positive rate, due to the reduced number of sequences [32], sites with *p*-values <0.1 for SLAC, FEL and MEME models, Bayes Factor >50 for REL model and a posterior probability >0.90 for FUBAR were accepted as candidates for selection.

From the HyPhy software available on the Datamonkey web server, we also run the PARRIS method used to detect if a proportion of sites in the alignment evolve with $d_N/d_S > 1$ and that accounts for synonymous rate variation and recombination [35].

3.4. Amino acid-based analyses of positive selection

By using TreeSAAP v3.2 (Selection of Amino Acid Properties based on Phylogenetic Trees) [36] it was possible to detect selection signatures at the amino acid level, more specifically, positively selected amino acid properties that result in radical structural and functional changes in local regions of the protein (destabilization). Properties that fell into categories 6 through 8 (the most radical values denoting positive destabilizing selection), presented z-score values of 3.09 and higher, and with a probability value of 0.001 were plotted in a sliding window (length=20).

Thirty-one amino acid properties were evaluated across SAMD9 and SAMD9L phylogenetic trees to identify protein regions that presented evidence of positive destabilization for each property. The thirty-one amino acid properties are the following: alpha-helical tendencies, average number of surrounding residues, beta-structure tendencies, bulkiness, buriedness, chromatographic index, coil tendencies, composition, compressibility, equilibrium constant (ionization of COOH), helical contact area, hydropathy, isoelectric point, long-range non-bonded energy, mean r.m.s. fluctuation displacement, molecular volume, molecular weight, normalized consensus hydrophobicity, partial specific volume, polar requirement, polarity, power to be at the C-terminal, power to be at the middle of alpha-helix, power to be at the N-terminal, refractive index, short and medium range non-bonded energy, solvent accessible reduction ratio, surrounding hydrophobicity, thermodynamic transfer hydrophobicity, total non-bonded energy and turn tendencies.

4. Results

4.1. *SAMD9* and *SAMD9L* genes prevalence in mammals

All the available and complete mammalian *SAMD9* and *SAMD9L* genes coding sequences in the NCBI and Ensembl databases were collected, resulting in a total of fifteen *SAMD9* and nineteen *SAMD9L* genomic sequences of different species indicated in Table 1. The species collected for *SAMD9* genes fit into seven Eutheria orders, commonly designated as placental mammals, while the taxa collected for *SAMD9L* genes fit into eight placental orders. The grey short-tailed opossum, a representative of the order Didelphimorphia traditionally included in Marsupialia (pouch mammals), was the only marsupial genome to possess a complete *SAMD9L* sequence.

Besides the complete *SAMD9* and *SAMD9L* coding sequences, several other non-complete *SAMD9* and *SAMD9L* mammalian genes, including full length mRNA-derived transcripts with many still-undetermined nucleotides (for example, the large

flying fox or the west European hedgehog *SAMD9* coding sequences, or the American pika *SAMD9L* sequence) or partial gene sequences (for example, the Ord's kangaroo rat *SAMD9L* or the Hoffmann's two-toed sloth *SAMD9* genes), have been already identified and annotated in Ensembl database. However, these incomplete sequences were not used in the phylogenetic and selection analyses performed in this study. Both the complete and the non-complete *SAMD9* and *SAMD9L* genes annotated in Ensembl are represented in Figure 1, allowing a broader view into this gene family distribution within the mammalian context.

Special reference has to be made to two particular complete sequences that were included in our evolutionary analyses: the northern white-cheeked gibbon (Nole) *SAMD9* and the domestic dog (Calu) *SAMD9L*. The northern white-cheeked gibbon has no *SAMD9* gene currently annotated in Ensembl. However, by comparing *SAMD9* sequences of other primates to the gibbon genome in Ensembl using BLAST analysis, we obtained a perfect match with a neighboring designated pseudogene of *SAMD9L*. Despite this biotype classification, we could not exclude this SAMD9 sequence from being considered as a *bona fide* gibbon *SAMD9* gene. Regarding the domestic dog *SAMD9L*, this gene is present in NCBI and is annotated in Ensembl, but in this latter database the sequence was missing seventy-four nucleotides when compared to the sequence in NCBI. Thus, for the subsequent analyses we used only the sequence from NCBI. It should also be noted that, despite not being annotated in Ensembl, an incomplete *SAMD9* sequence for the domestic dog is available in NCBI. However, when the NCBI sequence (XM_003639470.1) was analyzed by BLAST, it possessed 99 to 100% identity with a non-annotated region of chromosome 14. Since it is a non-complete nucleotide sequence, it was not used further for the study reported here.

When *SAMD9* and *SAMD9L* were mapped in human chromosome 7, orthologous counterparts of both genes were identified in the chimpanzee (Patr), dog (Calu) and rat (Rano), but in the house mouse (Mumu) genome there was only a single genetic correspondence to the *SAMD9L* open reading frame in chromosome 6 [1]. From what is currently available in Ensembl database, the absence of *SAMD9* for the house mouse (Mumu) is confirmed. We checked the other available rodents to confirm the presence or absence of *SAMD9* in this specific lineage. In Ensembl there is a single *SAMD9* annotation for the thirteen-lined ground squirrel (Ictr). In addition, what appear to be intact *SAMD9* genes have been deposited in NCBI database for the brown rat (Rano), the Chinese hamster (Crgr) and the domestic Guinea pig (Capo). On the other hand, like the house mouse (Mumu), the Ord's kangaroo rat (Dior) does not have *SAMD9* gene annotated in Ensembl database.

Forward strand
< CDK6
< SAMD9L
< CDK6
< SAMD9
< HEPACAM2
CCDC132 >
< CALCR
CALCR >
< CCDC132
HEPACAM2 >
SAMD9L >
CDK6 >
SAMD9
ORDER
SAMD9
SAMD9L
Didelphimorphia
Afrosoricida
Artiodactyla
Carnivora
Cingulata
Chiroptera
Erinaceomorpha
Hyracoidea
Lagomorpha
Perissodactyla
Pilosa
Primates
Proboscidea
Rodentia
Soricomorpha
Modo
Ecte
Bota
Susc
Vipa
Aime
Calu
Feca
Dano
Mylu
Ptva
Ereu
Prca
Ocpr
Orcu
Eqca
Chho
Caja
Gogo
Hosa
Mamu
Nole
Otga
Patr
Poab
Tasy
Loaf
Capo
Dior
Ictr
Mumu
Rano
Soar
Chromosome 8
GeneScaffold_7273
Chromosome 4
Chromosome 9
GeneScaffold_1246
Scaffold GL192585.1
Chromosome 14
GeneScaffold_4062
GeneScaffold_7428
Scaffold GL429782
GeneScaffold_3953
GeneScaffold_8766
GeneScaffold_6298
GeneScaffold_4408
Chromosome 10
Chromosome 4
Scaffold_65211
Chromosome 8
Chromosome 7
Chromosome 7
Chromosome 3
SuperContig GL397261.1
Scaffold GL873524.1
Chromosome 7
Chromosome 7
Scaffold_62435
SuperContig scaffold_5
Scaffold_11
GeneScaffold_5660
Scaffold JH393613.1
Chromosome 6
Chromosome 4
Scaffold_194773
Scaffold_257382
Scaffold JH393613.1
Scaffold AAPE02063303
GeneScaffold_8766
Chromosome 10
Chromosome 4
Chromosome 7
Chromosome 7
Chromosome 3
SuperContig GL397261.1
Chromosome 7
Chromosome 7
SuperContig scaffold_5
Scaffold_11
GeneScaffold_5660
Scaffold JH393613.1
Chromosome 6
Chromosome 4
Scaffold_194773
Scaffold_257382

Figure 1 - Ensembl annotation of mammalian *SAMD9* and *SAMD9L* genes, and neighboring genes.
Ensembl annotation of the available mammalian *SAMD9* and *SAMD9L* genes, both complete and incomplete coding region sequences, are represented. Complete sequences for both genes are highlighted (dark yellow for *SAMD9* and light yellow for *SAMD9L*) and were used in the evolutionary analyses. Other *SAMD9* and *SAMD9L* genes are already annotated in Ensembl, but at the time of these analyses were still incomplete, corresponding to the non-highlighted locations in the figure and were excluded from posterior analyses. Ensembl species by order of appearance: Modo - Opossum; Ecte - Lesser hedgehog tenrec; Bota - Cow; Susc - Pig; Vipa - Alpaca; Aime - Panda; Calu - Dog; Feca - Cat; Dano - Armadillo; Mylu - Microbat; Ptva - Megabat; Ereu - Hedgehog; Prca - Hyrax; Ocpr - Pika; Orcu - Rabbit; Eqca - Horse; Chho - Sloth; Caja - Marmoset; Gogo - Gorilla*;* Hosa - Human; Mamu - Macaque; Nole - Gibbon; Otga - Bushbaby; Patr - Chimpanzee; Poab - Orangutan; Tasy - Tarsier; Loaf - Elephant; Capo - Guinea pig; Dior - Kangaroo rat; Ictr - Squirrel; Mumu - Mouse; Rano - Rat; Soar - Shrew. To access the complete species name, the list of abbreviations should be consulted. Based on human chromosome 7 mapping, *SAMD9* and *SAMD9L* neighboring genes were identified and represented (*CDK6*, *HEPACAM2*, *CCDC132* and *CALCR*). Using Ensembl database, the same search was performed for the remaining species and the identified genes are represented under the forward strain arrow. The represented genes are, in most cases, the immediate neighboring genes, while for a reduced number of species some other genes are located in the same region, but were excluded for this purpose. The "<" symbol corresponds to the gene being located in the reverse strand, while the ">" symbol stands for a forward strand location.

4.2. Complete mammalian *SAMD9* and *SAMD9L* gene sequences: recombination and phylogenetic analyses

The complete nucleotide coding sequences from *SAMD9* and *SAMD9L* were aligned together (*SAMD9*+*SAMD9L*) and translated into deduced protein sequences (Supplementary Figure S1). Before further phylogenetic analyses, we used the software GARD [17, 18] to look for any evidence of recombination in the alignment. Three breakpoints were identified, but only one was strongly supported by the Kishino-Hasegawa (KH) test (Supplementary Table S1), which should result in the estimation of a phylogenetic tree for each segment. However, since the breakpoint was located on nucleotide 4755, the genomic segment to the right of the breakpoint was only composed of 150 nucleotides.

A Maximum Likelihood (ML) tree was estimated for the smallest genetic segment (not shown), but the nodes were weakly supported by low bootstrap values. Therefore, only the large segment with 4755 nucleotides was used to reconstruct a ML phylogenetic tree under the GTR+I+G nucleotide substitution model. The resulting tree is represented in Figure 2. Another ML phylogenetic tree was estimated, but without testing recombination, to compare differences in the tree topologies. The model used was again the GTR+I+G and resulted in a tree (Supplementary Figure S2) with a similar overall topology to the gene segment containing 4755 nucleotides.

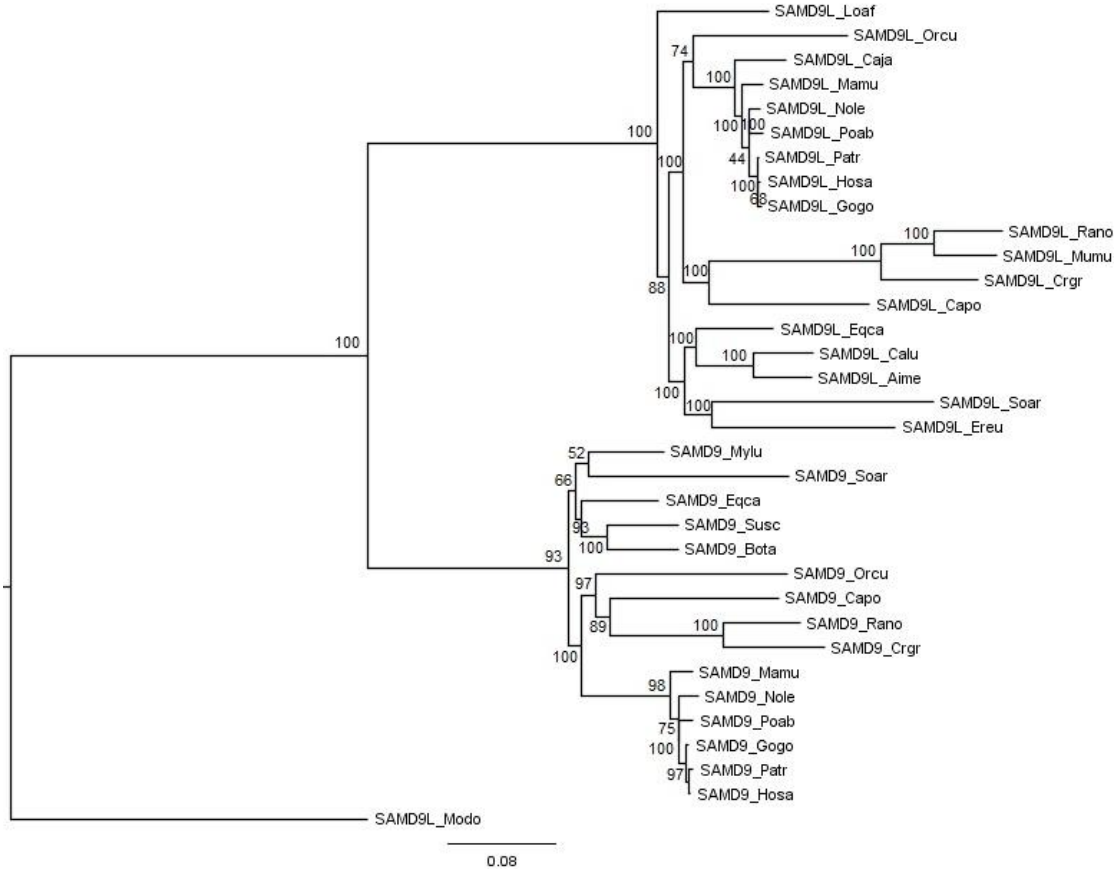


Figure 2 - Mammalian *SAMD9* and *SAMD9L* genes estimated Maximum Likelihood tree.

For the mammalian *SAMD9* and *SAMD9L* genes alignment and after GARD analysis [17, 18], a significant recombination breakpoint was detected (nucleotide position 4755) defining a left and a right segment. A phylogenetic tree was estimated for each segment using the Maximum Likelihood (ML) method. However, the resulting tree from the right segment presented weakly supported nodes and was discarded. On the other hand, the left segment with 4755 nucleotides was used to reconstruct a ML phylogenetic tree under the GTR+I+G nucleotide substitution model. The analyses were performed with 1,000,000 generations and 1,000 bootstrap searches. The bootstrap values are indicated on the branches. The abbreviations correspond to the following species common names: Aime - Giant panda; Bota - Cow; Caja - Common marmoset; Calu - Domestic dog; Capo - Domestic Guinea pig; Crgr - Chinese hamster; Eqca - Horse; Ereu - West European hedgehog; Gogo - Western gorilla; Hosa - Human; Loaf - African bush elephant; Mamu - Rhesus monkey; Modo - Grey short-tailed opossum; Mumu - House mouse; Mylu - Little brown myotis; Nole - Northern white-cheeked gibbon; Orcu - European rabbit; Patr - Common chimpanzee; Poab - Sumatran orangutan; Rano - Brown rat; Soar - Common shrew ; Susc - Pig. To access the species scientific names, the list of abbreviations should be consulted.

In the estimated ML phylogenetic tree (Figure 2), *SAMD9* and *SAMD9L* formed two well defined monophyletic groups, and within each clade we observed a concordant topology with the accepted evolutionary relationships of eutherian mammals [37] (Supplementary Figure S3). Interestingly, the marsupial grey short-tailed opossum (Modo) *SAMD9L* represented a highly divergent outgroup, even from the remaining *SAMD9L* species.

4.3. A gene duplication event after the split of marsupial and placental mammals originated *SAMD9*/*SAMD9L* gene family

It has been previously suggested that *SAMD9* and its paralogous *SAMD9L* may have originated from a common ancestor by a gene duplication event [1]. In our study, the ML tree (Figure 2) topology supports this view. However, the opossum (Modo) gene annotated as *SAMD9L* in NCBI database (XM_001378475.1) does not cluster in the placental mammal *SAMD9L* group. In fact, the opossum sequence can be recognized as being in a basal position. Two highly supported eutherian monophyletic clades in the ML tree, one corresponding to all *SAMD9* genes and the other one to all *SAMD9L* genes, were observed. The most likely evolutionary scenario can be described as following: an ancestral gene is present before the separation of marsupial from placental mammals in the common ancestor that originated the extant *SAMD9L* gene in the marsupial opossum (Modo) and the ancestral gene of placental *SAMD9*/*SAMD9L* gene family. Later, in placental mammals, this ancestral gene suffered an event of gene duplication resulting in the contemporary *SAMD9* and *SAMD9L* genes.

The conservation of similar arrangement of genes in the same relative locations on the chromosomes of different species, denominated as shared synteny, can indicate the existence of a common ancestor. In Ensembl, among the mammalian species where the presence of *SAMD9* and/or *SAMD9L* has been annotated, shared synteny can be readily observed in chromosomes and 'gene-scaffolds'. The consistent presence of the same common flanking genes (*CALCR*, *CCDC132*, *CDK6* and *HEPACAM2*) in different species supports the idea that *SAMD9* and *SAMD9L* are located in highly conserved regions throughout placental mammals' divergence and diversification (Figure 1).

4.4. Inference of positive selection at *SAMD9* and *SAMD9L* genes level

Placental SAMD9 and SAMD9L deduced protein sequences were aligned independently (Supplementary Figure S4; Supplementary Figure S5) and ML trees were estimated for each gene (Supplementary Figure S6; Supplementary Figure S7). Afterwards, we determined whether the *SAMD9* and *SAMD9L* genes might have been subject to positive selection pressures by comparing PAML codon-based nested models with and without positive selection using likelihood ratio tests (LRTs) [29, 30]. Both comparisons of M1 (nearly neutral) versus M2 (positive selection) and M7 (beta) versus M8 (beta and $\omega>1$) resulted in the rejection of the null hypothesis, strongly

supporting the finding of positive selection for both *SAMD9* and *SAMD9L* (<0.001; Table 2). We also used the PARRIS [35] method to detect if a proportion of sites in each gene alignment evolved under positive selection after accounting for the potentially confounding effects of recombination and synonymous site variation. Interestingly, only *SAMD9L* was found to be under selection when using this method (<0.05; Supplementary Table S2).

Table 2 - *SAMD9* and *SAMD9L* likelihood ratio test (LRT) for four site models from PAML software

Hypothesis		LRT		
Null Hypothesis	Alternative Hypothesis	-2ΔlnL	df	*p*-Value
Site Models				
SAMD9				
M1: nearly neutral	M2: positive selection	25.55	2	< 0.001***
M7: beta	M8: beta and ω>1	77.76	2	< 0.001***
SAMD9L				
M1: nearly neutral	M2: positive selection	51.10	2	< 0.001***
M7: beta	M8: beta and ω>1	97.44	2	< 0.001***

***, highly significant.

Six different methods were used to detect sites under selection for *SAMD9* and *SAMD9L* (Supplementary Table S3). For PAML software [29, 30], we used M8 model to detect sites under selection for *SAMD9* and *SAMD9L* phylogenetic trees, and the BEB approach was used to identify codons with a posterior probability >90%. The other five applied methods to detect sites under positive selection are available in the Datamonkey web server. In this study, we only considered a codon with evidence of selection when it was identified by at least three of the six used methods [27, 28] (Supplementary Table S3). Seventeen sites for SAMD9 and nineteen sites for SAMD9L were identified as candidates for sites under positive selection (Figure 3 and 4; Supplementary Table S3).

Amino acid substitutions can be either conservative or radical, depending on whether they lead to a change in a certain physicochemical property [38]. For the codons identified as being under selection, we investigated the alterations of charge and polarity between mammalian taxa. For SAMD9 all the detected codons (Figure 3) exhibited at least one physicochemical alteration across species and a maximum of five different combinations of properties were identified for codon 331. Primate species SAMD9 amino acid changes were quite conservative, since eleven codons exhibited

the same amino acid. Despite the low number of species available for Artiodactyla and Rodentia, we verified in each order a great number of amino acid physicochemical alterations *per* codon in the SAMD9 proteins. In addition, all SAMD9L codons under presumptive selection (Figure 4) exhibited physicochemical alterations across taxa and at least three properties were represented in each codon. A maximum of five different physicochemical properties were identified for codon position 452. In Primates, amino acid substitutions in SAMD9L were once again quite conservative, given that thirteen positions kept the same physicochemical properties even when amino acid substitutions happened. On the contrary, among the four Rodentia species, only three positions in SAMD9L presented the same physicochemical properties, but just one was in fact the same amino acid.

		Codons																
		48	88	279	331	352	383	491	513	731	872	993	1006	1116	1258	1320	1329	1398
Analyses	PAML M8																	
	SLAC																	
	FEL																	
	REL																	
	MEME																	
	FUBAR																	
Species	Hosa	W	M	L	Y	K	T	P	L	A	Q	N	E	G	P	V	S	S
	Patr	W	M	L	Y	K	T	P	L	A	Q	N	E	G	P	V	S	S
	Gogo	W	M	L	Y	K	T	P	L	A	Q	N	E	G	P	V	S	S
	Poab	W	M	L	Y	K	T	P	L	A	Q	N	K	D	P	V	S	S
	Nole	W	M	L	Y	K	T	S	L	A	Q	N	K	G	P	I	S	S
	Mamu	W	M	L	Y	K	T	P	S	A	Q	N	K	G	Q	T	S	S
	Bota	Y	S	L	V	P	R	-	L	E	Q	N	I	D	A	A	L	I
	Susc	Y	R	M	Q	P	K	-	L	T	N	N	T	N	P	A	S	S
	Eqca	W	K	L	D	S	T	S	I	E	N	K	D	S	S	V	S	P
	Mylu	F	K	L	H	K	A	P	L	R	Q	K	I	T	P	A	L	Q
	Orcu	F	Q	E	H	S	T	T	S	A	H	M	S	-	S	L	I	L
	Rano	W	K	L	E	P	A	V	L	K	Q	S	T	E	L	V	L	Q
	Crgr	W	R	L	E	S	E	L	S	K	Q	S	S	K	L	V	S	Q
	Capo	W	A	S	H	S	S	T	S	E	K	T	I	N	L	I	S	L
	Soar	L	K	Q	H	S	T	T	L	D	Q	C	L	E	S	G	T	T

Figure 3 - Positively-selected SAMD9 codons and respective physicochemical properties for each mammalian species. SAMD9 sites under positive selection identified by at least three of the six used Maximum Likelihood methods. The sites are numbered according to the SAMD9 deduced proteins alignment (Supplementary Figure S4). The abbreviations correspond to the following species common names: Hosa - Human; Patr - Common chimpanzee; Gogo - Western gorilla; Poab - Sumatran orangutan; Nole - Northern white-cheeked gibbon; Mamu - Rhesus monkey; Bota - Cow; Susc - Pig; Eqca - Horse; Mylu - Little brown myotis; Orcu - European rabbit; Rano - Brown rat; Crgr - Chinese hamster; Capo - Domestic Guinea pig; Soar - Common shrew. To access the species scientific names, the list of abbreviations should be consulted. The background colors represent amino acid properties: polar positive (yellow), polar negative (orange), polar neutral (green), non-polar neutral (purple), non-polar aliphatic (blue) and non-polar aromatic (pink).

		Codons																		
		39	156	260	267	340	357	362	452	586	606	653	776	978	1186	1229	1276	1308	1429	1474
Analyses	PAML M8																			
	SLAC																			
	FEL																			
	REL																			
	MEME																			
	FUBAR																			
Species	Hosa	S	L	A	V	L	S	V	M	H	T	V	A	A	P	K	I	T	S	R
	Patr	S	L	A	V	L	S	V	M	H	T	V	A	A	P	K	I	T	S	R
	Gogo	S	L	A	V	L	S	V	M	H	T	V	A	A	P	K	I	T	S	R
	Poab	S	L	S	V	L	S	V	M	H	T	V	A	A	P	K	I	T	S	R
	Nole	S	L	A	V	T	S	V	M	H	T	V	A	A	P	K	I	T	S	R
	Caja	N	L	V	L	S	P	V	Q	H	T	A	V	S	P	E	I	T	S	H
	Mamu	S	L	A	L	S	S	V	V	H	T	V	A	A	P	I	I	T	S	R
	Loaf	N	V	D	L	Y	V	A	V	Q	T	L	T	A	L	Q	V	T	T	H
	Eqca	N	L	S	V	H	G	T	Q	Q	A	I	A	T	P	K	A	M	R	N
	Calu	R	L	A	I	L	V	L	V	H	T	I	E	T	K	K	D	V	S	S
	Aime	T	T	A	V	L	V	I	V	Q	T	T	G	A	P	K	N	G	S	S
	Ereu	C	R	A	L	S	G	K	E	Q	T	F	A	A	T	Q	A	I	N	L
	Orcu	N	I	V	V	Y	A	P	G	Q	S	T	A	S	A	L	I	R	H	R
	Mumu	K	P	I	T	T	P	R	H	A	A	I	V	S	P	K	G	V	T	H
	Crgr	N	A	V	K	N	A	R	Q	H	A	T	L	L	P	I	L	A	G	H
	Rano	E	P	I	V	T	P	R	H	S	A	I	I	S	P	K	L	A	T	H
	Capo	N	L	A	L	L	V	I	R	K	S	L	E	A	P	L	S	T	N	R
	Soar	N	L	S	K	T	E	A	V	D	E	T	A	K	Q	E	A	I	D	R

Figure 4 - Positively-selected SAMD9L codons and respective physicochemical properties for each mammalian species. SAMD9L sites under positive selection identified by at least three of the six used Maximum Likelihood methods. The sites are numbered according to the SAMD9L deduced proteins alignment (Supplementary Figure S5). The abbreviations correspond to the following species common names: Hosa - Human; Patr - Common chimpanzee; Gogo - Western gorilla; Poab - Sumatran orangutan; Nole - Northern white-cheeked gibbon; Caja - Common marmoset; Mamu - Rhesus monkey; Loaf - African bush elephant; Eqca - Horse; Calu - Domestic dog; Aime - Giant panda; Ereu - West European hedgehog; Orcu - European rabbit; Mumu - House mouse; Crgr - Chinese hamster; Rano - Brown rat; Capo - Domestic Guinea pig; Soar - Common shrew. To access the species scientific names, the list of abbreviations should be consulted. The background colors represent amino acid properties: polar positive (yellow), polar negative (orange), polar neutral (green), non-polar neutral (purple), non-polar aliphatic (blue) and non-polar aromatic (pink).

To detect whether some sites along particular *SAMD9* and *SAMD9L* lineages were under positive selection, we employed branch-site Model A (Table 3). On the *SAMD9* phylogenetic tree we identified six branches (foreground branches) with ω ratio greater than 1, but only the common shrew (Soar) branch had a statistical significant LRT (<0.01). *SAMD9L* branch-site analysis revealed a total of twelve branches with ω ratio greater than 1, yet only four of those branches presented a statistical significant LRT. Both the Sumatran orangutan (Poab) and the domestic Guinea pig (Capo) branches had statistical significance <0.05, while the west European hedgehog (Ereu) and the common shrew (Soar) branches had statistical significance <0.01.

Table 3 - *SAMD9* and *SAMD9L* parameter estimates and likelihood ratio test (LRT) for branch-site model A (PAML)

Branch-site Model A		LRT			
Foreground branches [a]	**Parameter estimates**	**-2ΔlnL** [b]	**df** [c]	***p*-Value**	**Positively selected sites** [d]
SAMD9					
Gogo	p_0= 0.693 p_1= 0.291 p_{2a}= 0.011 p_{2b}= 0.005 ω_0= 0.096 ω_1= 1.000 ω_2= **6.192**	0.20	1	n.s.	none
Poab	p_0= 0.701 p_1= 0.293 p_{2a}= 0.004 p_{2b}= 0.002 ω_0= 0.096 ω_1= 1.000 ω_2= **21.339**	3.65	1	n.s.	none
Nole	p_0= 0.692 p_1= 0.291 p_{2a}= 0.012 p_{2b}= 0.005 ω_0= 0.095 ω_1= 1.000 ω_2= **2.555**	0.12	1	n.s.	none
Orcu	p_0= 0.696 p_1= 0.290 p_{2a}= 0.010 p_{2b}= 0.004 ω_0= 0.094 ω_1= 1.000 ω_2= **4.818**	3.30	1	n.s.	none
Capo	p_0= 0.702 p_1= 0.292 p_{2a}= 0.005 p_{2b}= 0.002 ω_0= 0.096 ω_1= 1.000 ω_2= **5.728**	1.25	1	n.s.	none
Soar	p_0= 0.696 p_1= 0.286 p_{2a}= 0.013 p_{2b}= 0.005 ω_0= 0.094 ω_1= 1.000 ω_2= **8.165**	**9.56**	1	**< 0.01**	288, 572
SAMD9L					
Poab	p_0= 0.729 p_1= 0.270 p_{2a}= 0.001 p_{2b}= 0.000 ω_0= 0.139 ω_1= 1.000 ω_2= **409.279**	**6.59**	1	**< 0.05**	888
Caja	p_0= 0.714 p_1= 0.263 p_{2a}= 0.016 p_{2b}= 0.006 ω_0= 0.138 ω_1= 1.000 ω_2= **3.169**	1.12	1	n.s.	none
Mamu	p_0= 0.727 p_1= 0.268 p_{2a}= 0.004 p_{2b}= 0.001 ω_0= 0.140 ω_1= 1.000 ω_2= **11.372**	1.22	1	n.s.	none
Loaf	p_0= 0.717 p_1= 0.262 p_{2a}= 0.015 p_{2b}= 0.006 ω_0= 0.139 ω_1= 1.000 ω_2= **2.244**	0.80	1	n.s.	none
Calu	p_0= 0.730 p_1= 0.269 p_{2a}= 0.013 p_{2b}= 0.000 ω_0= 0.140 ω_1= 1.000 ω_2= **20.273**	0.52	1	n.s.	none
Aime	p_0= 0.728 p_1= 0.269 p_{2a}= 0.003 p_{2b}= 0.001 ω_0= 0.139 ω_1= 1.000 ω_2= **16.318**	2.51	1	n.s.	none
Ereu	p_0= 0.725 p_1= 0.266 p_{2a}= 0.006 p_{2b}= 0.002 ω_0= 0.139 ω_1= 1.000 ω_2= **998.998**	**7.45**	1	**< 0.01**	none
Mumu	p_0= 0.716 p_1= 0.264 p_{2a}= 0.015 p_{2b}= 0.005 ω_0= 0.138 ω_1= 1.000 ω_2= **3.755**	0.34	1	n.s.	none
Crgr	p_0= 0.728 p_1= 0.268 p_{2a}= 0.003 p_{2b}= 0.001 ω_0= 0.139 ω_1= 1.000 ω_2= **38.672**	2.62	1	n.s.	none
Rano	p_0= 0.727 p_1= 0.268 p_{2a}= 0.004 p_{2b}= 0.001 ω_0= 0.139 ω_1= 1.000 ω_2= **11.843**	2.14	1	n.s.	none
Capo	p_0= 0.722 p_1= 0.261 p_{2a}= 0.012 p_{2b}= 0.004 ω_0= 0.139 ω_1= 1.000 ω_2= **6.984**	**6.36**	1	**< 0.05**	861
Soar	p_0= 0.717 p_1= 0.264 p_{2a}= 0.014 p_{2b}= 0.005 ω_0= 0.137 ω_1= 1.000 ω_2= **7.759**	**10.49**	1	**< 0.01**	84, 1338, 1346

[a] Species names on the foreground branches by order of appearance: Gogo - Western gorilla; Poab - Sumatran orangutan; Nole - Northern white-cheeked gibbon; Orcu - European rabbit; Capo - Domestic Guinea pig; Soar - Common shrew; Caja - Common marmoset; Mamu - Rhesus monkey; Loaf - African bush elephant; Calu - Domestic dog; Aime - Giant panda; Ereu - West European hedgehog; Mumu - House mouse; Crgr - Chinese hamster; Rano - Brown rat.

[b] -2ΔlnL: likelihood ratio test (LRT) to detect positive selection.

[c] df: degrees of freedom.

[d] Positively selected sites: posterior probabilities > 90% in the BEB (Bayes empirical Bayes) analyses.

4.5. Inference of positive selection at SAMD9 and SAMD9L deduced proteins level

The evaluation of destabilizing radical changes that may occur in specific regions of proteins should complement the information obtained from positive selection analyses at the gene level. Using TreeSAAP software, it is possible to estimate, from a phylogenetic tree, the amino acid properties under selection from the thirty-one available in the software [36] (see Materials and Methods section for full list of the thirty-one properties).

For both SAMD9 and SAMD9L phylogenetic trees, the two amino acid properties with the most radical value (category 8) denoting positive destabilizing selection were the isoelectric point (pI) and the equilibrium constant (ionization of COOH) (Supplementary Table S4). When comparing the pI values among species for each protein, we observed a high variability across them, especially for SAMD9L taxa (Figure 5). For SAMD9 proteins, both the cow (Bota) and the domestic Guinea pig (Capo) exhibited the lowest pI (7.60), while a pI of 8.11 for the northern white-cheeked gibbon was the highest observed in SAMD9 proteins. SAMD9L proteins from placental mammals exhibited a larger range for the pI values with the giant panda (Aime) presenting the lowest pI (6.85) and the horse (Eqca) exhibiting the highest pI (8.22). Interestingly, the marsupial grey short-tailed opossum SAMD9L deduced protein presented the lowest pI (6.74) of all. The differences in the pI, and especially in SAMD9L proteins, may cause dramatic effects on proteins folding, since those changes are caused by significant differences in the polarity of the amino acids that compose the proteins. Besides the pI and equilibrium constant, SAMD9 presented two other properties under strong positive destabilizing selection, while five more properties were identified as being under positive destabilizing selection for the SAMD9L alignment (Supplementary Table S4).

Regarding the SAMD9 sliding window, the four amino acid properties with significant z-Score values (>3.09) were evenly distributed along the SAMD9 proteins alignment (Figure 6). However, a superior concentration of higher z-Score values was observed in the region between amino acid 660 and 910, specifically for the pI. The SAMD9L sliding window showed a dense pattern for the seven amino acid properties under destabilizing selection (Figure 7). Yet, two regions of SAMD9L proteins alignment presented an even larger density of properties and the highest z-Score values for some of those properties: amino acid range of 208-431 and the range of 863-1430.

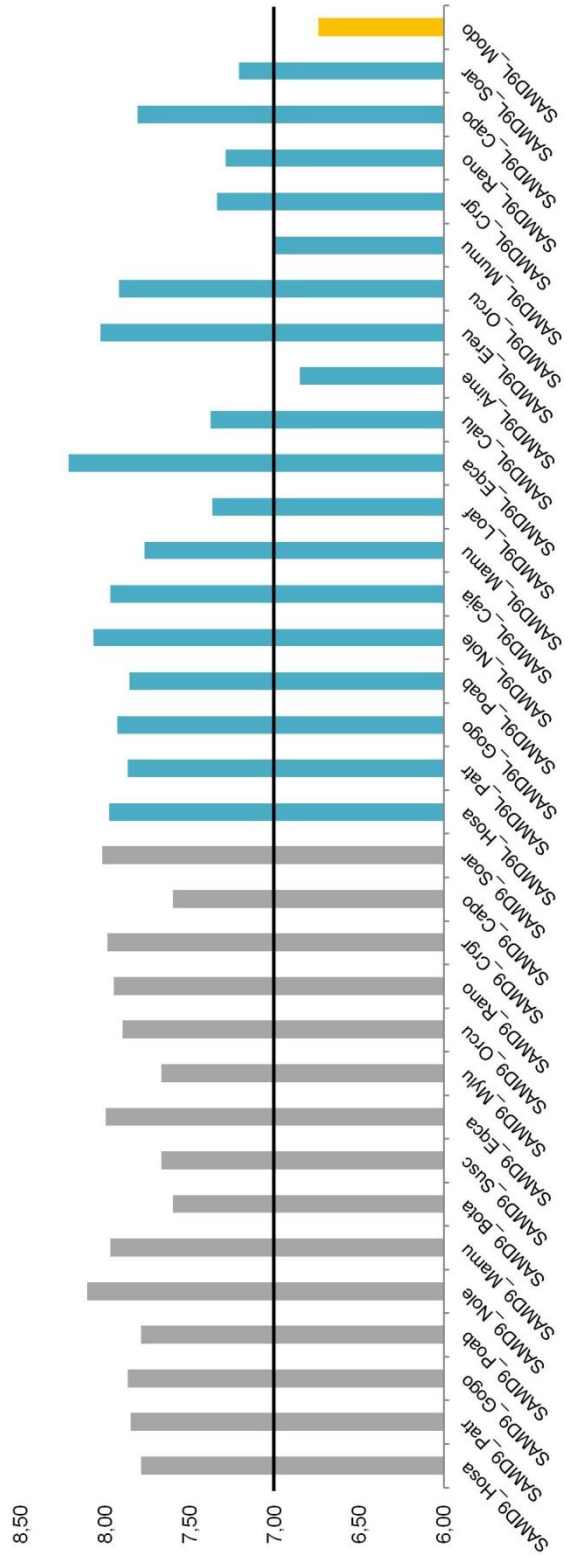


Figure 5 - Mammalian SAMD9 and SAMD9L deduced proteins isoelectric points (pI).

The grey bars correspond to the SAMD9 deduced proteins pI, the blue bars to the SAMD9L deduced proteins pI and the yellow bar to the opossum (Modo) SAMD9L deduced protein pI. The abbreviations correspond to the following species common names: Hosa - Human; Patr - Common chimpanzee; Gogo - Western gorilla; Poab - Sumatran orangutan; Nole - Northern white-cheeked gibbon; Mamu - Rhesus monkey; Bota - Cow; Susc - Pig; Eqca - Horse; Mylu - Little brown myotis; Orcu - European rabbit; Rano - Brown rat; Crgr - Chinese hamster; Capo - Domestic Guinea pig; Soar - Common shrew; Caja - Common marmoset; Loaf - African bush elephant; Calu - Domestic dog; Aime - Giant panda; Ereu - West European hedgehog; Mumu - House mouse; Modo - Grey short-tailed opossum. To access the species scientific names, the list of abbreviations should be consulted.

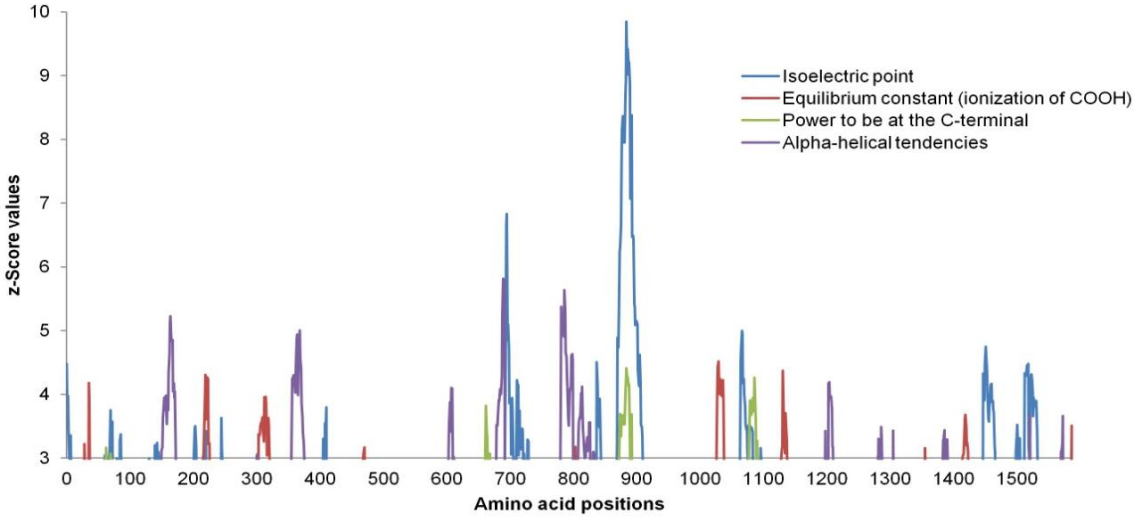


Figure 6 - Sliding window for SAMD9 amino acid properties under positive selection.
SAMD9 amino acid properties under destabilizing selection with significant z-Score values (>3.09).

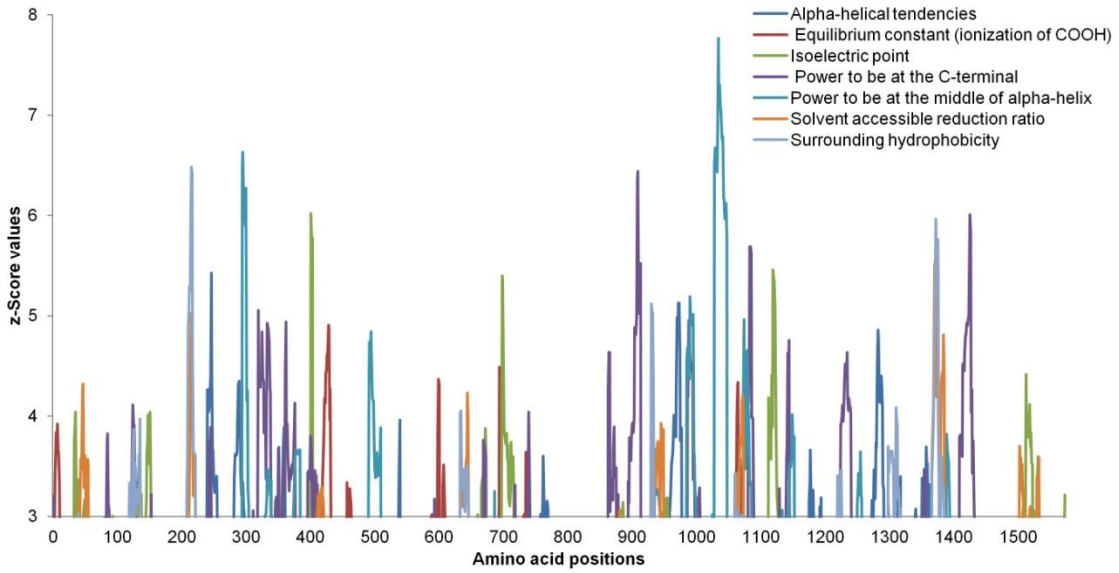


Figure 7 - Sliding window for SAMD9L amino acid properties under positive selection.
SAMD9L amino acid properties under destabilizing selection with significant z-Score values (>3.09).

5. Discussion

From a previous study, *SAMD9* and its paralogue *SAMD9L* have been identified in a variety of species, namely in human, chimpanzee, dog and rat. However, in the house

mouse (*Mus musculus*, Mumu) genome, *SAMD9* was uniquely lost [1]. The same study indicated the absence of both genes in chicken, frog and all currently sequenced fish species, suggesting that the *SAMD9*/*SAMD9L* genes originating event had occurred after the mammalian radiation. One of our goals was to intensify the identification of *SAMD9* and *SAMD9L* within different mammalian genomes and also verify whether the loss of mouse *SAMD9* was a unique event restricted to this taxon.

Despite the great number of morphological, molecular and phylogenetic studies for the order Rodentia, controversies relating to the divergence times between its major suborders still persist [39]. In a recent study on rodent evolution [40] some internal rodent branches have been resolved, where three main groups in the phylogenetic tree were supported: the Mouse-related clade, Ctenohystrica clade and the Squirrel-related clade. A scenario has been proposed where the pre-Squirrel-related clade diverged early from the common ancestor followed by a later separation of the pre-Mouse-related and pre-Ctenohystrica clade [40]. We gathered sequences for one or both *SAMD9* and *SAMD9L* genes for species representative of the three clades. The two genes were present in the thirteen-lined ground squirrel (Squirrel-related clade), the domestic Guinea pig (Ctenohystrica clade), the Chinese hamster and the brown rat (Mouse-related clade). Together with the absence of *SAMD9* in the house mouse genome, the Ord's kangaroo rat (Mouse-related clade) also did not have this gene annotated in Ensembl. With the apparent region synteny for the Ord's kangaroo rat when compared to the other mammals, this absence might just be the case of a genome still to be completely annotated, leaving the house mouse as the only rodent taxon that has lost *SAMD9*, at least from the currently available genomic sequence database.

A great number of the available mammalian genomes are still not completely annotated. Therefore, we made no assumptions regarding *SAMD9* and *SAMD9L* for those species. Nevertheless, we observed that the fairly well annotated cow and pig genomes (Order Artiodactyla) had no matches or annotations for *SAMD9L*. This information together with the absence of *SAMD9* in the house mouse and the already suggested origin of both genes from a common ancestor by ancient gene duplication [1] led us to the following hypothesis: in some lineages the presence of both genes might be costly for the genome, resulting in the loss of one of the genes that functionally would be overcome by the remaining paralogue. Although these observations support the potential existence of certain gene redundancy between *SAMD9* and *SAMD9L*, we also note the almost nonexistent recombination between them, despite the proximity in the location of these two genes in the genomes of all the annotated mammalian species. This genetic isolation of the two paralogues does not

support the existence of functional redundancy between *SAMD9* and *SAMD9L*. These apparent contradictory hypotheses have to be confirmed with the conduction of functional studies in different species.

With all the available mammalian sequences collected for both *SAMD9* and *SAMD9L* genes, the performed phylogenetic study resulted in a tree with a well-defined monophyletic group *per* gene gathering solely placental mammals and a single outgroup, the marsupial grey short-tailed opossum. This supported the speculative hypothesis of *SAMD9* and *SAMD9L* resulting from a gene duplication event, more precisely, after the divergence of Marsupialia from Placentalia 147.7 Mya [41]. Despite the common ancestor, when testing for the occurrence of potential positive selection acting at the gene and protein levels, we concluded that SAMD9L is under stronger selection than SAMD9. This is supported by the fact that a higher number of sites at the gene level and of specific lineages were positively selected in SAMD9L than SAMD9. Besides, a greater number of amino acid properties were under selection at the deduced protein level of SAMD9L than SAMD9.

When we examined the amino acid substitutions and changes on physicochemical properties for sites under selection, it was clear, for both proteins, that members of the Rodentia order presented the highest number of divergent alterations for the same residues compared to other mammalian orders. Since it is known that in many proteins the amino acid substitutions caused by positive selection are not random [38, 42], for instances the Primate APOBEC3G residues involved in HIV-1 Vif interaction [43], we hypothesize that any occurring alteration in rodents or even in other lineages may be the result of consistent arms race between the host and a pathogen stressor. This could be a significant observation, given that anti-viral properties have been already assigned to human SAMD9 in cultured human cells. Specifically, a unique viral gene product, M062 of myxoma virus, was found to antagonize the anti-viral properties of SAMD9 protein in order to permit the replication of this virus in cultured human cells [8].

Considering the mammalian species included in this study, selection analyses performed on *SAMD9* and/or *SAMD9L* genes for each species individually one may have different results from the obtained in our work, since recombination rates and effective population sizes are expected to differ among species. These species and population specific selection analyses should result in the identification of sites under selection in *SAMD9* and/or *SAMD9L* genes that can be used in genetic population studies by determining parameters like allele and genotype frequencies, and F_{ST} and nucleotide diversity values. This contributes to the definition of genotypes that might be favorable or not, for example, to the defense against certain pathogens.

Human SAMD9 and SAMD9L have solely one defined domain, the sterile alpha motif (SAM), a module of about 70 amino acid residues long [44], specifically 65 amino acids and 66 in SAMD9 and SAMD9L, respectively. SAM domains, one of the most common protein domains found in eukaryotic cells, are protein-protein interaction modules that perform a large number of different functions [45, 46] and are not easily categorized. Indeed, different SAM domains can self-associate, bind to other SAM domains and/or to non-SAM proteins, and even interact with RNA, DNA or lipids [46]. Because of the great variety of known functions, the presence of a SAM domain does not necessarily involve a specific function or pathway, but an array of possible functions. For both human SAMD9 and SAMD9L, no function has yet been assigned to their SAM domains, but for SAMD9 the ability to form SAM polymers has been suggested [47]. From our evolutionary study on both proteins, none of the identified sites or amino acid properties under positive selection overlapped with the deduced SAM domains, demonstrating a high level of conservation among the mammalian species.

6. Conclusions

Since the origin and evolution of the *SAMD9* and *SAMD9L* genes were first reported, a great number of mammalian genomes have been sequenced, allowing now a more detailed view into the evolutionary history of both genes. Our study supports the previously suggested origin of *SAMD9* and *SAMD9L* from a mammalian ancestral duplication event. Specifically, according to the results from our study, this event occurred after the divergence of Marsupialia from Placentalia. When considering the mostly complete mammalian genomes collected for this study, the apparent loss of *SAMD9* or *SAMD9L* in some species led us to propose that some overlapping functional redundancy exists between the two proteins, despite the almost nonexistent recombination between the two closely located genes from other species. From the positive selection analyses performed, both at gene and protein levels, we demonstrate that SAMD9 and SAMD9L continue to be under long-term selective pressure, with even stronger evidence for positive selection in SAMD9L.

Both *SAMD9* and *SAMD9L* genes are upregulated by type I interferon, a classic feature associated with many innate pathogen-response genes called interferon-stimulated genes (ISGs). Indeed, human SAMD9 has already been shown to be a functional inhibitor for at least one viral pathogen, a poxvirus called myxoma virus, that expresses a specific viral inhibitor (M062) that counteracts the anti-viral properties of

SAMD9 [8]. Our results suggest that at least the SAMD9 genes may have been under sustained selection pressure exerted by viral pathogens.

Our work is the first complete study to investigate the evolutionary history of mammalian SAMD9 and SAMD9L.

7. List of abbreviations

Aime (Giant panda - *Ailuropoda melanoleuca*); Bota (Cow - *Bos taurus*); Caja (Common marmoset - *Callithrix jacchus*); Calu (Domestic dog - *Canis lupus familiaris*); Capo (Domestic Guinea pig - *Cavia porcellus*); Chho (Hoffmann's two-toed sloth - *Choloepus hoffmanni*); Crgr (Chinese hamster - *Cricetulus griseus*); Dano (Nine-banded armadillo - *Dasypus novemcinctus*); Dior (Ord's kangaroo rat - *Dipodomys ordii*); Ecte (Lesser hedgehog tenrec - *Echinops telfairi*); Eqca (Horse - *Equus caballus*); Ereu (West European hedgehog - *Erinaceus europaeus*); Feca (Domestic cat - *Felis catus*); Gogo (Western gorilla - *Gorilla gorilla*); Hosa (Human - *Homo sapiens*); Ictr (Thirteen-lined ground squirrel - *Ictidomys tridecemlineatus*); Loaf (African bush elephant - *Loxodonta africana*); Mamu (Rhesus monkey - *Macaca mulatta*); Modo (Grey short-tailed opossum - *Monodelphis domestica*); Mumu (House mouse - *Mus musculus*); Mylu (Little brown myotis - *Myotis lucifugus*); Nole (Northern white-cheeked gibbon - *Nomascus leucogenys*); Ocpr (American pika - *Ochotona princeps*); Orcu (European rabbit - *Oryctolagus cuniculus*); Otga (Northern greater galago - *Otolemur garnettii*); Patr (Common chimpanzee - *Pan troglodytes*); Poab (Sumatran orangutan - *Pongo abelii*); Prca (Rock hyrax - *Procavia capensis*); Ptva (Large flying fox - *Pteropus vampyrus*); Rano (Brown rat - *Rattus norvegicus*); Soar (Common shrew - *Sorex araneus*); Susc (Pig - *Sus scrofa*); Tasy (Philippine tarsier - *Tarsius syrichta*); Vipa (Alpaca - *Vicugna pacos*).

8. Supplementary material

Supplementary material is appended to the present document by order of appearance in the main text.

Supplementary Figure S1 - Mammalian SAMD9 and SAMD9L deduced protein sequences alignment.
SAMD9 and *SAMD9L* genes coding sequences were collected for fifteen and nineteen species, respectively. Sequences were aligned with Clustal W [13] implemented in BioEdit [14]. The abbreviations correspond to the following species common names:

Hosa - Human; Patr - Common chimpanzee; Gogo - Western gorilla; Poab - Sumatran orangutan; Nole - Northern white-cheeked gibbon; Mamu - Rhesus monkey; Bota - Cow; Susc - Pig; Eqca - Horse; Mylu - Little brown myotis; Orcu - European rabbit; Rano - Brown rat; Crgr - Chinese hamster; Capo - Domestic Guinea pig; Soar - Common shrew; Caja - Common marmoset; Loaf - African bush elephant; Calu - Domestic dog; Aime - Giant panda; Ereu - West European hedgehog; Mumu - House mouse; Modo - Grey short-tailed opossum. To access the species scientific names, the list of abbreviations should be consulted. Residues positions are numbered according to human SAMD9 protein. "?" represents undetermined codons; "." represents identity with the reference sequence of human SAMD9 protein.

Supplementary Table S1 - Detection of recombination breakpoints from *SAMD9* and *SAMD9L* genes alignment using GARD analysis.
SAMD9 and *SAMD9L* complete coding sequences were aligned together and the software GARD [17, 18] was used to look for any evidence of recombination. Three breakpoints were identified, but only one was strongly supported by the Kishino-Hasegawa (KH) test.

Supplementary Figure S2 - Mammalian *SAMD9* and *SAMD9L* genes estimated Maximum Likelihood tree without testing recombination.
A phylogenetic tree was estimated for the mammalian *SAMD9* and *SAMD9L* genes alignment using the Maximum Likelihood method and under the GTR+I+G nucleotide substitution model. The analyses were performed with 1,000,000 generations and 1,000 bootstrap searches. The bootstrap values are indicated on the branches. The abbreviations correspond to the following species common names: Aime - Giant panda; Bota - Cow; Caja - Common marmoset; Calu - Domestic dog; Capo - Domestic Guinea pig; Crgr - Chinese hamster; Eqca - Horse; Ereu - West European hedgehog; Gogo - Western gorilla; Hosa - Human; Loaf - African bush elephant; Mamu - Rhesus monkey; Modo - Grey short-tailed opossum; Mumu - House mouse; Mylu - Little brown myotis; Nole - Northern white-cheeked gibbon; Orcu - European rabbit; Patr - Common chimpanzee; Poab - Sumatran orangutan; Rano - Brown rat; Soar - Common shrew ; Susc - Pig. To access the species scientific names, the list of abbreviations should be consulted.

Supplementary Figure S3 - Evolutionary relationships of eutherian mammals.
Placental mammals' evolutionary relationships tree retrieved and adapted from Song *et al.* [37].

Supplementary Figure S4 - Mammalian SAMD9 deduced protein sequences alignment.
SAMD9 deduced protein sequences from fifteen species were aligned with Clustal W [13] implemented in BioEdit [14]. The abbreviations correspond to the following species common names: Hosa - Human; Patr - Common chimpanzee; Gogo - Western gorilla; Poab - Sumatran orangutan; Nole - Northern white-cheeked gibbon; Mamu - Rhesus monkey; Bota - Cow; Susc - Pig; Eqca - Horse; Mylu - Little brown myotis; Orcu - European rabbit; Rano - Brown rat; Crgr - Chinese hamster; Capo - Domestic Guinea pig; Soar - Common shrew. To access the species scientific names, the list of abbreviations should be consulted. Residues positions are numbered according to human SAMD9 protein. "?" represents undetermined codons; "." represents identity with the reference sequence of human SAMD9 protein.

Supplementary Figure S5 - Mammalian SAMD9L deduced protein sequences alignment.
SAMD9L deduced protein sequences from eighteen species were aligned with Clustal W [13] implemented in BioEdit [14]. The abbreviations correspond to the following species common names: Hosa - Human; Patr - Common chimpanzee; Gogo - Western gorilla; Poab - Sumatran orangutan; Nole - Northern white-cheeked gibbon; Caja - Common marmoset; Mamu - Rhesus monkey; Loaf - African bush elephant; Eqca - Horse; Calu - Domestic dog; Aime - Giant panda; Ereu - West European hedgehog; Orcu - European rabbit; Mumu - House mouse; Crgr - Chinese hamster; Rano - Brown rat; Capo - Domestic Guinea pig; Soar - Common shrew. To access the species scientific names, the list of abbreviations should be consulted. Residues positions are numbered according to human SAMD9L protein. "?" represents undetermined codons; "." represents identity with the reference sequence of human SAMD9L protein.

Supplementary Figure S6 - Mammalian *SAMD9* gene estimated Maximum Likelihood tree.
The phylogenetic tree of mammalian *SAMD9* gene alignment was estimated using the Maximum Likelihood method and the nucleotide substitution model TVM+G. The analyses were performed with 1,000,000 generations and 1,000 bootstrap searches. The bootstrap values are indicated on the branches. The abbreviations correspond to the following species common names: Hosa - Human; Patr - Common chimpanzee; Gogo - Western gorilla; Poab - Sumatran orangutan; Nole - Northern white-cheeked gibbon; Mamu - Rhesus monkey; Bota - Cow; Susc - Pig; Eqca - Horse; Mylu - Little brown myotis; Orcu - European rabbit; Rano - Brown rat; Crgr - Chinese hamster; Capo

- Domestic Guinea pig; Soar - Common shrew. To access the species scientific names, the list of abbreviations should be consulted.

Supplementary Figure S7 - Mammalian *SAMD9L* gene estimated Maximum Likelihood tree.
The phylogenetic tree of mammalian *SAMD9L* gene alignment was estimated using the Maximum Likelihood method and the nucleotide substitution model GTR+G. The analyses were performed with 1,000,000 generations and 1,000 bootstrap searches. The bootstrap values are indicated on the branches. The abbreviations correspond to the following species common names: Hosa - Human; Patr - Common chimpanzee; Gogo - Western gorilla; Poab - Sumatran orangutan; Nole - Northern white-cheeked gibbon; Caja - Common marmoset; Mamu - Rhesus monkey; Loaf - African bush elephant; Eqca - Horse; Calu - Domestic dog; Aime - Giant panda; Ereu - West European hedgehog; Orcu - European rabbit; Mumu - House mouse; Crgr - Chinese hamster; Rano - Brown rat; Capo - Domestic Guinea pig; Soar - Common shrew. To access the species scientific names, the list of abbreviations should be consulted.

Supplementary Table S2 - *SAMD9* and *SAMD9L* likelihood ratio test (LRT) for PARRIS analysis from HyPhy software.
Only *SAMD9L* was found to be under selection when using this specific method.

Supplementary Table S3 - Positively-selected codon positions in *SAMD9* and *SAMD9L* determined by six different Maximum Likelihood methods.
The six methods correspond to PAML M8, SLAC, FEL, REL, MEME and FUBAR. Codons positions are numbered according to human SAMD9 and SAMD9L proteins (Supplementary Figure S4 and Supplementary Figure S5).

Supplementary Table S4 - SAMD9 and SAMD9L amino acid properties under positive selection determined in TreeSAAP.
SAMD9 exhibited three and SAMD9L evidenced seven amino acid properties under positive selection.

9. References

1. Li CF, MacDonald JR, Wei RY, Ray J, Lau K, Kandel C, Koffman R, Bell S, Scherer SW, Alman BA: Human sterile alpha motif domain 9, a novel gene

identified as down-regulated in aggressive fibromatosis, is absent in the mouse. *BMC Genomics* 2007, 8:92.

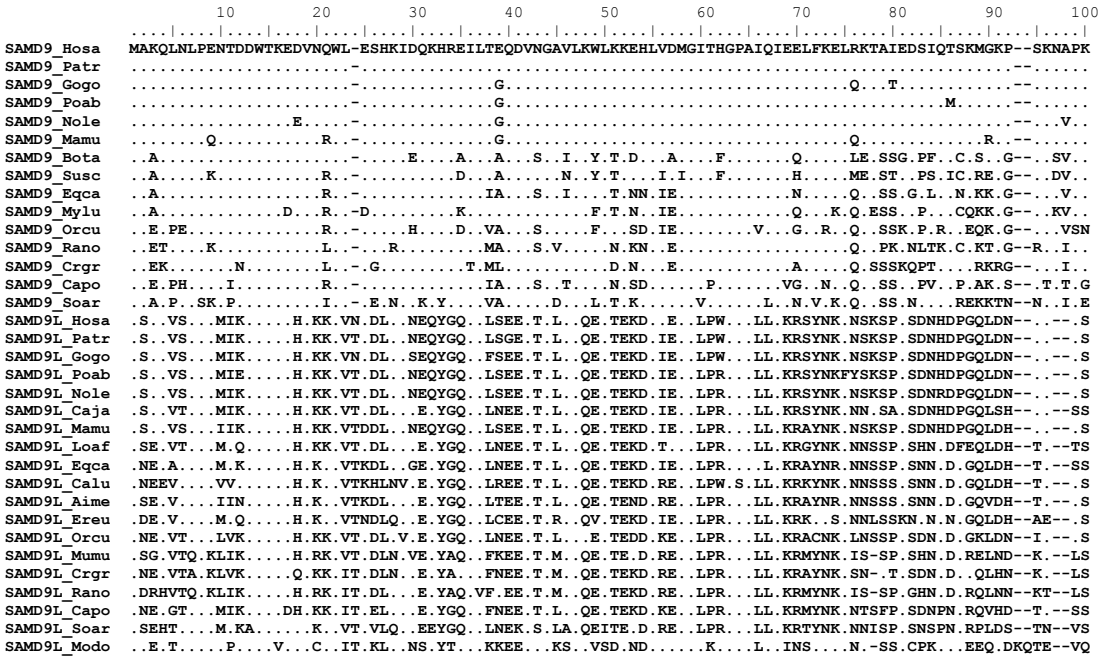
2. Asou H, Matsui H, Ozaki Y, Nagamachi A, Nakamura M, Aki D, Inaba T: Identification of a common microdeletion cluster in 7q21.3 subband among patients with myeloid leukemia and myelodysplastic syndrome. *Biochemical and Biophysical Research Communications* 2009, 383:245-251.
3. Topaz O, Indelman M, Chefetz I, Geiger D, Metzker A, Altschuler Y, Choder M, Bercovich D, Uitto J, Bergman R, Richard G, Sprecher E: A deleterious mutation in SAMD9 causes normophosphatemic familial tumoral calcinosis. *The American Journal of Human Genetics* 2006, 79:759-764.
4. Chefetz I, Ben Amitai D, Browning S, Skorecki K, Adir N, Thomas MG, Kogleck L, Topaz O, Indelman M, Uitto J, Richard G, Bradman N, Sprecher E: Normophosphatemic familial tumoral calcinosis is caused by deleterious mutations in SAMD9, encoding a TNF-alpha responsive protein. *Journal of Investigative Dermatology* 2008, 128:1423-1429.
5. Tanaka M, Shimbo T, Kikuchi Y, Matsuda M, Kaneda Y: Sterile alpha motif containing domain 9 is involved in death signaling of malignant glioma treated with inactivated Sendai virus particle (HVJ-E) or type I interferon. *International Journal of Cancer* 2010, 126:1982-1991.
6. Hershkovitz D, Gross Y, Nahum S, Yehezkel S, Sarig O, Uitto J, Sprecher E: Functional characterization of SAMD9, a protein deficient in normophosphatemic familial tumoral calcinosis. *Journal of Investigative Dermatology* 2011, 131:662-669.
7. Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, Rice CM: A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* 2011, 472:481-485.
8. Liu J, Wennier S, Zhang L, McFadden G: M062 is a host range factor essential for myxoma virus pathogenesis and functions as an antagonist of host SAMD9 in human cells. *Journal of Virology* 2011, 85:3270-3282.
9. Zhang LK, Chai F, Li HY, Xiao G, Guo L: Identification of host proteins involved in Japanese encephalitis virus infection by quantitative proteomics analysis. *Journal of Proteome Research* 2013, 12:2666-2678.
10. Pappas DJ, Coppola G, Gabatto PA, Gao F, Geschwind DH, Oksenberg JR, Baranzini SE: Longitudinal system-based analysis of transcriptional responses to type I interferons. *Physiological Genomics* 2009, 38:362-371.

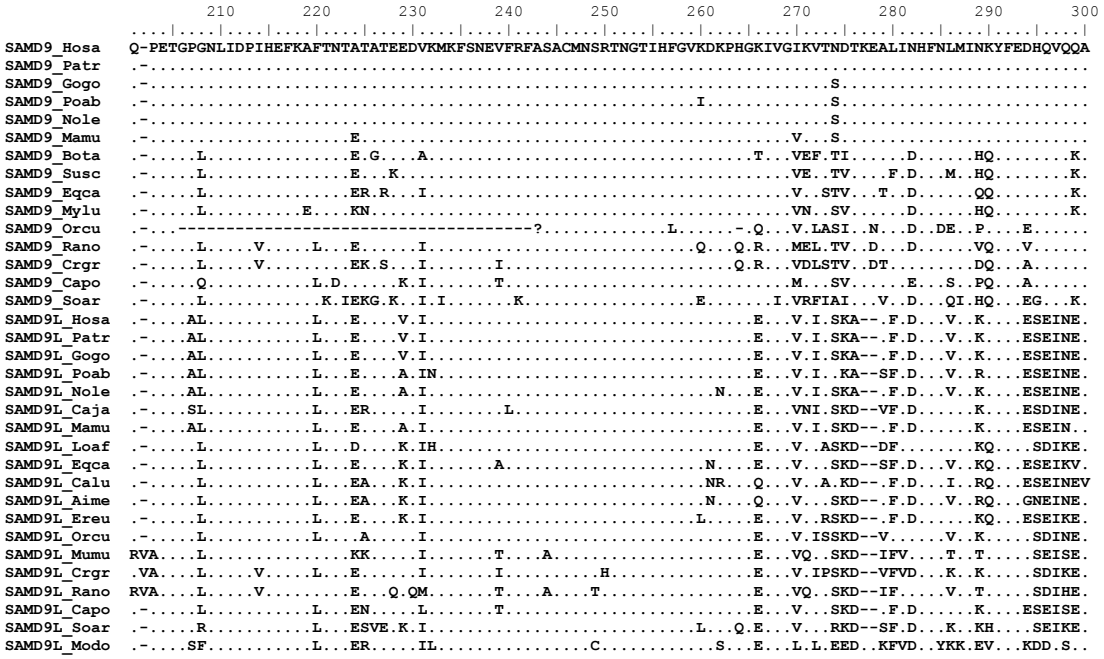
11. Jiang Q, Quaynor B, Sun A, Li Q, Matsui H, Honda H, Inaba T, Sprecher E, Uitto J: The Samd9L gene: transcriptional regulation and tissue-specific expression in mouse development. *Journal of Investigative Dermatology* 2011, 131:1428-1434.
12. Critchley-Thorne RJ, Yan N, Nacu S, Weber J, Holmes SP, Lee PP: Down-regulation of the interferon signaling pathway in T lymphocytes from patients with metastatic melanoma. *PLoS Medicine* 2007, 4:e176.
13. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Symposium Series* 1994, 22:4673-4680.
14. Hall T: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 1999, 41:95-98.
15. Xia X, Xie Z: DAMBE: software package for data analysis in molecular biology and evolution. *Journal of Heredity* 2001, 92:371-373.
16. Posada D, Crandall KA: The effect of recombination on the accuracy of phylogeny estimation. *Journal of Molecular Evolution* 2002, 54:396-402.
17. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD: GARD: a genetic algorithm for recombination detection. *Bioinformatics* 2006, 22:3096-3098.
18. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD: Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular Biology and Evolution* 2006, 23:1891-1901.
19. Pond SL, Frost SD: Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 2005, 21:2531-2533.
20. Posada D: jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 2008, 25:1253-1256.
21. Zwickl DJ: Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. *PhD Thesis.* University of Texas, 2006.
22. Miyata T, Yasunaga T: Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *Journal of Molecular Evolution* 1980, 16:23-36.
23. Nei M, Gojobori T: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* 1986, 3:418-426.

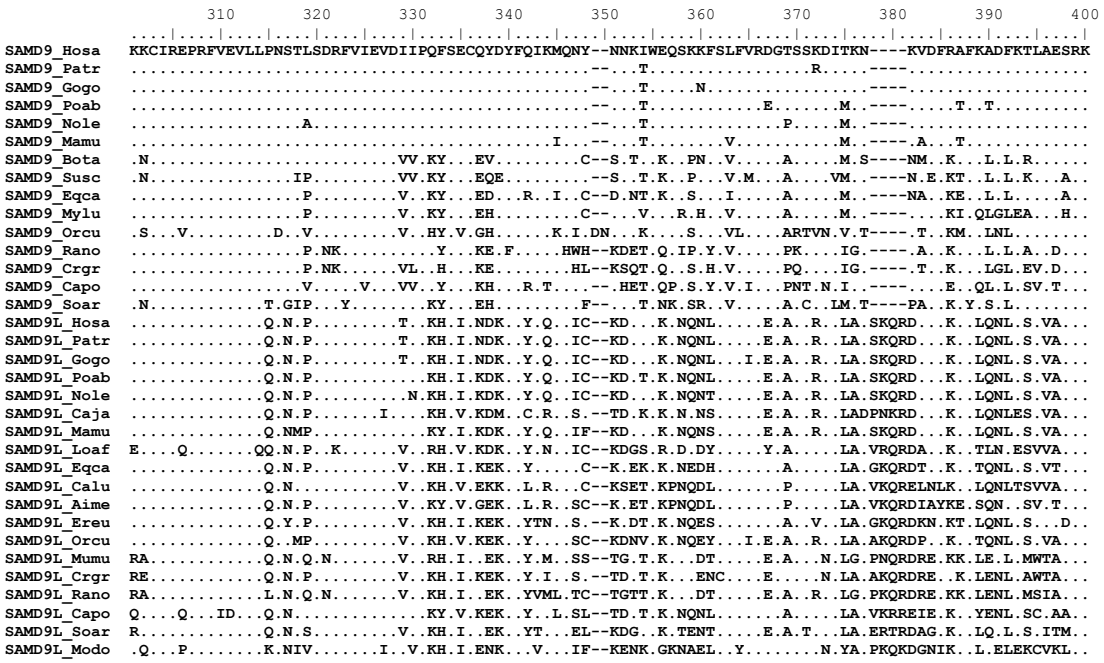
24. Crandall KA, Kelsey CR, Imamichi H, Lane HC, Salzman NP: Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Molecular Biology and Evolution* 1999, 16:372-382.
25. Nielsen R, Yang Z: Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 1998, 148:929-936.
26. Yang Z, Nielsen R, Goldman N, Pedersen AM: Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 2000, 155:431-449.
27. Wlasiuk G, Nachman MW: Adaptation and constraint at Toll-like receptors in primates. *Molecular Biology and Evolution* 2010, 27:2172-2186.
28. Areal H, Abrantes J, Esteves PJ: Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evolutionary Biology* 2011, 11:368.
29. Yang Z: PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* 1997, 13:555-556.
30. Yang Z: PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 2007, 24:1586-1591.
31. Yang Z, Wong WS, Nielsen R: Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* 2005, 22:1107-1118.
32. Kosakovsky Pond SL, Frost SD: Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* 2005, 22:1208-1222.
33. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL: Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics* 2012, 8:e1002764.
34. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K: FUBAR: A Fast, Unconstrained Bayesian AppRoximation for inferring selection. *Molecular Biology and Evolution* 2013, 30:1196-1205.
35. Scheffler K, Martin DP, Seoighe C: Robust inference of positive selection from recombining coding sequences. *Bioinformatics* 2006, 22:2493-2499.
36. Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA: TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* 2003, 19:671-672.
37. Song S, Liu L, Edwards SV, Wu S: Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy of Sciences of the United States of America* 2012, 109:14942-14947.

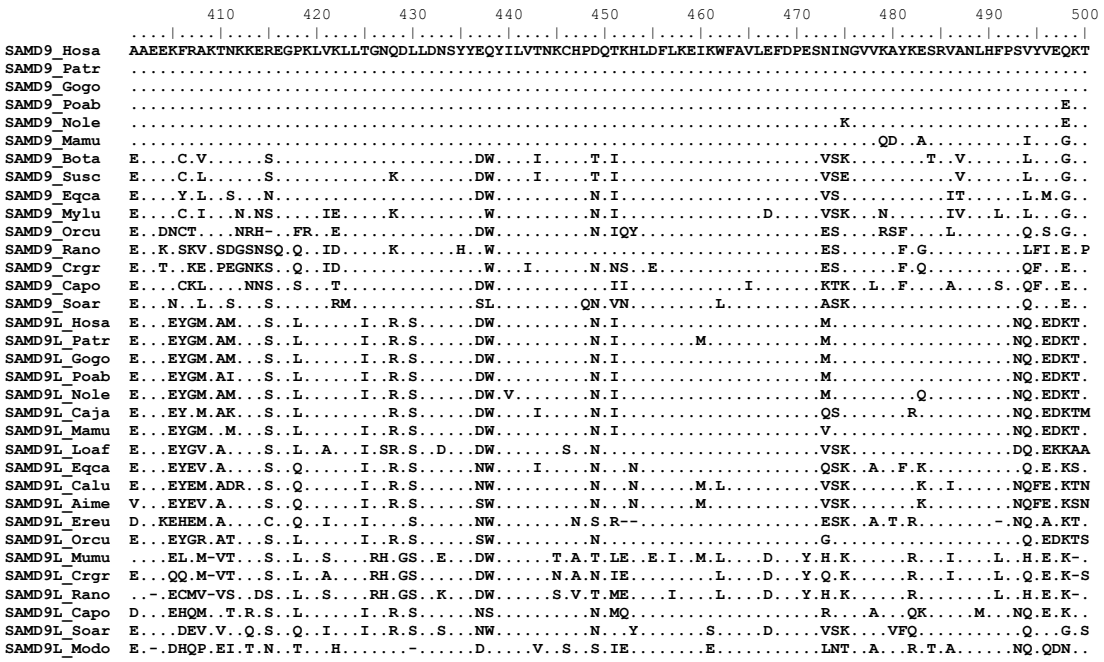
38. Zhang J: Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *Journal of Molecular Evolution* 2000, 50:56-68.
39. Adkins RM, Gelke EL, Rowe D, Honeycutt RL: Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes. *Molecular Biology and Evolution* 2001, 18:777-791.
40. Churakov G, Sadasivuni MK, Rosenbloom KR, Huchon D, Brosius J, Schmitz J: Rodent evolution: back to the root. *Molecular Biology and Evolution* 2010, 27:1315-1326.
41. Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A: The delayed rise of present-day mammals. *Nature* 2007, 446:507-512.
42. Hughes AL, Ota T, Nei M: Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Molecular Biology and Evolution* 1990, 7:515-524.
43. Zhang J, Webb DM: Rapid evolution of primate antiviral enzyme APOBEC3G. *Human Molecular Genetics* 2004, 13:1785-1791.
44. Ponting CP: SAM: a novel motif in yeast sterile and *Drosophila* polyhomeotic proteins. *Protein Science* 1995, 4:1928-1930.
45. Qiao F, Bowie JU: The many faces of SAM. *Science Signal Transduction Knowledge Environment* 2005, 2005:re7.
46. Meruelo AD, Bowie JU: Identifying polymer-forming SAM domains. *Proteins* 2009, 74:1-5.
47. Knight MJ, Leettola C, Gingery M, Li H, Bowie JU: A human sterile alpha motif domain polymerizome. *Protein Science* 2011, 20:1697-1706.

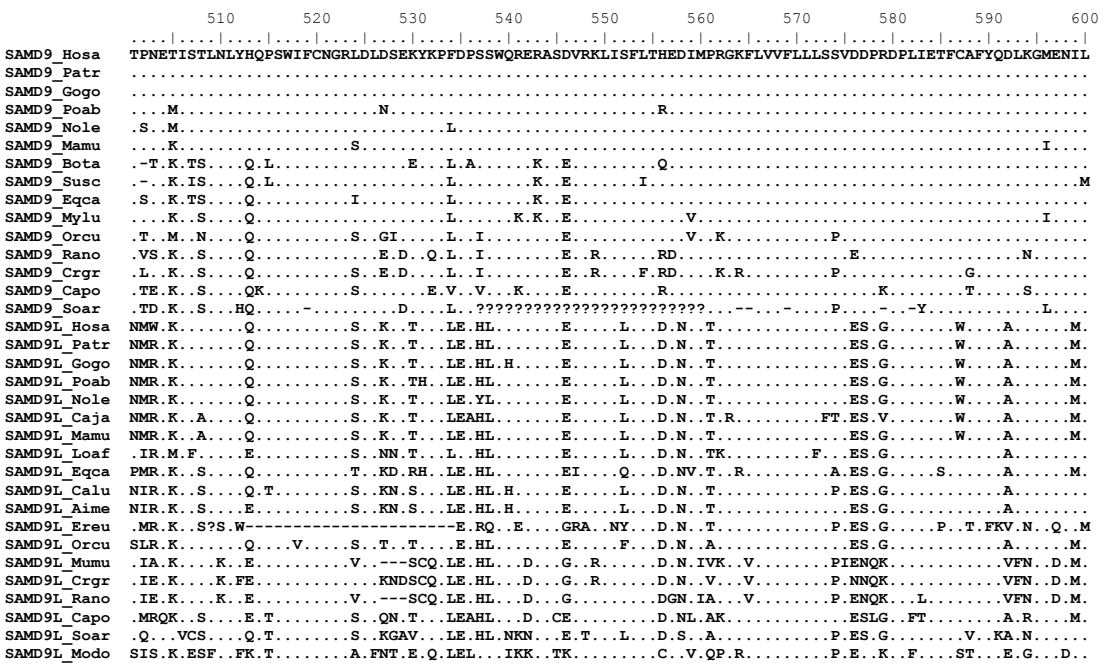
Supplementary Figure S1




210 220 230 240 250 260 270 280 290 300
SAMD9_Hosa Q-PETGPGNLIDPIHEFKAFTNTATATEEDVKMKFSNEVFRFASACMNSRTNGTIHFGVKDKPHGKIVGIKVTNDTKEALINHFNLMINKYFEDHQVQQA
SAMD9_Patr .-
SAMD9_Gogo .-
SAMD9_Poab .-
SAMD9_Nole .-
SAMD9_Mamu .-
SAMD9_Bota .-
SAMD9_Susc .-
SAMD9_Eqca .-
SAMD9_Mylu .-
SAMD9_Orcu .-
SAMD9_Rano .-
SAMD9_Crgr .-
SAMD9_Capo .-
SAMD9_Soar .-
SAMD9L_Hosa .-
SAMD9L_Patr .-
SAMD9L_Gogo .-
SAMD9L_Poab .-
SAMD9L_Nole .-
SAMD9L_Caja .-
SAMD9L_Mamu .-
SAMD9L_Loaf .-
SAMD9L_Eqca .-
SAMD9L_Calu .-
SAMD9L_Aime .-
SAMD9L_Ereu .-
SAMD9L_Orcu .-
SAMD9L_Mumu RVA
SAMD9L_Crgr .VA
SAMD9L_Rano RVA
SAMD9L_Capo .-
SAMD9L_Soar .-
SAMD9L_Modo .-

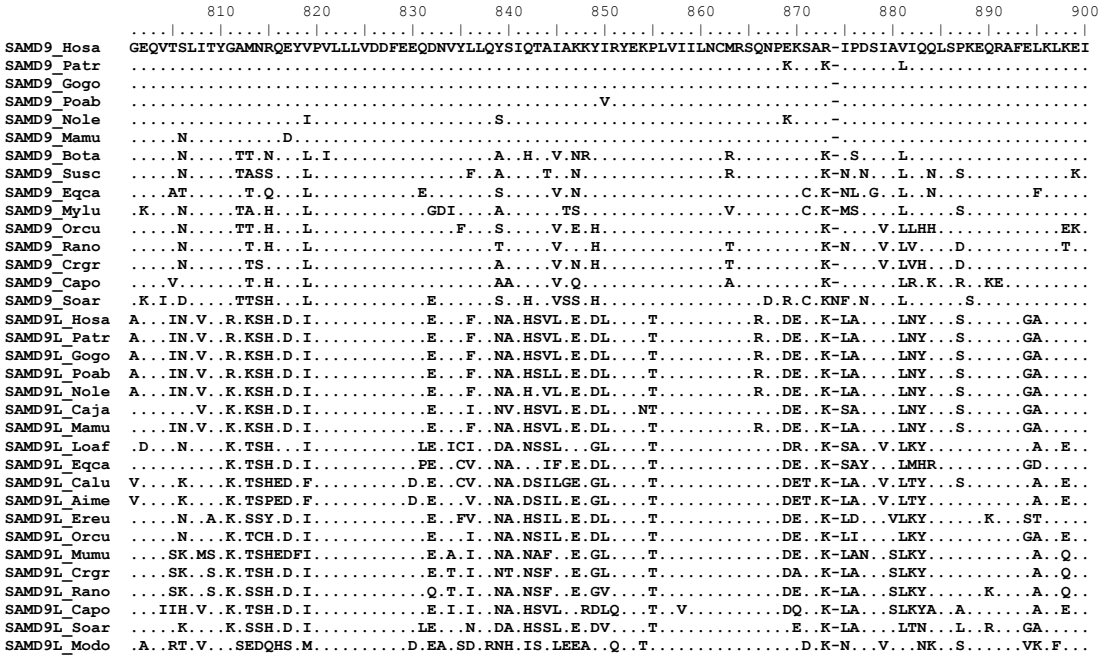

310 320 330 340 350 360 370 380 390 400
SAMD9_Hosa KKCIREPRFVEVLLPNSTLSDRFVIEVDIIPQFSECQYDYFQIKMQNY--NNKIWEQSKKFSLFVRDGTSSKDITKN----KVDFRAFKADFKTLAESRK
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
SAMD9L_Hosa
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
SAMD9L_Modo

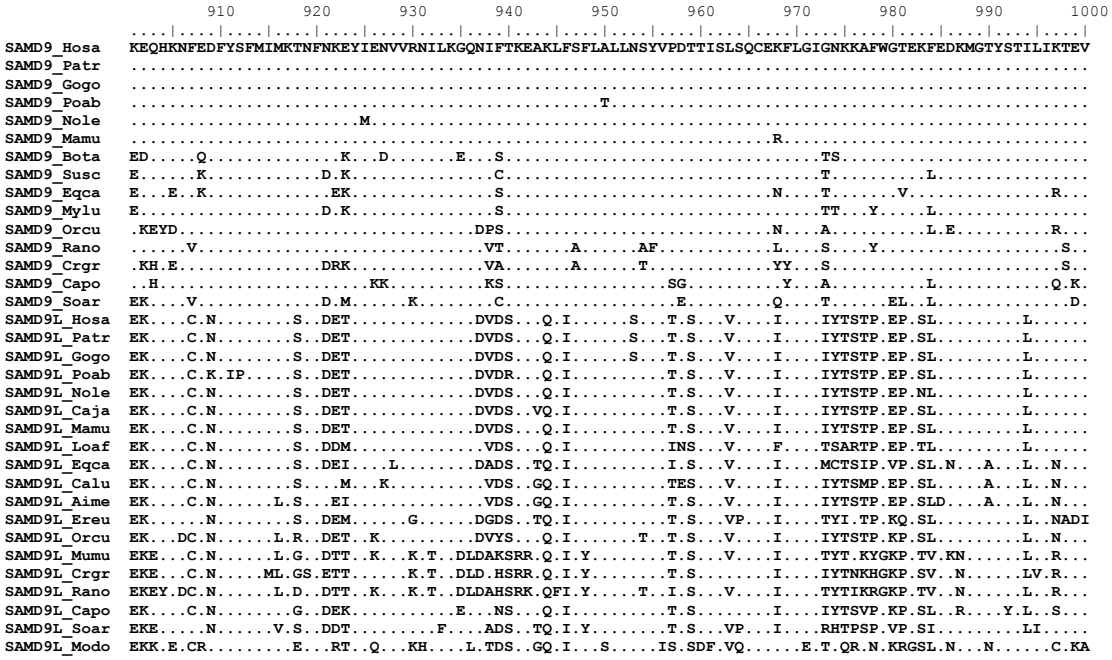

410 420 430 440 450 460 470 480 490 500
SAMD9_Hosa AAEEKFRAKTNKKEREGPKLVKLLTGNQDLLDNSYYEQYILVTNKCHPDQTKHLDFLKEIKWFAVLEFDPESNINGVVKAYKESRVANLHFPSVYVEQKT
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
SAMD9L_Hosa
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
SAMD9L_Modo

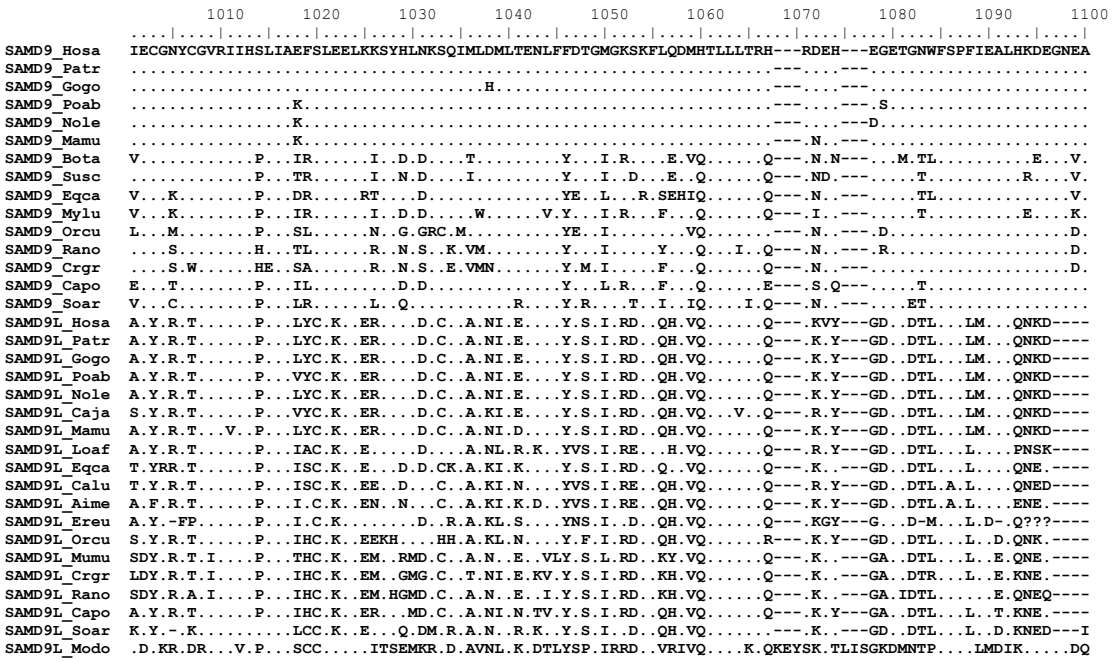

510 520 530 540 550 560 570 580 590 600
SAMD9_Hosa TPNETISTLNLYHQPSWIFCNGRLDLDSEKYKPFDPSSWQRERASDVRKLISFLTHEDIMPRGKFLVVFLLLSSVDDPRDPLIETFCAFYQDLKGMENIL
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
SAMD9L_Hosa
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
SAMD9L_Modo

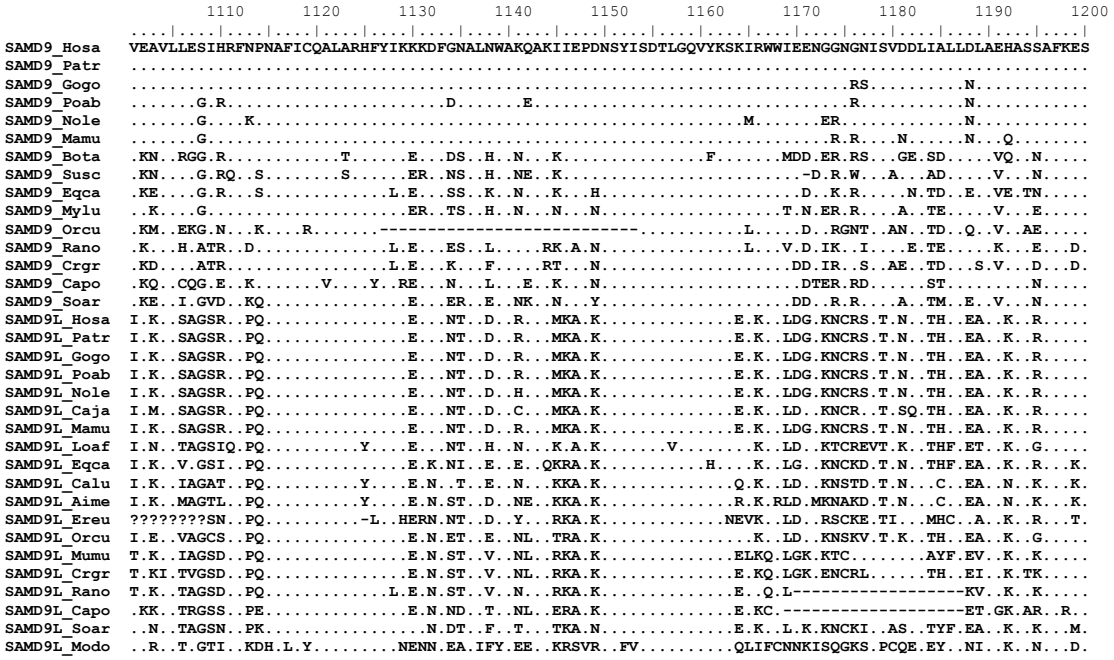
```
                610       620       630       640       650       660       670       680       690       700
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SAMD9_Hosa CICVHPHIFQGWKDLLEARLIKHQDEISSQCISALSLEEINGTILKLKSVTQSSKRLLPSIGLSTVLL-KKEEDIMTALEIICENECEGTLLEKDKNKFL
SAMD9_Patr ..........................................................-..........................................
SAMD9_Gogo ..........................................................-..........................................
SAMD9_Poab ..........................................................-..........................................
SAMD9_Nole ....Q.....................................................-..........................................
SAMD9_Mamu ....Q.....................................................-........S..................................
SAMD9_Bota ....DSR.C...........TTQ...L.N..V.S.....................F.....S.....-...........L......D.I.....K.L.
SAMD9_Susc ...A.SR.C...........A.QE..L.........................F...V.S.....-.R..........L.............EK...
SAMD9_Eqca .....S..C.R.........T.Q...L.N.....................L..F.....S.....-........S...L.............K...
SAMD9_Mylu .V...S..C.R.........T.-...L......S.................K..Q.F.....S.....-EM.........L.....K.....N......
SAMD9_Orcu ......Q.C...........T.E...L.N......N.Q.............F.....S.....-..Q..T......L......D...KD..K...
SAMD9_Rano .........Y.V......V..SSQ...L...S.FS.N.A..........L......F...V.S.....-.R..E.......L.............RE.LF
SAMD9_Crgr ..S.FSS.Y.V.........AGQ...L.NRS....N.A..........L......F.....ST....-....-.......L...........N..K.LF
SAMD9_Capo ....Q.N.C...........TNE...L.DK.....N..Q......N...I....E......CS.....-.....-......L.V........H...T.LQ
SAMD9_Soar V.-LNSQLC.I.........KIQ...LAN......N.................V.K......S.....-............L......D.D.........
SAMD9L_Hosa ..S.NS..Y.R.....QT.MKM-E..LTNHS..T.NI.LV.S.........R..R.F..AR.S.S.I.E..K..VL.....L.....TE.DI....S...
SAMD9L_Patr ..S.NS..Y.R.....QT.MKM-E..LTNHS..T.NI.LV.S.........R..R.F..AR.S.S.I.E..K..VL.....L.....TE.DI....S...
SAMD9L_Gogo ..S.NS..Y.R.....QT.MKM-E..LTNHS..T.NI.LV.S.........R..R.F..AR.S.S.I.E..K..VL.....L.....TE.DI....S...
SAMD9L_Poab ..S.NS..Y.R.....QT.MKM-E..LTNHS..T.NI.LV.S.............R.F..AH.S.S.I.E..K..VL.....L.....KD.DI...ES...
SAMD9L_Nole ..S.NS..Y.R.....QT.MKM-E..LTNHS..T.NI.LV.S...........PR.F..AH.S.S.I.E..K..VL.....L.....RD.DI...ES...
SAMD9L_Caja ..S.SS..Y.R.....QT.MNI-E..LTNHS..T.NI.LV.SS..........R.F..AH.S.S.I.K..K..AL.....L.....RD.DI...ES...
SAMD9L_Mamu ..S.NS..Y.R.....QT.MKM-E..LTNHS..T.NI.LV.S.............R.F..AR.S.S.I.E..K..VL.....L.....RD.DI...ES...
SAMD9L_Loaf ....KSQ.Y.R.....RT..TI-A..LTNHS..T.N..Q..S........I...R.F...R.S.S.I.E.....LF.....L.....KD.DI....S...
SAMD9L_Eqca ....NSQ.Y.R.....QT..TV-A..LANHS..T.N..L..S......P..E..R.F...S...S.I.E......L.....L.....RD.DI...EF..Q
SAMD9L_Calu ....NS..Y.R.....QT..TI-A..LTNHS..T.N..L..S......P....LR.F...H.F.SII.E......L.....L.....KD.DI....S..Q
SAMD9L_Aime ....NSQ.Y.R.....QT..TA-A..LTNHS..T.N..L..S......P.....R.F...H.FAS.I.E.....TL.....L.....KD.DI...ES..Q
SAMD9L_Ereu ....T.Q.Y.R.....QT.FVM-S..LTNHS..T....LL.S..............F...R.S.S.I.E.....FL.....L.....GD.DI...EC..E
SAMD9L_Orcu ....NSQ.Y.R.....KT..TV-AH.L.NHS..T.N..LV.S.........R..R.F...R.S.S.I.E..K..TL.....L.....RD.DI...ES...
SAMD9L_Mumu ....NSA.Y.Q.S...QV..EI-K.DLAKHS..T.NI.LV.N........I...R.F...C.S.S.I.E.MD....S....L.....KD.DI...ESQ..
SAMD9L_Crgr ....N...Y.R.S...QV..EI-KG.LAEHS..T.NIQLV.N........I...R.F...C.S.S.I.E.I...TL.....L.....RD.DI..NESE..
SAMD9L_Rano ....N.S.Y.Q.S...QV..EI-K.DLAKHS..T.NI.LV.S........I...R.F...C.S.S.I.E.MD....S....L.....RD.DI...ESQ..
SAMD9L_Capo ....NSK.Y.R.....QT..TI-G..L.KHSV.T.NI.L..S.......L....RKF...Y.S.S.I.E.....LL...Q.LS....KE.DI....S...
SAMD9L_Soar ....NLD.YKQ.....K....E-ES.LEDHT..T.N..FL.S.........R....F...K.S.S.T.E.....TF.T...L.....KD.DI..NAS..Q
SAMD9L_Modo S.S.DQQ.GSQ..E...GKFSS-I.-LCNR..TT....QL........PQ.E.PI.F...N.S.-II.Q..D..S.S....L.....KD.AI...E-..T
```

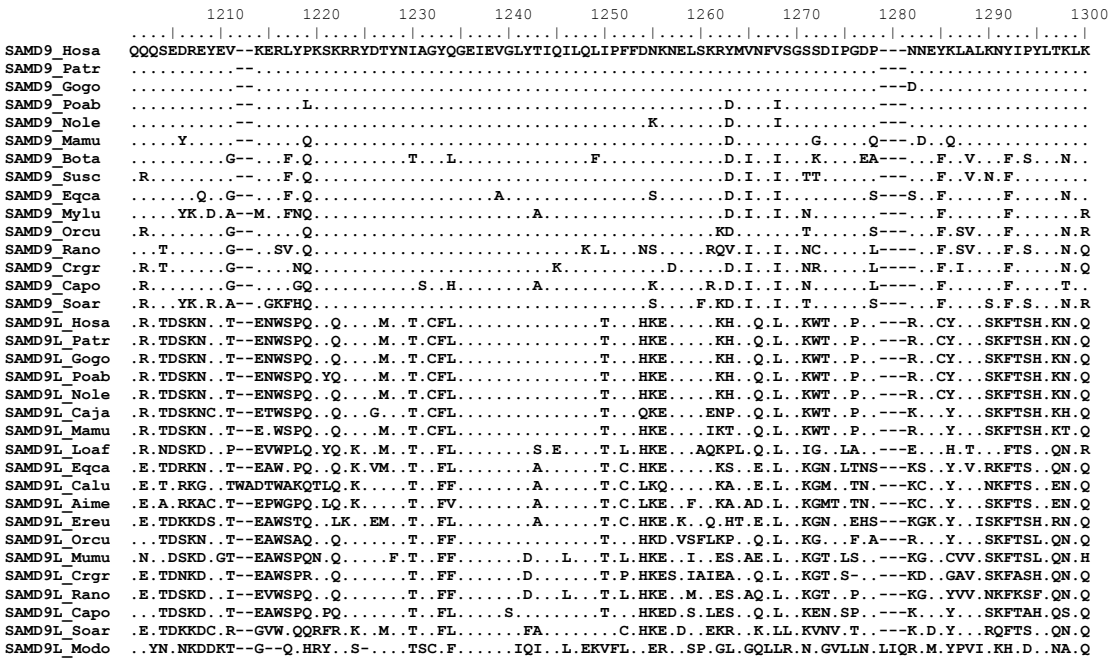
```
                710       720       730       740       750       760       770       780       790       800
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SAMD9_Hosa EFKASKEEDFYRGGKVSWWNFYFSSESYSSPFVKRDKYERLEAMIQNCADSSKPTSTKIIHLYHHPGCGGTTLAMHILWELRKKFRCAVLKNKTVDFSEI
SAMD9_Patr ..........................................................................................
SAMD9_Gogo ..........................................................................................A.....
SAMD9_Poab ..........................................................................................
SAMD9_Nole ..........................................................................................A.....
SAMD9_Mamu ..........................................................C.......................D.....................
SAMD9_Bota ....L...........T.........N............K..E...RG.....S.CV.....................................
SAMD9_Susc ....L.....................N........ER..K..T...E..A.....CV...................V...............M...C..
SAMD9_Eqca D..T.....................K...L.........I..E....W.......CV.....................................M.....
SAMD9_Mylu ....L.....................N............K..R..H.......S.CS.....................................
SAMD9_Orcu .......KA.................NH..A.................W......ICA.....................................
SAMD9_Rano ...T...K.................N...S.......VK..KK..EW....Q.VCA.....................................E.....
SAMD9_Crgr ...TA..Q.................N...S........K..KK..EW....Q.VCA.....................D.............E.....
SAMD9_Capo ...T.....................-..Y...S........K.KE....Y..P..LMCV.....................................
SAMD9_Soar A..TL....................KN...A........K.KD....W-....S.CV...Q.....................N.....................
SAMD9L_Hosa ...K....H.................N...D.....S..K.KDL.HCW.E.P..IFA...N...............V..D.K.N..........T..A..
SAMD9L_Patr ...K....H.................N...D.....S..K.KDL.HCW.E.P..IFA...N...............V..D.K.N..........T..A..
SAMD9L_Gogo ...K....H.................N...D.....S..K.KDLVHCW.E.P..IFA...N...............V..D.K.N..........T..A..
SAMD9L_Poab ...K....H.....C...........N...D.....S..K.KDL.RCW.E.P..IFA...N...............V..D.K.N..........T..A..
SAMD9L_Nole ...K....H.................N...D.....S..K.KDL.HCW.E.P..IFA...N...............V..D.K.N..........T..A..
SAMD9L_Caja ..RK....H.................N.P.D.....S..K.KDL..CL.E.P..IFA...N...............V..D.KE...........T..V..
SAMD9L_Mamu ...K....H.................N...D.....S..K.KDL..CW.E.P..IFA...N...............V..D.K.N..........TG.A..
SAMD9L_Loaf ..RK....H.....R...........N...A.....S..H.KDL..CW.E.P..VFA...N...............V..N.K.N.........AT..T..
SAMD9L_Eqca ...K...KY.................N...A.....S..K.KDL.LCR.Q.P..PFA..VN..............NV..D.K.N.........AT..A..
SAMD9L_Calu ...KL...H.................N...A.....S..K.KVL.KCW.E.P..EFA...N...............NV..D.K.N..........T..E..
SAMD9L_Aime ...T....H.................N...A.....S..K.KDL..CW...P..VFA...N...............NV..D.K.S..........T..G..
SAMD9L_Ereu ..RK....H.................N...-.I...R..E.RSL.EDW.E.P..VFA...N...............NV..DF............AT..A..
SAMD9L_Orcu ...K....H.................N...A.....S.KE.KNL..CW.E.PR.VFA...N...............V..D.K.N..........AT..A..
SAMD9L_Mumu ...K.R..H.....R...........N...A.....SF.E.TTL..Q....P..VFV.V.N...............V..D.KQ...........AT..V..
SAMD9L_Crgr K..KLR..H.....RA..........N...A.....G..E.TTL..Q....PE.VFA.V.N...............V..D.K...........AT..L..
SAMD9L_Rano ...KLR..H.....R...........N...A.....NF.E.TTL..Q....P..VFA.V.N...............V..D.KQ...........AT..I..
SAMD9L_Capo ..RK.R..H.................N...A.....M.KN.KDL.ER..ECP..EFA...N.................D.KQ.Y.........AI..E.V
SAMD9L_Soar ..IK.N..Q....S...........NHT.A.....R..N.KKL.EDW.E.P..LFA...N...............NV..............I.A..A..
SAMD9L_Modo ..QMN..KH..L...........Y..NHT.D.IR..S..I..K...SL.N.PRESGAEV.N.....................K.N.......S..FASED.
```

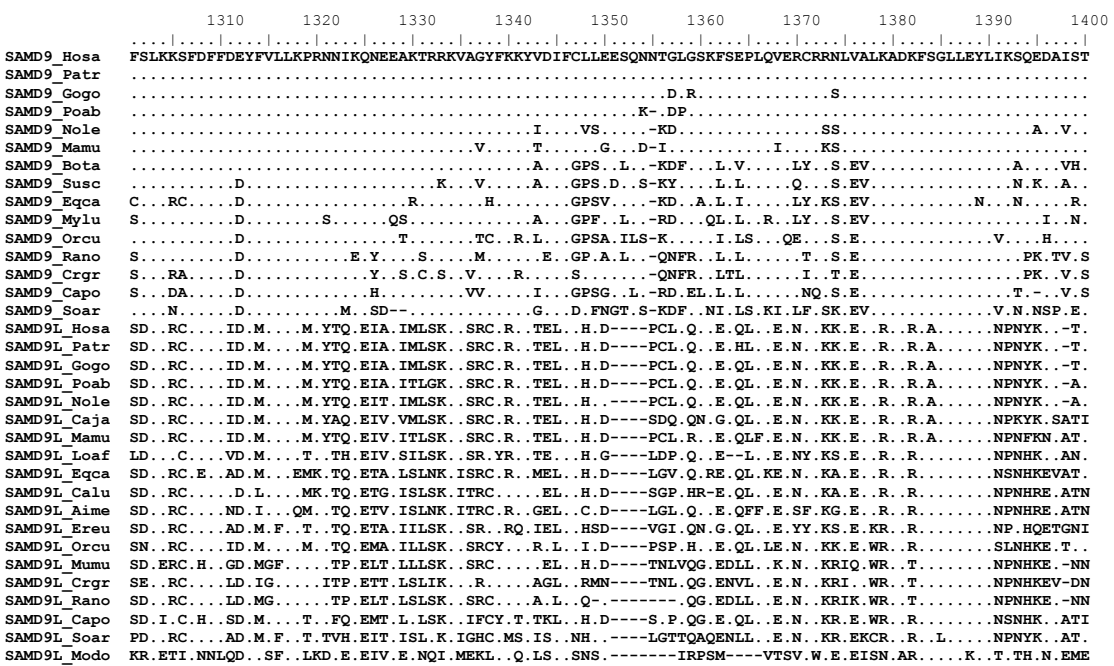

810 820 830 840 850 860 870 880 890 900
SAMD9_Hosa GEQVTSLITYGAMNRQEYVPVLLLVDDFEEQDNVYLLQYSIQTAIAKKYIRYEKPLVIILNCMRSQNPEKSAR-IPDSIAVIQQLSPKEQRAFELKLKEI
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
SAMD9L_Hosa
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
SAMD9L_Modo

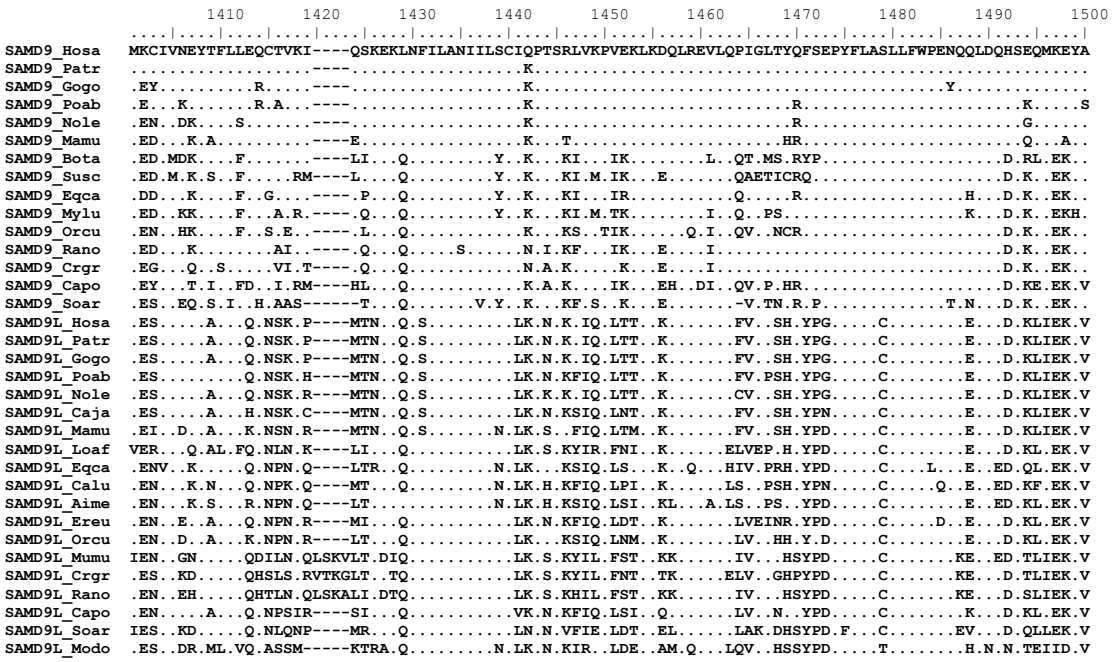

910 920 930 940 950 960 970 980 990 1000
SAMD9_Hosa KEQHKNFEDFYSFMIMKTNFNKEYIENVVRNILKGQNIFTKEAKLFSFLALLNSYVPDTTISLSQCEKFLGIGNKKAFWGTEKFEDKMGTYSTILIKTEV
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
SAMD9L_Hosa
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
SAMD9L_Modo

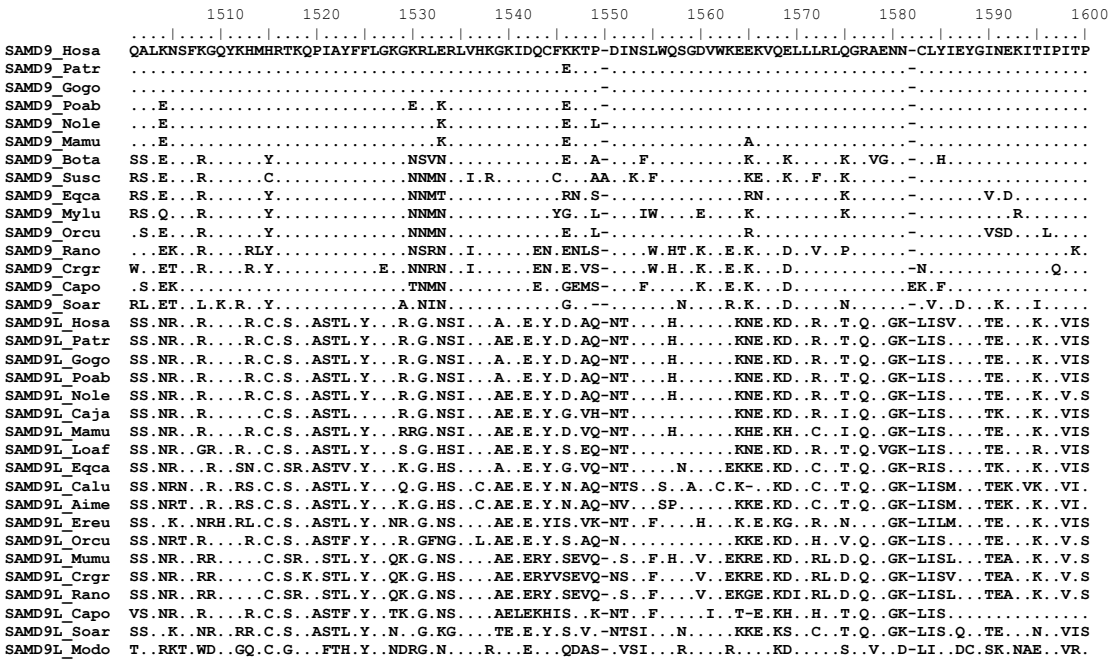

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
SAMD9_Hosa IECGNYCGVRIIHSLIAEFSLEELKKSYHLNKSQIMLDMLTENLFFDTGMGKSKFLQDMHTLLLTRH---RDEH---EGETGNWFSPFIEALHKDEGNEA
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
SAMD9L_Hosa
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
SAMD9L_Modo


1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
SAMD9_Hosa VEAVLLESIHRFNPNAFICQALARHFYIKKKDFGNALNWAKQAKIIEPDNSYISDTLGQVYKSKIRWWIEENGGNGNISVDDLIALLDLAEHASSAFKES
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
SAMD9L_Hosa
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
SAMD9L_Modo


1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
SAMD9_Hosa QQQSEDREYEV--KERLYPKSKRRYDTYNIAGYQGEIEVGLYTIQILQLIPFFDNKNELSKRYMVNFVSGSSDIPGDP---NNEYKLALKNYIPYLTKLK
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
SAMD9L_Hosa
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Capo
SAMD9L_Soar
SAMD9L_Modo


1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
SAMD9_Hosa FSLKKSFDFFDEYFVLLKPRNNIKQNEEAKTRRKVAGYFKKYVDIFCLLEESQNNTGLGSKFSEPLQVERCRRNLVALKADKFSGLLEYLIKSQEDAIST
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
SAMD9L_Hosa
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
SAMD9L_Modo


1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
SAMD9_Hosa MKCIVNEYTFLLEQCTVKI----QSKEKLNFILANIILSCIQPTSRLVKPVEKLKDQLREVLQPIGLTYQFSEPYFLASLLFWPENQQLDQHSEQMKEYA
SAMD9_Patr .....----.....K.....
SAMD9_Gogo .EY.....R....----.....K.....Y.....
SAMD9_Poab .E...K.....R.A..----.....K.....R.....K.....S
SAMD9_Nole .EN..DK....S.....----.....K.....R.....G.....
SAMD9_Mamu .ED...K.A.....----E.....K...T.....HR.....Q...A..
SAMD9_Bota .ED.MDK....F.....----LI...Q.....Y..K...KI...IK.....L..QT.MS.RYP.....D.RL.EK..
SAMD9_Susc .ED.M.K.S..F.....RM----L....Q.....Y..K...KI.M.IK...E.....QAETICRQ.....D.K..EK..
SAMD9_Eqca .DD...K....F..G....----.P...Q.....Y..K...KI...IR.....Q.....R.....H...D.K..EK..
SAMD9_Mylu .ED..KK....F...A.R.----.Q...Q.....Y..K...KI.M.TK.....I..Q..PS.....K...D.K..EKH.
SAMD9_Orcu .EN..HK....F..S.E..----.L...Q.....K...KS..TIK.....Q.I..QV..NCR.....D.K..EK..
SAMD9_Rano .ED...K.....AI..----.Q...Q.....S.....N.I.KF...IK...E....I.....D.K..EK..
SAMD9_Crgr .EG...Q..S.....VI.T----.Q...Q.....N.A.K.....K...E....I.....D.K..EK..
SAMD9_Capo .EY...T.I..FD..I.RM----HL...Q.....K.A.K....IK...EH..DI..QV.P.HR.....D.KE.EK.V
SAMD9_Soar .ES..EQ.S.I..H.AAS------T...Q.....V.Y..K...KF.S..K...E.....-V.TN.R.P.....T.N...D.K..EK..
SAMD9L_Hosa .ES.....A...Q.NSK.P----MTN..Q.S.....LK.N.K.IQ.LTT..K.....FV..SH.YPG.....C.....E...D.KLIEK.V
SAMD9L_Patr .ES.....A...Q.NSK.P----MTN..Q.S.....LK.N.K.IQ.LTT..K.....FV..SH.YPG.....C.....E...D.KLIEK.V
SAMD9L_Gogo .ES.....A...Q.NSK.P----MTN..Q.S.....LK.N.K.IQ.LTT..K.....FV..SH.YPG.....C.....E...D.KLIEK.V
SAMD9L_Poab .ES.....Q.NSK.H----MTN..Q.S.....LK.N.KFIQ.LTT..K.....FV.PSH.YPG.....C.....E...D.KLIEK.V
SAMD9L_Nole .ES.....A...Q.NSK.R----MTN..Q.S.....LK.K.K.IQ.LTT..K.....CV..SH.YPG.....C.....E...D.KLIEK.V
SAMD9L_Caja .ES.....A...H.NSK.C----MTN..Q.S.....LK.N.KSIQ.LNT..K.....FV..SH.YPN.....C.....E...D.KLIEK.V
SAMD9L_Mamu .EI..D..A...K.NSN.R----MTN..Q.S.....N.LK.S..FIQ.LTM..K.....FV..SH.YPD.....C.....E...D.KLIEK.V
SAMD9L_Loaf VER...Q.AL.FQ.NLN.K----LI...Q.....LK.S.KYIR.FNI..K.....ELVEP.H.YPD.....C.....E...D.KL.EK.V
SAMD9L_Eqca .ENV..K.....Q.NPN.Q----LTR..Q.....N.LK...KSIQ.LS...K..Q...HIV.PRH.YPD.....C....L...E..ED.QL.EK.V
SAMD9L_Calu .EN...K.N...Q.NPK.Q----MT...Q.....N.LK.H.KFIQ.LPI..K.....LS..PSH.YPN.....C.....Q..E..ED.KF.EK.V
SAMD9L_Aime .EN...K.S...R.NPN.Q----LT.....N.LK.H.KSIQ.LSI..KL...A.LS..PS..YPD.....C.....E..ED.KL.EK.V
SAMD9L_Ereu .EN..E..A...Q.NPN.R----MI...Q.....LK.N.KFIQ.LDT..K.....LVEINR.YPD.....C.....D..E...D.KL.EK.V
SAMD9L_Orcu .EN..D..A...K.NPN.R----LT...Q.....LK...KSIQ.LNM..K.....LV..HH.Y.D.....C.....E...D.KL.EK.V
SAMD9L_Mumu IEN..GN.....QDILN.QLSKVLT.DIQ.....LK.S.KYIL.FST..KK.....IV...HSYPD.....C.....KE..ED.TLIEK.V
SAMD9L_Crgr .ES..KD.....QHSLS.RVTKGLT..TQ.....LK.S.KYIL.FNT..TK.....ELV..GHPYPD.....C.....KE...D.TLIEK.V
SAMD9L_Rano .EN..EH.....QHTLN.QLSKALI.DTQ.....LK.S.KHIL.FST..KK.....IV...HSYPD.....C.....KE...D.SLIEK.V
SAMD9L_Capo .EN.....A...Q.NPSIR----SI...Q.....VK.N.KFIQ.LSI..Q.....LV..N..YPD.....C.....K...D.KL.EK.V
SAMD9L_Soar IES..KD.....Q.NLQNP----MR...Q.....LN.N.VFIE.LDT..EL.....LAK.DHSYPD.F...C.....EV...D.QLLEK.V
SAMD9L_Modo .ES..DR.ML.VQ.ASSM-----KTRA.Q.....N.LK.N.KIR..LDE..AM.Q...LQV..HSSYPD.....T.....H.N.N.TEIID.V


1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
SAMD9_Hosa QALKNSFKGQYKHMHRTKQPIAYFFLGKGKRLERLVHKGKIDQCFKKTP-DINSLWQSGDVWKEEKVQELLLRLQGRAENN-CLYIEYGINEKITIPITP
SAMD9_Patr .....E...-.....-.....
SAMD9_Gogo .....-.....-.....
SAMD9_Poab ...E.....E..K.....E...-.....-.....
SAMD9_Nole ...E.....K.....E..L-.....-.....
SAMD9_Mamu ...E.....K.....E...-.....A.....-.....
SAMD9_Bota SS.E...R.....Y.....NSVN.....E..A-...F.....K...K.....K..VG..-..H.....
SAMD9_Susc RS.E...R.....C.....NNMN..I.R.....C...AA..K.F.....KE..K..F..K.....-.....
SAMD9_Eqca RS.E...R.....Y.....NNMT.....RN.S-.....RN.....K.....-.....V.D.....
SAMD9_Mylu RS.Q...R.....Y.....NNMN.....YG..L-...IW....E....K.....K.....-.....R.....
SAMD9_Orcu .S.E...R.....Y.....NNMN.....E..L-.....R.....-.....VSD...L....
SAMD9_Rano ...EK..R....RLY.....NSRN..I.....EN.ENLS-....W.HT.K..E.K...D..V..P.....-.....K.
SAMD9_Crgr W..ET..R....R.Y.....E..NNRN..I.....EN.E.VS-....W.H..K..E.K...D.....-N.....Q...
SAMD9_Capo .S.EK.....TNMN.....E..GEMS-...F.....K..E.K...D.....EK.F.....
SAMD9_Soar RL.ET..L.K.R..Y.....A.NIN.....G..--.....N....R.K...D.....N.....-.V..D...K...I.....
SAMD9L_Hosa SS.NR..R....R.C.S..ASTL.Y...R.G.NSI...A..E.Y.D.AQ-NT....H.....KNE.KD..R..T.Q..GK-LISV...TE...K..VIS
SAMD9L_Patr SS.NR..R....R.C.S..ASTL.Y...R.G.NSI...AE.E.Y.D.AQ-NT....H.....KNE.KD..R..T.Q..GK-LIS....TE...K..VIS
SAMD9L_Gogo SS.NR..R....R.C.S..ASTL.Y...R.G.NSI...A..E.Y.D.AQ-NT....H.....KNE.KD..R..T.Q..GK-LIS....TE...K..VIS
SAMD9L_Poab SS.NR..R....R.C.S..ASTL.Y...R.G.NSI...A..E.Y.D.AQ-NT....H.....KNE.KD..R..T.Q..GK-LIS....TE...K..VIS
SAMD9L_Nole SS.NR..R....R.C.S..ASTL.Y...R.G.NSI...AE.E.Y.D.AQ-NT....H.....KNE.KD..R..T.Q..GK-LIS....TE...K..V.S
SAMD9L_Caja SS.NR..R.....C.S..ASTL.....R.G.NSI...AE.E.Y.G.VH-NT.....KNE.KD..R..I.Q..GK-LIS....TK...K..VIS
SAMD9L_Mamu SS.NR..R....R.C.S..ASTL.Y...RRG.NSI...AE.E.Y.D.VQ-NT....H.....KHE.KH..C..I.Q..GK-LIS....TE...K..VIS
SAMD9L_Loaf SS.NR..GR..R..C.S..ASTL.Y...S.G.HSI...AE.E.Y.S.EQ-NT.....KNE.KD..R..T.Q.VGK-LIS....TE...R..VIS
SAMD9L_Eqca SS.NR...R..SN.C.SR.ASTV.Y...K.G.HS....A..E.Y.G.VQ-NT.....N....EKKE.KD..C..T.Q..GK-RIS....TK...K..VIS
SAMD9L_Calu SS.NRN..R..RS.C.S..ASTL.Y...Q.G.HS..C.AE.E.Y.N.AQ-NTS..S..A..C.K-..KD..C..T.Q..GK-LISM...TEK.VK..VI.
SAMD9L_Aime SS.NRT..R..RS.C.S..ASTL.Y...K.G.HS..C.AE.E.Y.N.AQ-NV...SP.....KKE.KD..C..T.Q..GK-LISM...TEK..K..VI.
SAMD9L_Ereu SS..K..NRH.RL.C.S..ASTL.Y..NR.G.NS....AE.E.YIS.VK-NT..F....H...K.E.KG..R..N....GK-LILM...TE...K..VIS
SAMD9L_Orcu SS.NRT.R....R.C.S..ASTF.Y...R.GFNG..L.AE.E.Y.S.AQ-N.....KKE.KD..H..V.Q..GK-LIS....TE...K..V.S
SAMD9L_Mumu SS.NR..RR.....C.SR..STL.Y..QK.G.NS....AE.ERY.SEVQ-.S..F.H..V..EKRE.KD..RL.D.Q..GK-LISL...TEA..K..V.S
SAMD9L_Crgr SS.NR..RR.....C.S.K.STL.Y..QK.G.HS....AE.ERYVSEVQ-NS..F....V..EKRE.KD..RL.D.Q..GK-LISV...TEA..K..V.S
SAMD9L_Rano SS.NR..RR.....C.SR..STL.Y..QK.G.NS....AE.ERY.SEVQ-.S..F....V..EKGE.KDI.RL.D.Q..GK-LISL...TEA..K..V.S
SAMD9L_Capo VS.NR..R....R.C.S..ASTF.Y..TK.G.NS....AELEKHIS..K-NT..F.....I..T-E.KH..H..T.Q..GK-LIS.....
SAMD9L_Soar SS..K..NR..RR.C.S..ASTL.Y..N..G.KG....TE.E.Y.S.V.-NTSI...N.....KKE.KS..C..T.Q..GK-LIS.Q..TE...N..VIS
SAMD9L_Modo T..RKT.WD..GQ.C.G...FTH.Y..NDRG.N....R...E...QDAS-.VSI...R....R....KD.....S..V..D-LI..DC.SK.NAE..VR.

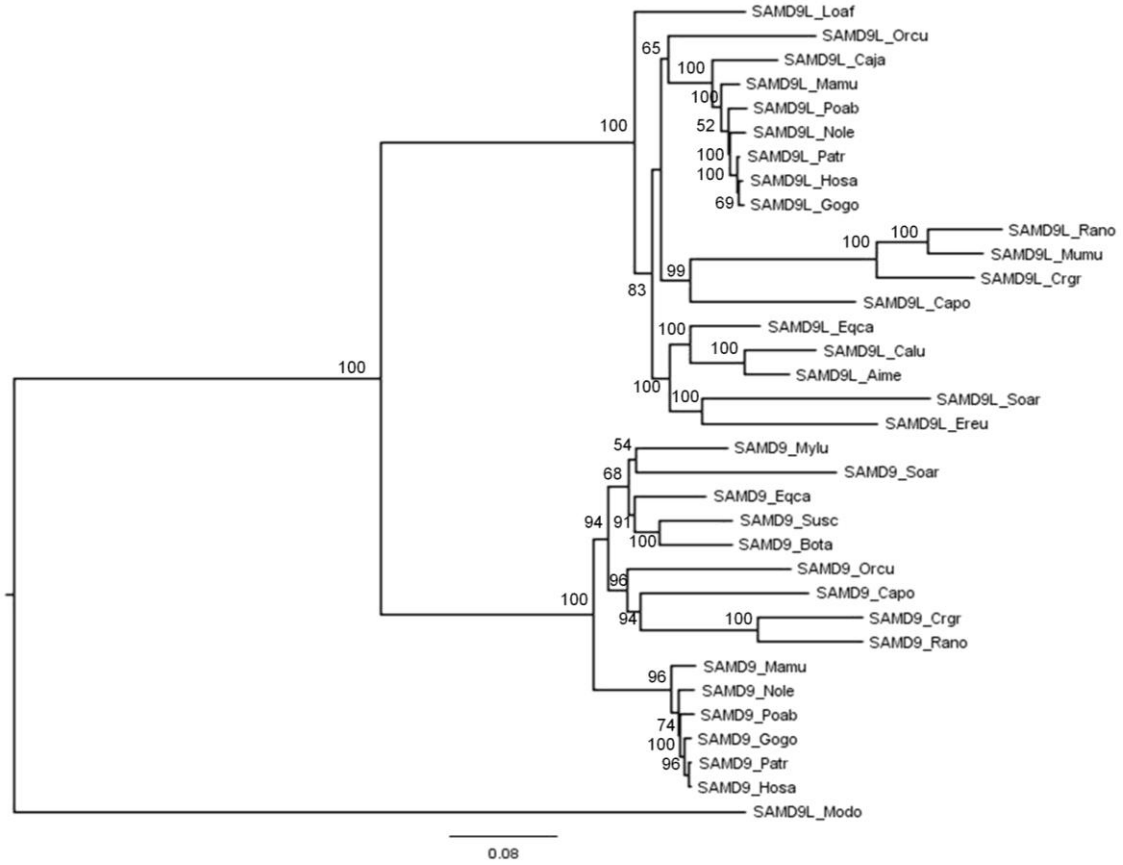
```
                1610      1620      1630
            ....|....|....|....|....|....|....
SAMD9_Hosa  AFLGQLRSGRSIEKVSFYLGFSIGGPLAYDIEIV
SAMD9_Patr  ................................
SAMD9_Gogo  ................................
SAMD9_Poab  ................................
SAMD9_Nole  ................................
SAMD9_Mamu  T...............................
SAMD9_Bota  ...............................I
SAMD9_Susc  ............D.........T.........VI
SAMD9_Eqca  ..F............................I
SAMD9_Mylu  ..W............................I
SAMD9_Orcu  ..W.........D..................I
SAMD9_Rano  THF......................L.....K..
SAMD9_Crgr  TH.......................L.....K..
SAMD9_Capo  ..F..........................Q.L
SAMD9_Soar  T.F..........................QVI
SAMD9L_Hosa VYS.P.....N..R.........E........VI
SAMD9L_Patr VYS.P.....N..R.........E........VI
SAMD9L_Gogo VYS.P.....N..R.........E........VI
SAMD9L_Poab VYS.P.....N..R.........E........VI
SAMD9L_Nole VYS.P.....N..R.........E.......KVI
SAMD9L_Caja VYS.P.....N..R.........E.......QVI
SAMD9L_Mamu VYS.P.....N..R.........E........VI
SAMD9L_Loaf VYS.P....GN..R.........E........VI
SAMD9L_Eqca VYS.P.Q...N..R........ME........VI
SAMD9L_Calu VYS.P....GN..R........ME.LQ..E...I
SAMD9L_Aime VYS.P....GN..R.....A...E..Q......I
SAMD9L_Ereu VYS.P.....N.QR........ME........VI
SAMD9L_Orcu VYS.P.....N............E........V.
SAMD9L_Mumu VYSAP.....N..R.........E.....G.KVI
SAMD9L_Crgr VYS.P.....N..R.V.......E.....G.KVI
SAMD9L_Rano VYS.P.....N..R.........E.....G.KVI
SAMD9L_Capo ..F..........................Q.L
SAMD9L_Soar VYS.P.....H.QS........ME........LI
SAMD9L_Modo .V.....R.Q.M.R........TA..V....KNI
```

Supplementary Table S1 - Detection of recombination breakpoints from *SAMD9* and *SAMD9L* genes alignment using GARD analysis

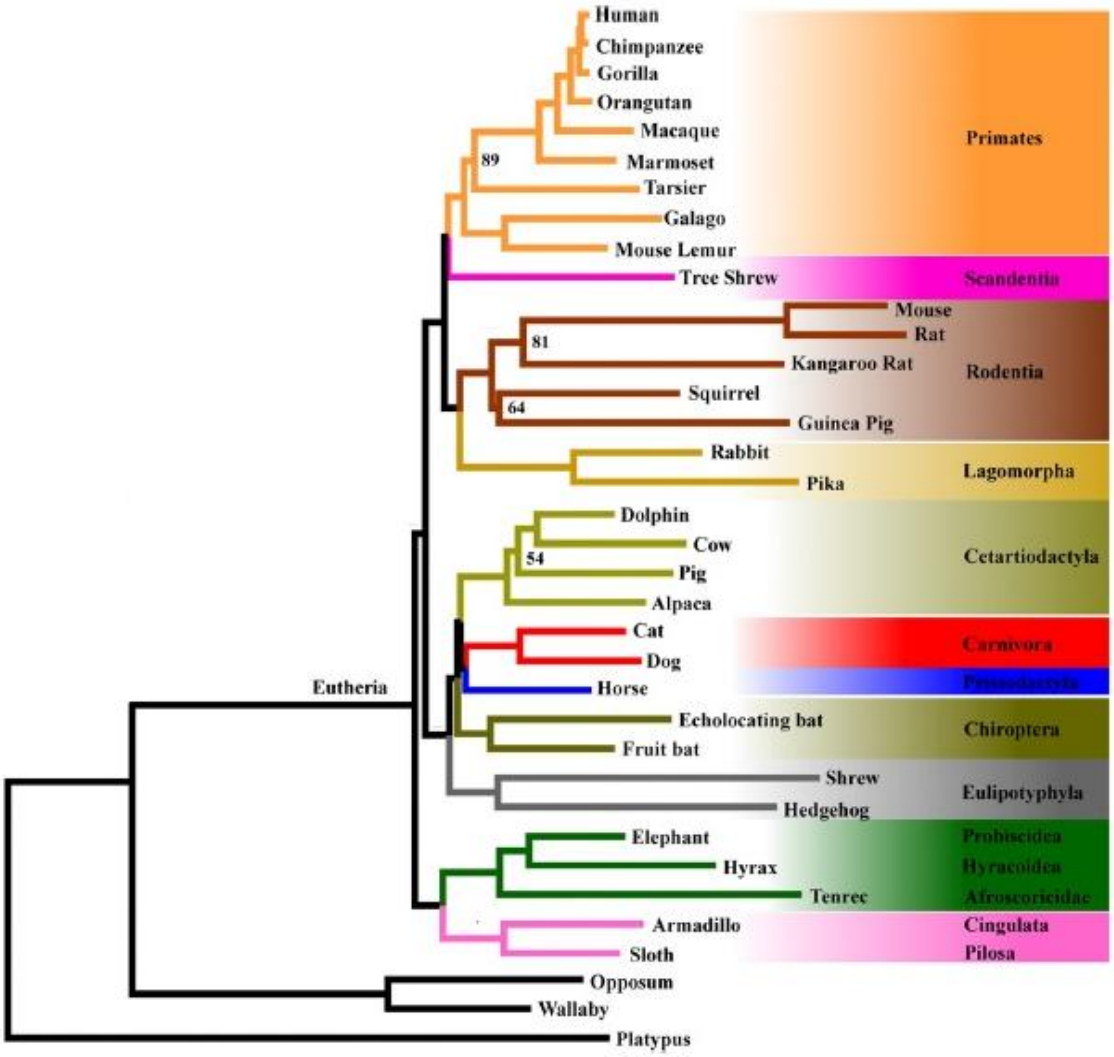
***SAMD9+SAMD9L* alignment**				
Breakpoint	LHS[a] Raw *p*	LHS[a] adjusted *p*	RHS[b] Raw *p*	RHS[b] adjusted *p*
1179	0.94650	1.00000	0.00010	0.00060
2404	0.00010	0.00060	0.40590	1.00000
4755	**0.00010**	**0.00060**	**0.00010**	**0.00060**

[a] LHS: segment to the left of the breakpoint.
[b] RHS: segment to the right of the breakpoint .

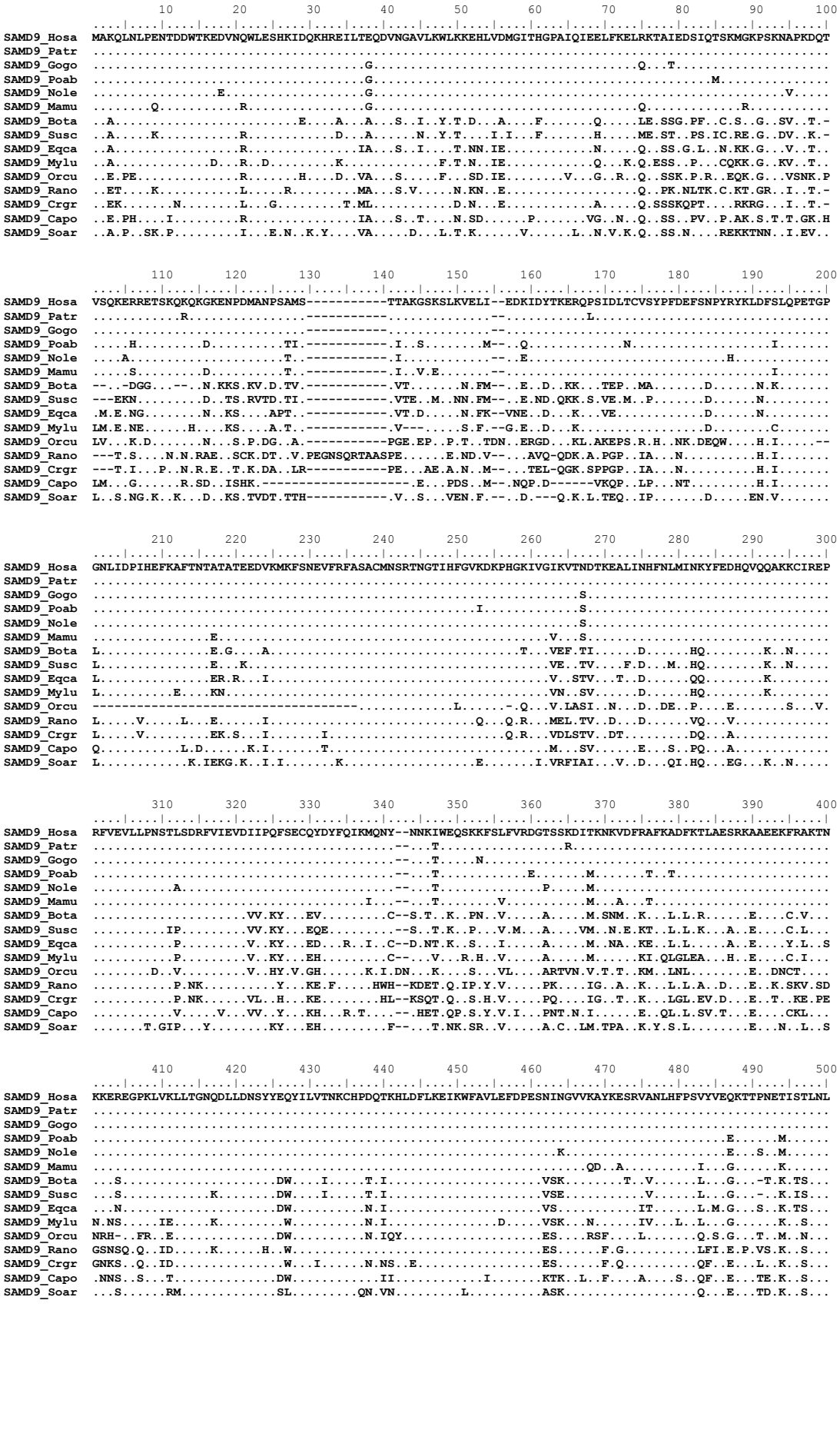
Supplementary Figure S2

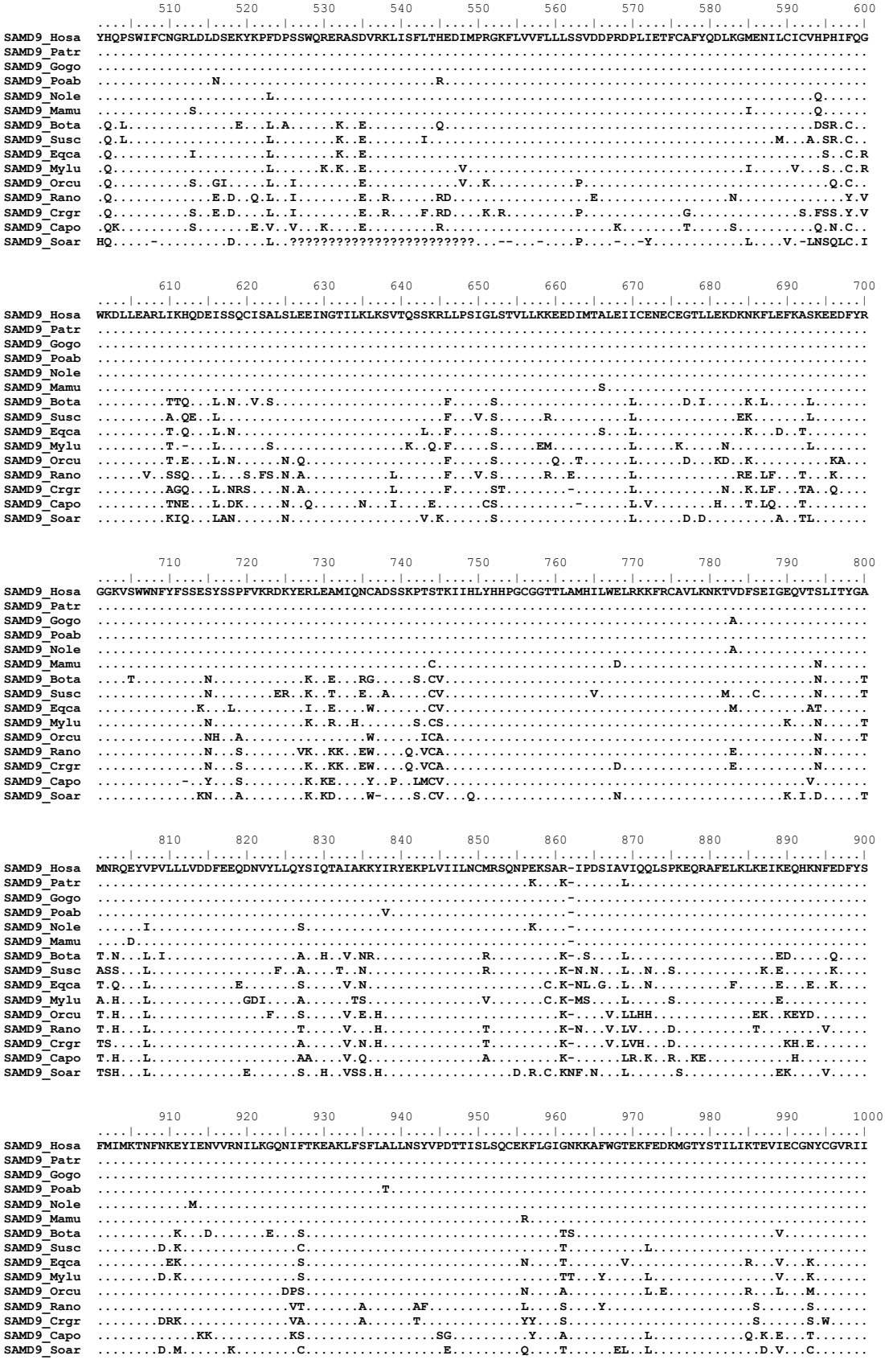


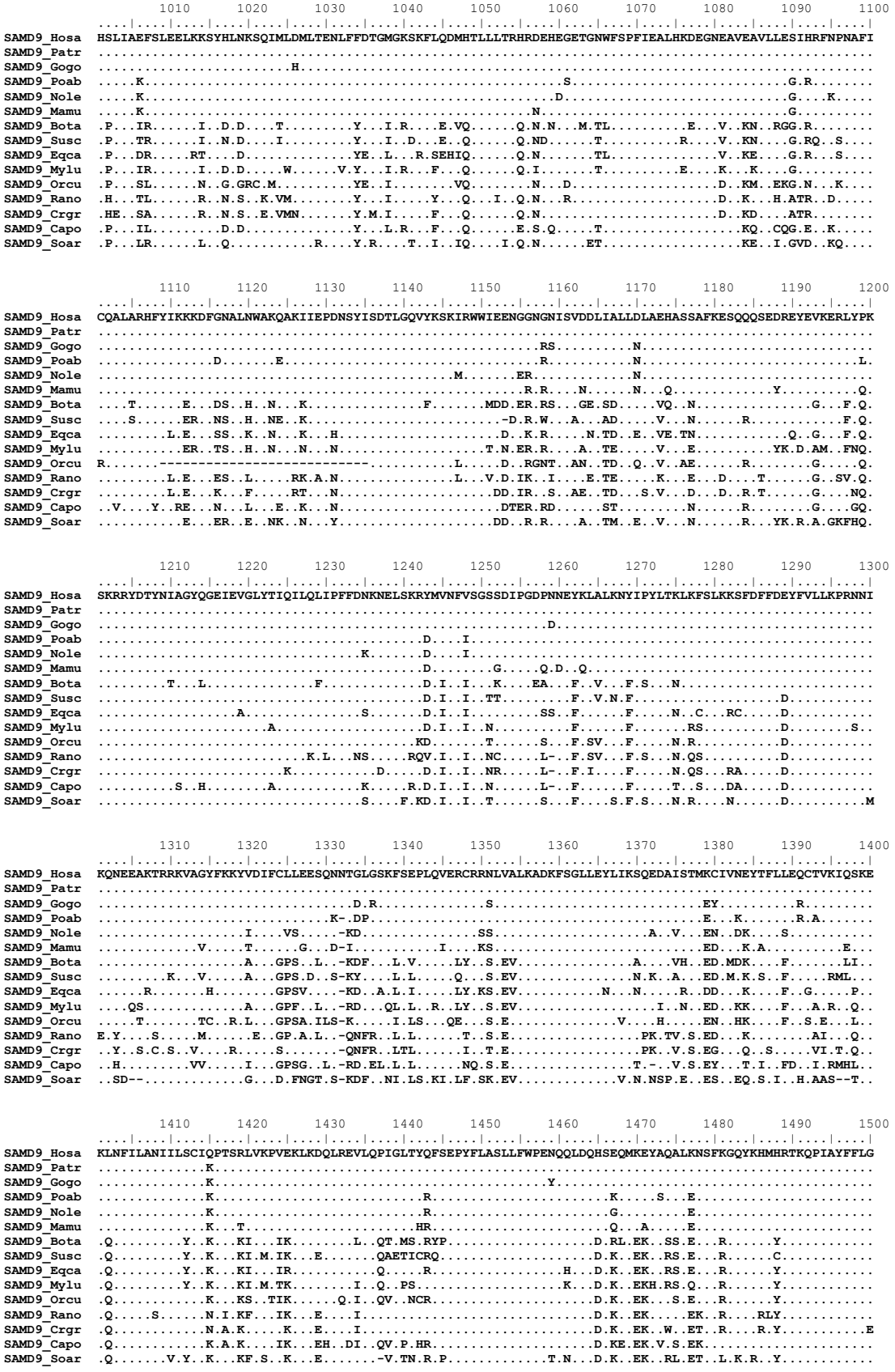
Supplementary Figure S3



Supplementary Figure S4

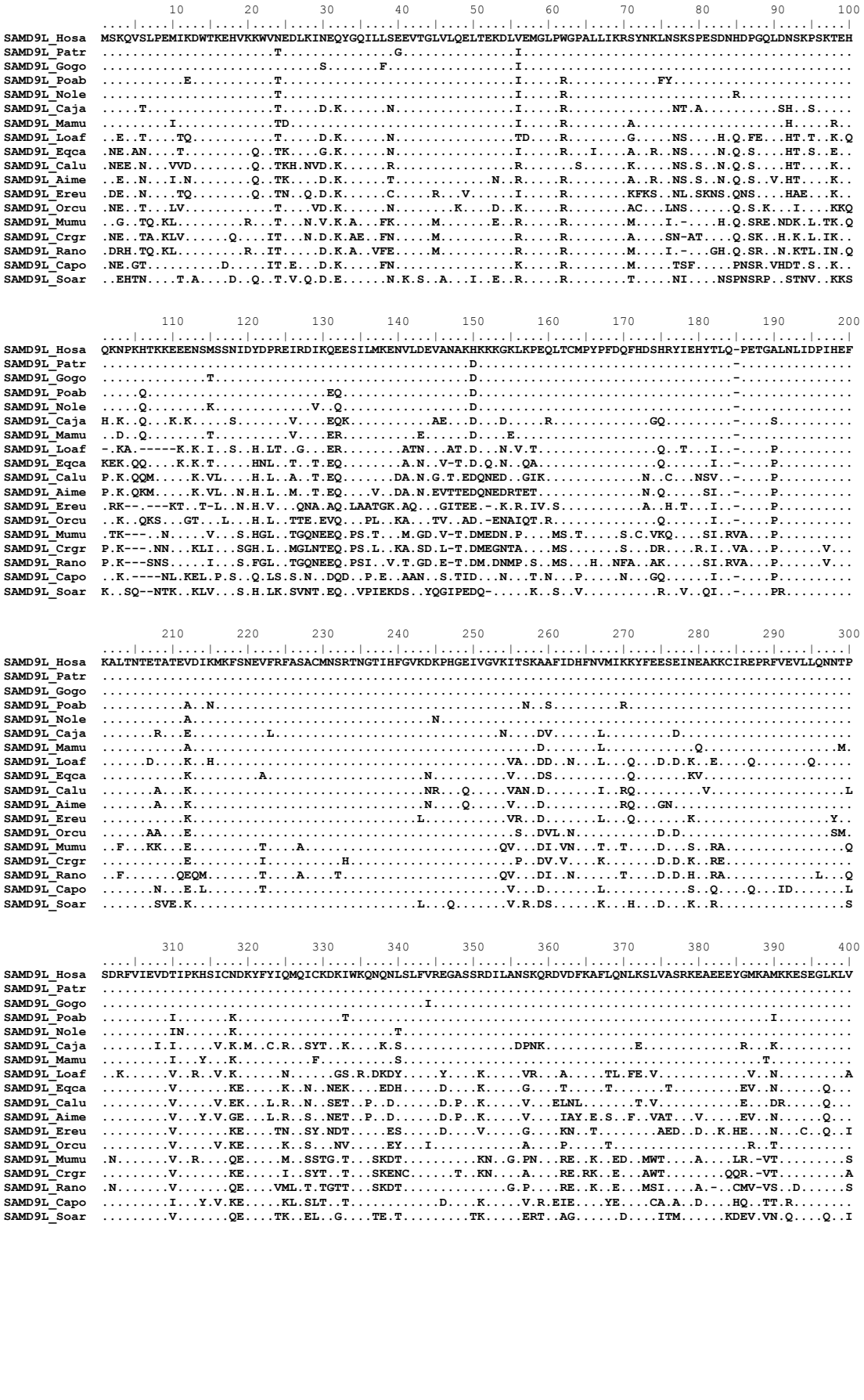


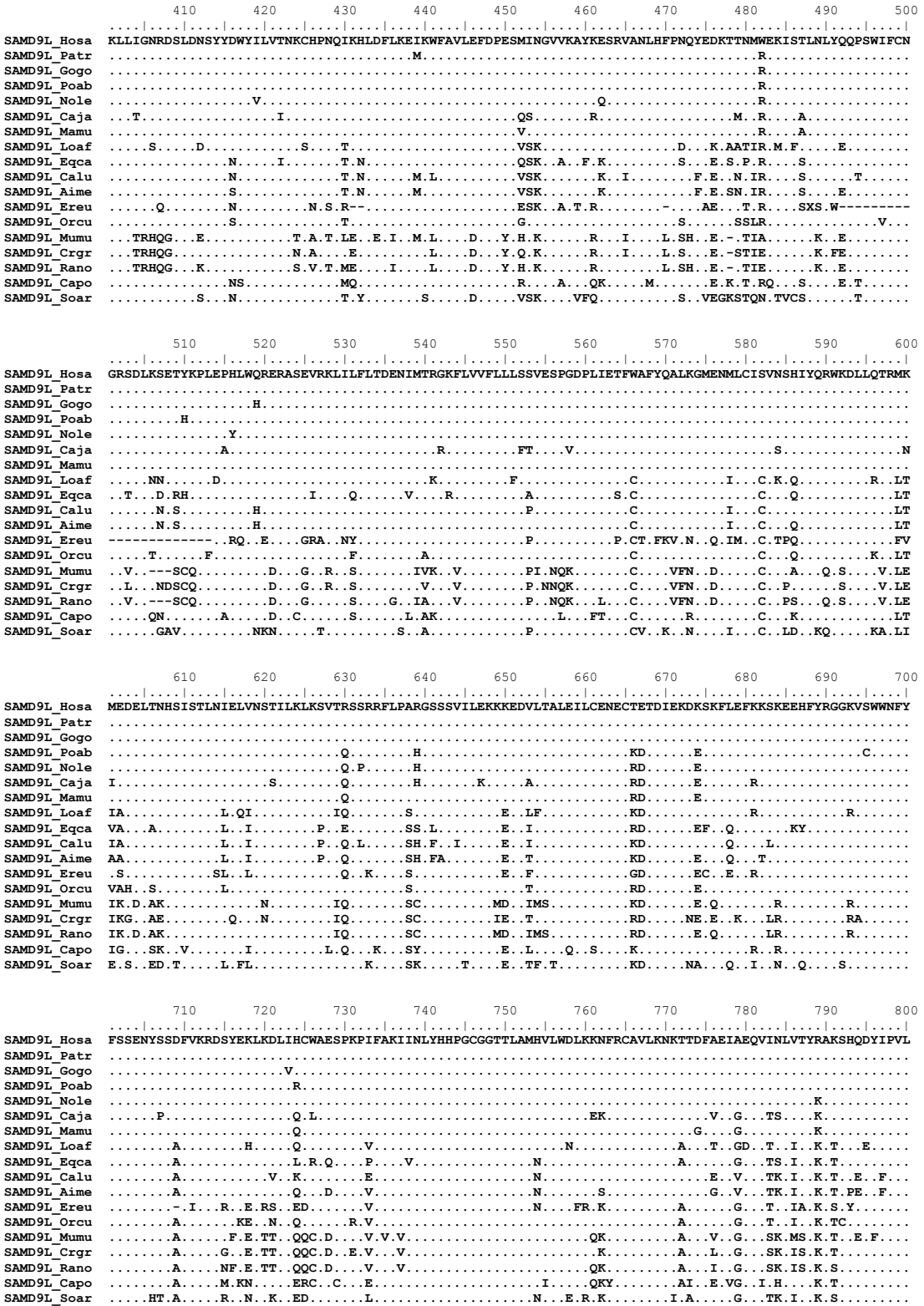

SAMD9_Hosa YHQPSWIFCNGRLDLDSEKYKPFDPSSWQRERASDVRKLISFLTHEDIMPRGKFLVVFLLLSSVDDPRDPLIETFCAFYQDLKGMENILCICVHPHIFQG
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
SAMD9_Hosa WKDLLEARLIKHQDEISSQCISALSLEEINGTILKLKSVTQSSKRLLPSIGLSTVLLKKEEDIMTALEIICENECEGTLLEKDKNKFLEFKASKEEDFYR
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
SAMD9_Hosa GGKVSWWNFYFSSESYSSPFVKRDKYERLEAMIQNCADSSKPTSTKIIHLYHHPGCGGTTLAMHILWELRKKFRCAVLKNKTVDFSEIGEQVTSLITYGA
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
SAMD9_Hosa MNRQEYVPVLLLVDDFEEQDNVYLLQYSIQTAIAKKYIRYEKPLVIILNCMRSQNPEKSAR-IPDSIAVIQQLSPKEQRAFELKLKEIKEQHKNFEDFYS
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
SAMD9_Hosa FMIMKTNFNKEYIENVVRNILKGQNIFTKEAKLFSFLALLNSYVPDTTISLSQCEKFLGIGNKKAFWGTEKFEDKMGTYSTILIKTEVIECGNYCGVRII
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar

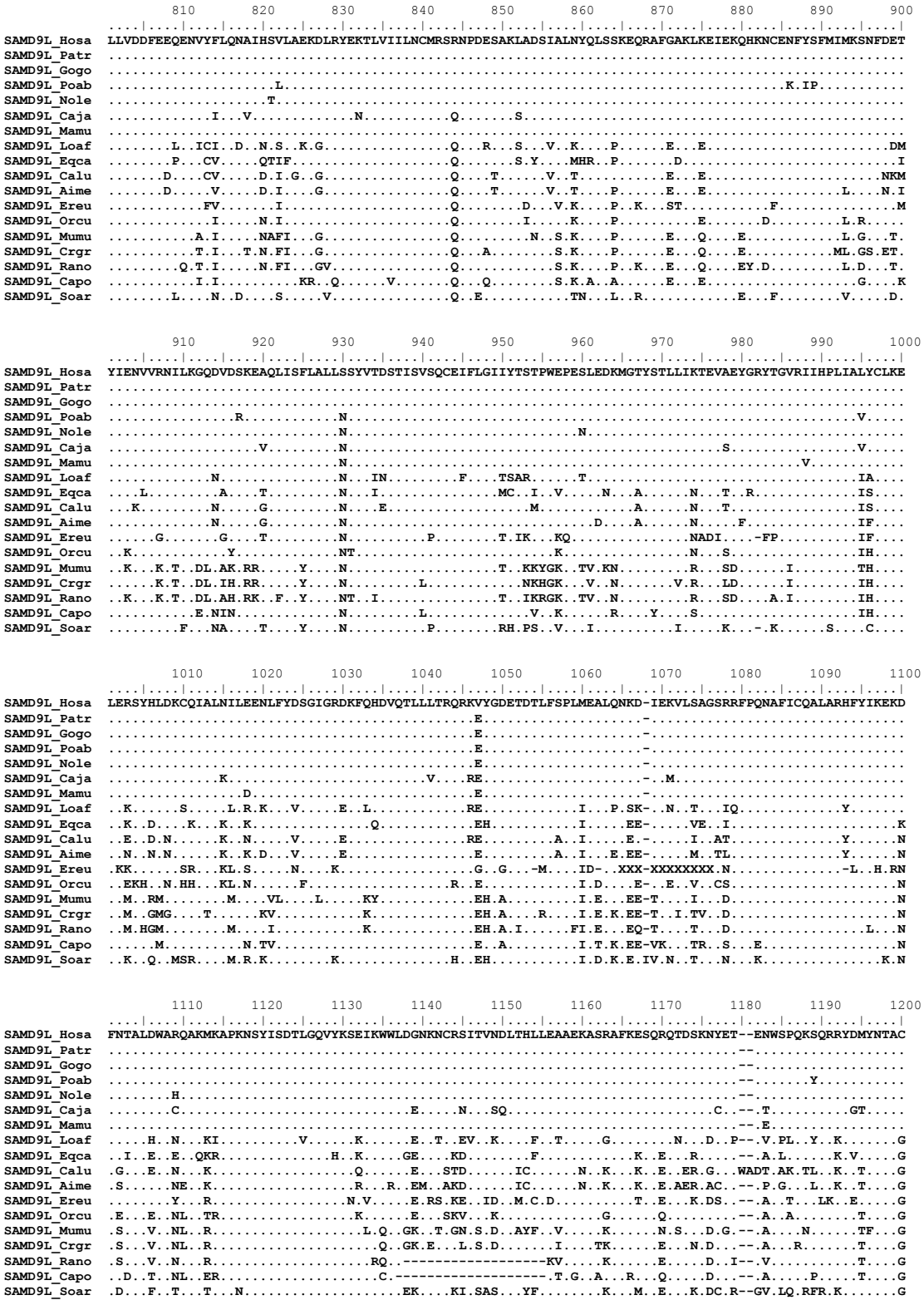

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
SAMD9_Hosa HSLIAEFSLEELKKSYHLNKSQIMLDMLTENLFFDTGMGKSKFLQDMHTLLLTRHRDEHEGETGNWFSPFIEALHKDEGNEAVEAVLLESIHRFNPNAFI
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
SAMD9_Hosa CQALARHFYIKKKDFGNALNWAKQAKIIEPDNSYISDTLGQVYKSKIRWWIEENGGNGNISVDDLIALLDLAEHASSAFKESQQQSEDREYEVKERLYPK
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
SAMD9_Hosa SKRRYDTYNIAGYQGEIEVGLYTIQILQLIPFFDNKNELSKRYMVNFVSGSSDIPGDPNNEYKLALKNYIPYLTKLKFSLKKSFDFFDEYFVLLKPRNNI
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
SAMD9_Hosa KQNEEAKTRRKVAGYFKKYVDIFCLLEESQNNTGLGSKFSEPLQVERCRRNLVALKADKFSGLLEYLIKSQEDAISTMKCIVNEYTFLLEQCTVKIQSKE
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
SAMD9_Hosa KLNFILANIILSCIQPTSRLVKPVEKLKDQLREVLQPIGLTYQFSEPYFLASLLFWPENQQLDQHSEQMKEYAQALKNSFKGQYKHMHRTKQPIAYFFLG
SAMD9_Patr
SAMD9_Gogo
SAMD9_Poab
SAMD9_Nole
SAMD9_Mamu
SAMD9_Bota
SAMD9_Susc
SAMD9_Eqca
SAMD9_Mylu
SAMD9_Orcu
SAMD9_Rano
SAMD9_Crgr
SAMD9_Capo
SAMD9_Soar

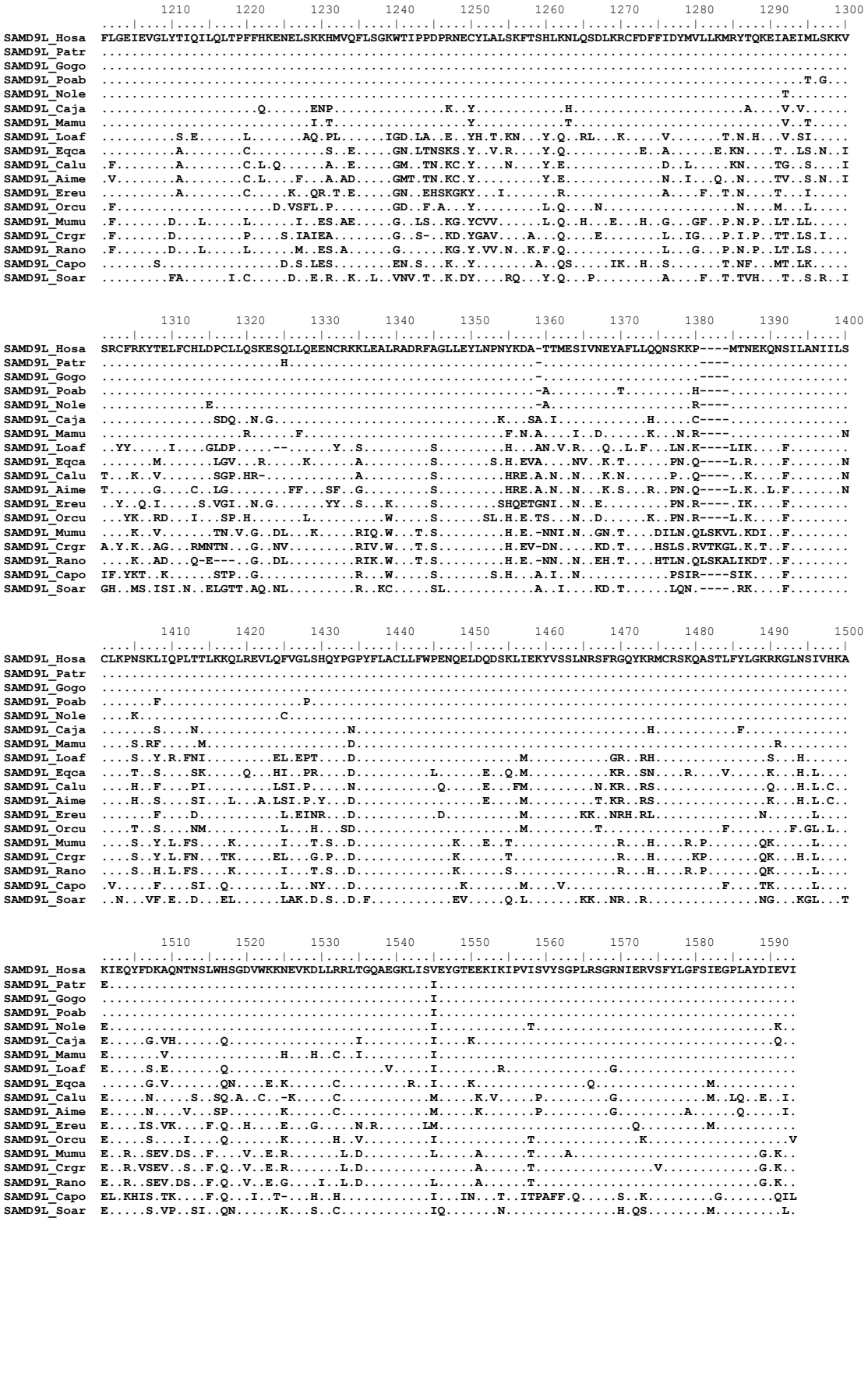
```
                1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
SAMD9_Hosa KGKRLERLVHKGKIDQCFKKTP-DINSLWQSGDVWKEEKVQELLLRLQGRAENN-CLYIEYGINEKITIPITPAFLGQLRSGRSIEKVSFYLGFSIGGPL
SAMD9_Patr ..................E...-............................-..........................................
SAMD9_Gogo ......................-............................-..........................................
SAMD9_Poab ..E..K............E...-............................-..........................................
SAMD9_Nole .....K............E..L-............................-..........................................
SAMD9_Mamu .....K............E...-.............A.............-................T.........................
SAMD9_Bota ..NSVN............E..A-...F.........K...K.....K..VG..-..H.....................................
SAMD9_Susc ..NNMN..I.R......C...AA..K.F.........KE..K..F..K......-.........................D.........T....
SAMD9_Eqca ..NNMT............RN.S-.............RN........K......-.......V.D..........F.....................
SAMD9_Mylu ..NNMN...........YG..L-...IW....E....K.........K......-..........R.........W.....................
SAMD9_Orcu ..NNMN............E..L-.............R.................-.......VSD...L......W.........D............
SAMD9_Rano ..NSRN..I......EN.ENLS-....W.HT.K..E.K...D..V..P......-................K.THF.....................L.
SAMD9_Crgr ..NNRN..I......EN.E.VS-....W.H..K..E.K...D............-N.............Q...TH.....................L.
SAMD9_Capo ..TNMN.........E..GEMS-...F.....K..E.K...D............EK.F.................F.....................
SAMD9_Soar .A.NIN............G..--.......N....R.K...D.....N......-.V..D...K...I.....T.F.....................

           ....|..
SAMD9_Hosa AYDIEIV
SAMD9_Patr .......
SAMD9_Gogo .......
SAMD9_Poab .......
SAMD9_Nole .......
SAMD9_Mamu .......
SAMD9_Bota ......I
SAMD9_Susc .....VI
SAMD9_Eqca ......I
SAMD9_Mylu ......I
SAMD9_Orcu ......I
SAMD9_Rano ....K..
SAMD9_Crgr ....K..
SAMD9_Capo ....Q.L
SAMD9_Soar ....QVI
```

Supplementary Figure S5

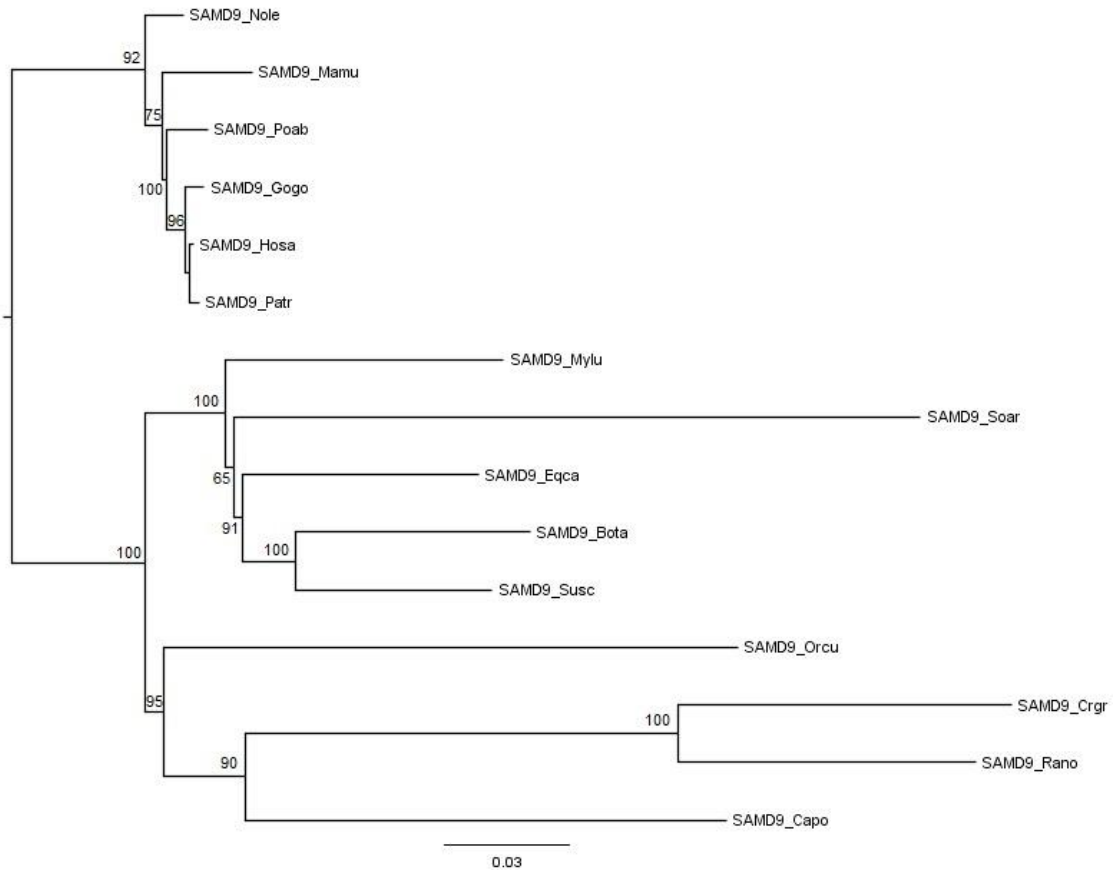



410 420 430 440 450 460 470 480 490 500
SAMD9L_Hosa KLLIGNRDSLDNSYYDWYILVTNKCHPNQIKHLDFLKEIKWFAVLEFDPESMINGVVKAYKESRVANLHFPNQYEDKTTNMWEKISTLNLYQQPSWIFCN
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
510 520 530 540 550 560 570 580 590 600
SAMD9L_Hosa GRSDLKSETYKPLEPHLWQRERASEVRKLILFLTDENIMTRGKFLVVFLLLSSVESPGDPLIETFWAFYQALKGMENMLCISVNSHIYQRWKDLLQTRMK
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
610 620 630 640 650 660 670 680 690 700
SAMD9L_Hosa MEDELTNHSISTLNIELVNSTILKLKSVTRSSRRFLPARGSSSVILEKKKEDVLTALEILCENECTETDIEKDKSKFLEFKKSKEEHFYRGGKVSWWNFY
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
710 720 730 740 750 760 770 780 790 800
SAMD9L_Hosa FSSENYSSDFVKRDSYEKLKDLIHCWAESPKPIFAKIINLYHHPGCGGTTLAMHVLWDLKKNFRCAVLKNKTTDFAEIAEQVINLVTYRAKSHQDYIPVL
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar

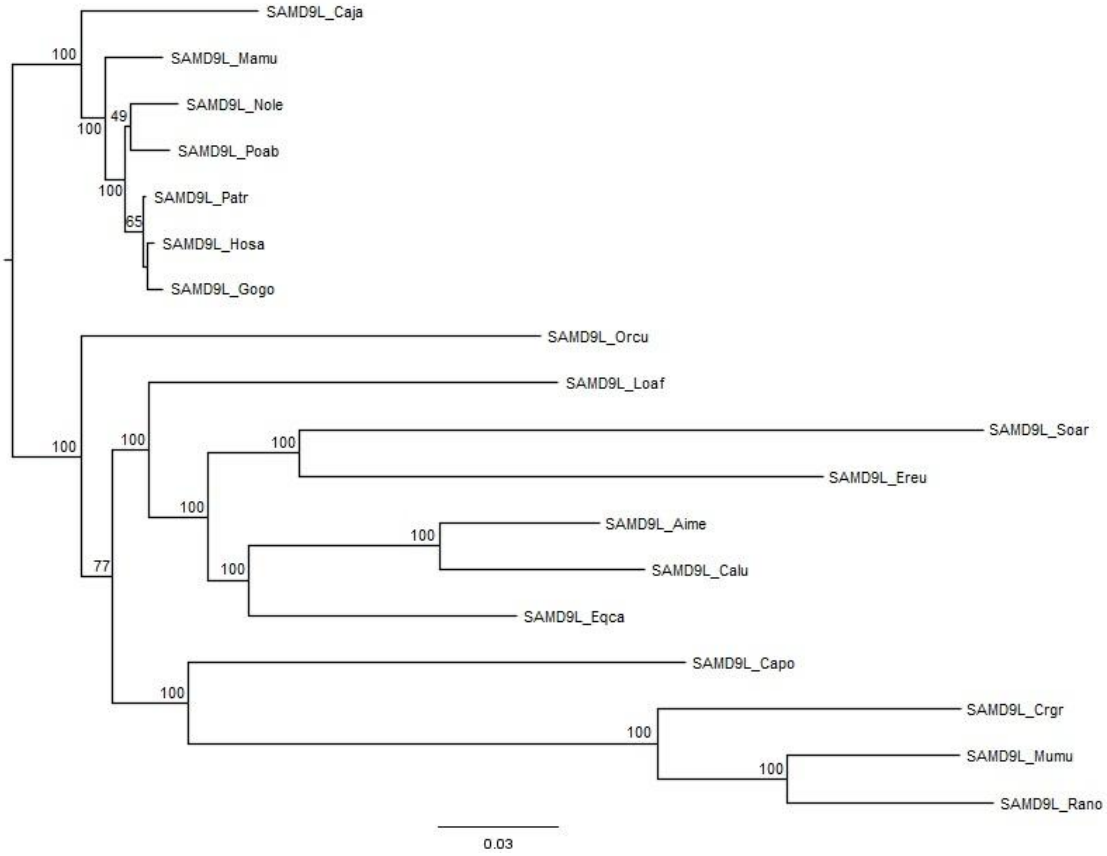

810 820 830 840 850 860 870 880 890 900
SAMD9L_Hosa LLVDDFEEQENVYFLQNAIHSVLAEKDLRYEKTLVIILNCMRSRNPDESAKLADSIALNYQLSSKEQRAFGAKLKEIEKQHKNCENFYSFMIMKSNFDET
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
910 920 930 940 950 960 970 980 990 1000
SAMD9L_Hosa YIENVVRNILKGQDVDSKEAQLISFLALLSSYVTDSTISVSQCEIFLGIIYTSTPWEPESLEDKMGTYSTLLIKTEVAEYGRYTGVRIIHPLIALYCLKE
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
SAMD9L_Hosa LERSYHLDKCQIALNILEENLFYDSGIGRDKFQHDVQTLLLTRQRKVYGDETDTLFSPLMEALQNKD-IEKVLSAGSRRFPQNAFICQALARHFYIKEKD
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
SAMD9L_Hosa FNTALDWARQAKMKAPKNSYISDTLGQVYKSEIKWWLDGNKNCRSITVNDLTHLLEAAEKASRAFKESQRQTDSKNYET--ENWSPQKSQRRYDMYNTAC
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar


1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
SAMD9L_Hosa FLGEIEVGLYTIQILQLTPFFHKENELSKKHMVQFLSGKWTIPPDPRNECYLALSKFTSHLKNLQSDLKRCFDFFIDYMVLLKMRYTQKEIAEIMLSKKV
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
SAMD9L_Hosa SRCFRKYTELFCHLDPCLLQSKESQLLQEENCRKKLEALRADRFAGLLEYLNPNYKDA-TTMESIVNEYAFLLQQNSKKP----MTNEKQNSILANIILS
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
SAMD9L_Hosa CLKPNSKLIQPLTTLKKQLREVLQFVGLSHQYPGPYFLACLLFWPENQELDQDSKLIEKYVSSLNRSFRGQYKRMCRSKQASTLFYLGKRKGLNSIVHKA
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar
1510 1520 1530 1540 1550 1560 1570 1580 1590
SAMD9L_Hosa KIEQYFDKAQNTNSLWHSGDVWKKNEVKDLLRRLTGQAEGKLISVEYGTEEKIKIPVISVYSGPLRSGRNIERVSFYLGFSIEGPLAYDIEVI
SAMD9L_Patr
SAMD9L_Gogo
SAMD9L_Poab
SAMD9L_Nole
SAMD9L_Caja
SAMD9L_Mamu
SAMD9L_Loaf
SAMD9L_Eqca
SAMD9L_Calu
SAMD9L_Aime
SAMD9L_Ereu
SAMD9L_Orcu
SAMD9L_Mumu
SAMD9L_Crgr
SAMD9L_Rano
SAMD9L_Capo
SAMD9L_Soar

Supplementary Figure S6



Supplementary Figure S7



Supplementary Table S2 - *SAMD9* and *SAMD9L* likelihood ratio test (LRT) for PARRIS analysis from HyPhy software

Hypothesis		LRT		
Null Hypothesis	Alternative Hypothesis	-2ΔlnL	df	p Value
PARRIS				
SAMD9				
M1: no selection	M2: selection	0.2	2	n.s.
SAMD9L				
M1: no selection	M2: selection	8.4	2	**< 0.05***

*, significant; n.s., not significant.

Supplementary Table S3 - Positively-selected codon positions in *SAMD9* and *SAMD9L* determined by six different Maximum Likelihood methods

	SAMD9	*SAMD9L*
PAML M8[a]	21*, **48***, 52, **88**, **491**, 736, **1006***, 1085**, 1088, **1116***, 1120*, **1258***, 1348, 1380, **1398***, 1440	71**, 126**, 146*, **156**, **260**, **267***, 329*, **340****, **357***, 368*, **452****, **586***, 639, **653***, 1145*, 1183, **1276***, 1295**, 1318*, 1339, 1380*, 1408*
SLAC[b]	**352**, **383**, **491***, **513**, **731**, **1006***, **1116**, **1320***, **1329**, **1398***	**156**, **267***, **340****, **362***, **586***, **606**, **776***, **978***, 1018, **1186**, **1229**, **1429**, **1474**
FEL[b]	38, 45*, 75*, **88***, 151, 170*, 176, **279***, **331**, **352***, **383***, 395*, **491***, **513***, 596*, 610, 618*, **731**, 735, 783*, **872**, 875*, **993***, **1006****, 1095, **1116****, 1127, **1258***, **1320***, **1329***, 1333, 1351, 1376, 1393, **1398****, 1424, 1503, 1528	5, **39***, 56, 76, 83, 88, 92, **156***, **260**, **267****, 282, **340****, **357***, **362****, **452****, 519, 540, 585, **586***, 602, **606***, **653***, **776***, 783, **978***, 1006, 1154*, **1186**, 1189*, **1229***, 1255, **1276**, **1308**, 1360*, 1364, 1374, 1413, **1429***, **1474***
REL[c]	**48**, 52, 69, 106, 116, 193, **279**, 312, **331**, 372, 376, 397, **491**, 594, 611, **731**, 736, 827, **872**, 956, **993**, **1006**, 1017, 1085, 1088, 1092, **1116**, 1120, 1167, 1177, **1258**, 1314, **1320**, 1327, **1329**, 1348, 1380, **1398**, 1519, 1576	**39**, 126, **156**, **260**, **267**, 279, 325, 328, **340**, **357**, **362**, **452**, **586**, 601, 639, **653**, 721, **776**, 822, 953, **978**, 996, 1018, 1078, 1106, 1109, 1183, **1229**, 1231, **1276**, 1295, **1308**, 1316, 1335, 1367, 1380, 1413, **1429**, 1535
MEME[b]	21, 38**, 45*, **48**, 61*, 75*, 77*, 79**, 85**, **88***, 96, 109, 121, 143, 145, 170, 173, 176, 187, 191, 193, 201**, 218*, **279**, **331***, **352****, 353, 365, 369, **383**, 387, 395, **491***, **513***, 596, 610, 618, 623*, 642*, **731**, 735**, 744, 783, 859, **872**, 875*, 969**, **993***, 1003*, **1006****, 1007, 1041, 1063, **1116***, 1174*, 1199, 1252, **1258***, **1320***, **1329***, 1341*, 1347*, 1351*, 1376*, 1393, **1398****, 1415, 1422, 1444**, 1467*, 1528, 1556**, 1566*	5, 24*, **39***, 76*, 92, 142*, 149*, **156***, 170, 181*, 212*, 213*, **267****, 282, 299, 318, 323*, **340****, 341**, 350, 351, **357***, 359, **362***, 366, 373*, 377*, 426*, **452****, 468*, 509, 519*, 540, 555*, 585, **586***, 602, **606**, 632**, **653**, 723, 725*, 754, **776***, 789*, 791*, 848, 856, 859, 861*, 888**, 889*, 945*, 963, **978***, 1004*, 1006*, 1109*, 1150*, 1154, 1170, **1186***, 1189, **1229***, 1250, 1252*, **1276***, 1284, 1287, **1308**, 1335**, 1339*, 1343, 1346, 1360*, 1374*, 1385, 1390, **1429***, **1474***, 1535
FUBAR[d]	**1006**, **1398**	**156**, **267***, **340****, **357**, **452**

[a] Codons with posterior probabilities >90% in the BEB analyses (*: P>95%; **: P>99%).
[b] Codons with significance level <0.1 (*: p<0.05; **: p<0.01).
[c] Codons with Bayes Factor >50.
[d] Codons with posterior probabilities >0.90 (*: P>0.95; **: P>0.99).
Codons identified by three or more than three methods are underlined and in bold.

Supplementary Table S4 - SAMD9 and SAMD9L amino acid properties under positive selection determined in TreeSAAP

Protein	Property	Category[a]	z-Score[b]
SAMD9	Isoelectric point	8	8.218***
	Equilibrium constant (ionization of COOH)	8	5.001***
	Power to be at the C-terminal	7	4.032***
	Alpha-helical tendencies	6	5.581***
SAMD9L	Isoelectric point	8	6.922***
	Equilibrium constant (ionization of COOH)	8	4.797***
	Solvent accessible reduction ratio	7	4.366***
	Power to be at the middle of alpha-helix	7	3.554***
	Power to be at the C-terminal	6	10.442***
	Alpha-helical tendencies	6	5.228***
	Surrounding hydrophobicity	6	3.921***

[a] Between 6 and 8 (the most radical values denoting positive destabilizing selection).
[b] z-Score > 3.09, level of significance p<0.001 (***).

Chapter 4

Innate anti-viral factors – Evolution and genetic characterization of RIG-I-like receptors

Paper 3

Positive evolutionary selection on the RIG-I-like receptor genes in mammals

A. Lemos de Matos, G. McFadden & P. J. Esteves

1. Summary

The mammalian RIG-I-like receptors, RIG-I, MDA5 and LGP2, are a family of DExD/H box RNA helicases responsible for the cytoplasmic detection of viral RNA. These receptors detect a variety of RNA viruses, or DNA viruses that express unusual RNA species, many of which are responsible for a great number of severe and lethal diseases. Host innate sentinel proteins involved in pathogen recognition must rapidly evolve in a dynamic arms race with pathogens, and thus are subjected to long-term positive selection pressures to avoid potential infections. Using six codon-based Maximum Likelihood methods, we were able to identify specific codons under positive selection in each of these three genes. The highest number of positively selected codons was detected in MDA5, but a great percentage of these codons were located outside of the currently defined protein domains for MDA5, which likely reflects the imposition of both functional and structural constraints. Additionally, our results support LGP2 as being the least prone to evolutionary change, since the lowest number of codons under selection was observed for this gene. On the other hand, the preponderance of positively selected codons for RIG-I were detected in known protein functional domains, suggesting that pressure has been imposed by the vast number of viruses that are recognized by this RNA helicase. Furthermore, the RIG-I repressor domain, the region responsible for recognizing and binding to its RNA substrates, exhibited the strongest evidence of selective pressures. Branch-site analyses were performed and several species branches on the three receptor gene trees showed evidence of episodic positive selection. In conclusion, by looking for evidence of positive evolutionary selection on mammalian RIG-I-like receptor genes, we propose that a multitude of viruses have crafted the receptors biological function in host

defense, specifically for the RIG-I gene, contributing to the innate species-specific resistance/susceptibility to diverse viral pathogens.

2. Introduction

The mammalian innate immune system operates as the first line of defense against microbial pathogen invasion [1-3]. This system recognizes infectious agents through a limited number of germline-encoded pattern-recognition receptors (PRRs) predominantly expressed on sentinel cells [2, 4-6]. The host PRRs recognize and react with specific microbial components, known as pathogen-associated molecular patterns (PAMPs), which includes bacterial lipopolysaccharides, peptidoglycans, lipoteichoic acids and cell-wall lipoproteins, fungal β-glucan and viral nucleic acids [2, 3, 5, 6]. The host PRRs exhibit distinct expression patterns and following sensing of their cognate ligands, activate specific signaling pathways that lead to the expression of a variety of inducible self-defense genes involved in the collective inflammatory and immune responses [2]. To date, four different classes of PRRs have been identified, including the cell membrane-associated C-type lectin receptors (CLRs), the Toll-like receptors (TLRs) at the cell surface and at the membrane of intracellular vesicles (endosomes and lysosomes), and the cytoplasmic detection systems for intracellular PAMPs, namely the RIG-I-like receptors (RLRs) and the NOD-like receptors (NLRs) [3, 6-8].

The RLRs are a family of DExD/H box RNA helicases critically and exclusively involved in the recognition of "nonself" RNA from actively replicating viruses in the cytoplasm of infected cells [9]. This receptor family consists of three members, the retinoic acid-inducible gene-I (RIG-I/*DDX58*), the melanoma differentiation associated factor protein 5 (MDA5/*IFIH1*) and the laboratory of genetics and physiology 2 (LGP2/ *DHX58*) [e.g. 10-14]. RIG-I and MDA5 share high sequence similarity and several structural features, including an N-terminal region consisting of tandem caspase activation and recruitment domains (CARDs), a central DExD/H box RNA helicase domain and a C-terminal domain (CTD). The two N-terminally located CARDs function as a signaling and interaction domain with other CARD-containing proteins [13, 15, 16]. The helicase domain retains catalytic activity to bind and unwind double stranded RNA (dsRNA) in an ATP hydrolysis-dependent manner [10, 17]. The CTD plays a predominant role in high-affinity binding with dsRNA, encoding a repressor domain (RD) in RIG-I, but not in MDA5, which harbors an RD-like domain that does not participate in autoregulation [18]. These two RLRs detect a variety of both DNA and RNA viruses, particularly at early phase of viral replication, and signal the production of type I interferons (IFNs) and induction of an anti-viral response [10, 17]. The third element of the RLR family,

the LGP2 protein, lacks any CARDs but contains the helicase domain and the CTD also harbors a RD. The role of LGP2 in anti-viral immunity is less clear, but it has been suggested in different studies to serve both as a negative and a positive regulator of RIG-I and MDA5 signaling [10, 19-21].

Despite the similarities between RIG-I and MDA5, they were shown to play different roles in anti-viral immunity by recognizing and protecting from specific classes of viruses [22]. RIG-I detects preferentially and most effectively short RNA sequences marked with 5'-triphosphate group (5'-ppp) and blunt-end of short double-stranded RNAs (dsRNAs) or single-stranded RNA (ssRNA) hairpins [23-27]. As a key sensor of ssRNA viruses, RIG-I is implicated in the response to Arenaviridae [28], Bunyaviridae [28], Filoviridae [28], Flaviviridae [18, 29], Orthomyxoviridae [22, 30], Paramyxoviridae [22, 28, 30, 31] and Rhabdoviridae [22, 23]. On the other hand, MDA5 is activated by high-molecular-weight poly(I:C) fragments [22, 32], and also by long-duplex RNAs from the genomes of dsRNA viruses [30] or dsRNA replication intermediates of positive-strand viruses, such as Caliciviridae [33], Coronaviridae [34] and Picornaviridae [22, 32]. Regardless the virus recognition specificity by RIG-I and MDA5, some viruses are redundantly sensed by both RLRs, such as the West Nile virus and the Dengue virus [30, 35]. In addition to the extensively described recognition of RNA viruses by RIG-I and MDA5, a role in anti-viral signaling in response to several dsDNA viruses has also been observed. As an RNA sensor, RIG-I does not detect DNA directly; however, not only do many DNA viruses create dsRNA products by virtue of convergent transcriptional units derived from opposite strands, but also the host RNA polymerase III can mediate the transcription of cytoplasmic DNA templates (such as transfected poly dA:dT) into dsRNA containing 5'-triphosphate, which will activate RIG-I and trigger the production of type I IFN [36, 37]. Both Epstein-Barr virus and myxoma virus are detected by RIG-I, while vaccinia virus is sensed by MDA5 [38-40]. It is also likely that the precise RLRs utilized for the sensing of specific viruses also operate within cell-specific contexts as well.

Interaction between host and pathogen results in a dynamic arms race. Whenever pathogens develop strategies to overtake the host immune system, the host proteins involved in pathogen recognition have to respond by evolving to avoid or reduce potential infections. These dynamics result in host-pathogen adaptation and counter-adaptation, which in turns lead to the rapid co-evolution of both parties. Particularly for the host, this accelerated molecular evolution is often reflected in host defense genes that exhibit strong signatures of ongoing diversifying selection [41, 42]. Because viruses are responsible for a great number of severe and lethal diseases, together with the important role that RLRs play in mammalian innate immune system,

we expect that RIG-I, MDA5 and LGP2 genes may have been under intense selective pressures in all mammals. We have previously demonstrated that one other class of mammalian PRRs, the TLRs, exhibit striking evidence of positive genetic selection as a result of selective pressures exerted by pathogens [43]. Using six different codon-based Maximum Likelihood (ML) methods, we searched for evidence of long-term selective pressures in the three RLR genes present in the available sequenced mammalian genomes and, where possible, pinpoint positively selected residues that might be involved in the host-virus interactions that have shaped their rapid diversification. Specific lineages subject to episodic positive selection have also been identified in the three RLR genes by using two different branch-site models.

3. Materials and Methods

3.1. Mammalian RIG-I-like receptor gene sequences

The coding region of the three RLR genes, *RIG-I*, *MDA5* and *LGP2*, were collected for different mammalian species from NCBI (http://www.ncbi.nlm.nih.gov) and Ensembl (http://www.ensembl.org/index.html) databases (Supplementary Table S1). Each set of mammalian orthologous gene sequences was aligned with Clustal W [44] implemented in BioEdit v7.0.9 [45]. The nucleotide sequences alignment corresponding to each gene coding region is represented in Supplementary Figure S1 (*RIG-I* alignment), Supplementary Figure S2 (*MDA5* alignment) and Supplementary Figure S3 (*LGP2* alignment). Translation into protein sequences was performed using also BioEdit [45]. Supplementary Figure S4, Supplementary Figure S5 and Supplementary Figure S6 represent the alignments of the deduced protein sequences for RIG-I, MDA5 and LGP2, respectively. For the evolutionary analyses, representative alignment gaps in Supplementary Figure S1, Supplementary Figure S2 and Supplementary Figure S3 had to be removed: gaps present in all sequences, with the exception of one or two, have been removed, while gaps present in only one or two sequences were kept. Supplementary Figure S7 (*RIG-I* alignment), Supplementary Figure S8 (*MDA5* alignment) and Supplementary Figure S9 (*LGP2* alignment) correspond to trimmed versions of the nucleotide sequences alignment of each RLR gene.

3.2. Phylogenetic reconstruction analyses

The nucleotide sequences alignment of each gene was firstly tested for recombination, as this biological process can mislead phylogenetic and positive selection analyses [46]. We used the software GARD (Genetic Algorithm for Recombination Detection)

[47, 48], implemented in the Datamonkey web server [49, 50], to detect possible recombination breakpoints on each alignment. The nucleotide substitution model for each phylogenetic reconstruction was indicated by the Akaike Information Criterion (AIC) implemented in jModelTest v0.1.1 [51].

Regarding the *RIG-I* gene one breakpoint was identified, but it was not supported by the Kishino-Hasegawa test. Therefore, the complete alignment was used for the gene phylogeny reconstruction and GTR+G nucleotide substitution model was indicated as the best-fitting model. On the other hand, the software GARD found evidence of two breakpoints in the *MDA5* gene alignment. However, only one of the breakpoints (nucleotide 903) reflected a significant topological incongruence (Kishino-Hasegawa test, $p<0.01$), suggesting that the multiple tree model can be preferred over the single tree model. We reconstructed *MDA5* phylogeny for the first 903 nucleotides of the mammalian alignment as also for the remaining 2211 nucleotides. To compare the different *MDA5* trees topology, we also used the complete alignment (no recombination testing) to reconstruct the gene phylogeny and GTR+G nucleotide substitution model was indicated by the AIC as the best-fit. For the *MDA5* segments which resulted from recombination detection, the best-fitting nucleotide substitution models were TIM3+G (first segment) and TIM3+I+G (second segment). Finally, we found no evidence of recombination for the *LGP2* gene alignment. The best-fitting nucleotide substitution model determined for this alignment was the TPM2uf+I+G model.

ML phylogenetic reconstruction was performed for the three genes using GARLI v2.0 (Genetic Algorithm for Rapid Likelihood Inference) [52]. The analyses were performed with 1,000,000 generations and 1,000 bootstrap searches. ML trees were displayed using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/).

3.3. Molecular evolutionary analyses

Codon substitution models implemented in the CODEML program in PAML v4.4 (Phylogenetic Analysis by Maximum Likelihood) package [53, 54] were applied to the trimmed alignments of *RIG-I* (Supplementary Figure S7), *MDA5* (Supplementary Figure S8) and *LGP2* (Supplementary Figure S9). The codon frequency model F3x4 was fitted to all the alignments. Two pairs of site-specific models were used, M1a (nearly neutral) versus M2a (selection) and M7 (neutral, beta) versus M8 (selection, beta & ω). In these comparisons, M1a and M7 neutral models (null hypothesis) do not admit positive selection, while M2a and M8 alternative models allow positive selection. A likelihood ratio test (LRT) with 2 degrees of freedom was performed, where a significant LRT

demonstrates that the selection model fits better than the neutral model [55-57]. From the HyPhy software available on the Datamonkey web server [49, 50], the PARRIS method [58] was also applied to detect if a proportion of sites in each RLR alignment evolved with ω (d_N/d_S) > 1.

Six different codon-based ML methods were applied to detect codons under positive selection on mammalian *RIG-I*, *MDA5* and *LGP2* trimmed alignments, and based on the methodology adopted by other authors and in previous studies [43, 59, 60], only codons identified by at least three of the six used methods were considered to be under positive selection. Model M8 from PAML package [53, 54] was one of the codon-based methods used to detect codons under positive selection, and a Bayes empirical Bayes (BEB) approach was employed to detect codons with a posterior probability >90% of being under selection [61]. Five other methods, using HyPhy software implemented in the Datamonkey web server [49, 50], were also applied to detect sites under selection for the three genes: the Single Likelihood Ancestor Counting (SLAC) method, the Fixed Effect Likelihood (FEL) method, the Random Effect Likelihood (REL) method [62] and the recently described Mixed Effects Model of Evolution (MEME) [63] and Fast Unbiased Bayesien AppRoximation (FUBAR) [64] methods. To avoid a high false-positive rate [62], sites with *p*-values <0.1 for SLAC, FEL and MEME models, Bayes Factor >50 for REL model and a posterior probability >0.90 for FUBAR were accepted as candidates for selection.

Two branch-site models allowing ω ratios to vary both among lineages and amino acid sites were performed: the PAML branch-site model A [65] and the Hyphy branch-site REL method [66]. When performing PAML branch-site model A [65], every species branch was analyzed as a foreground branch independently. For each analysis of a foreground branch, the remaining lineages were denominated as background branches. In branch-site model A, three ω ratios are assumed for foreground ($0 < \omega_0 < 1$, $\omega_1 = 1$, $\omega_2 > 1$) and two ω ratios for background ($0 < \omega_0 < 1$, $\omega_1 = 1$). The null model is the same as model A, but $\omega_2 = 1$ is fixed [65]. The BEB approach was also used to calculate the posterior probability of a specific codon site and to identify those most likely to be under positive selection (posterior probability >90%) [61]. On the other hand, the branch-site REL method [66] was applied to identify branches where a proportion of sites evolved under episodic diversifying selection.

4. Results

4.1. Mammalian RIG-I-like receptor gene sequences

Publicly available mammalian *RIG-I*, *MDA5* and *LGP2* gene sequences were collected from Ensembl and NCBI databases (Supplementary Table S1) for phylogenetic and selection analyses. The nucleotide coding sequences for each of the three RLR gene orthologous were aligned and are represented in Supplementary Figure S1 (*RIG-I* alignment), Supplementary Figure S2 (MDA5 alignment) and Supplementary Figure S3 (*LGP2* alignment). The translation into deduced protein sequences is also represented in Supplementary Figure S4 (RIG-I alignment), Supplementary Figure S5 (MDA5 alignment) and Supplementary Figure S6 (LGP2 alignment).

The inherent limitations of using solely publicly available mammalian RLRs sequences should be highlighted, although several studies have used the same source of data for general conclusions about other genes in mammals [43, 60, 67-71]. The analyses performed ahead use only an individual representative of each included species and therefore, any drawn conclusions should be carefully considered.

4.2. Phylogenetic reconstruction of mammalian RIG-I-like receptors

ML trees were reconstructed for *RIG-I*, *MDA5* and *LGP2* genes after looking for evidence of recombination using the software GARD [47, 48] implemented in the Datamonkey web server [49, 50].

In mammalian *RIG-I* phylogenetic reconstruction, the monophyly of six of a total of eight taxonomic orders was observed (Figure 1). However, an interesting fact was registered for the two remaining orders, Rodentia and Lagomorpha, when the European rabbit (order Lagomorpha) grouped with the rodent cluster. When looking carefully at the European rabbit RIG-I deduced protein sequence (Supplementary Figure S4), a great number of conserved regions between this species and the other mammalian species was observed, with the exception of a region between codons 840 and 879. This 40 amino acid domain region was unique to the European rabbit RIG-I. We originally speculated that this difference might have been the result of a gene conversion event with adjacent genes. However, when we examined the genes that are chromosomally adjacent to European rabbit *RIG-I* (*NDUFB6*, *TOPORS* and *FRP*), no clear evidence of gene conversion was detected by the software GARD [47, 48].

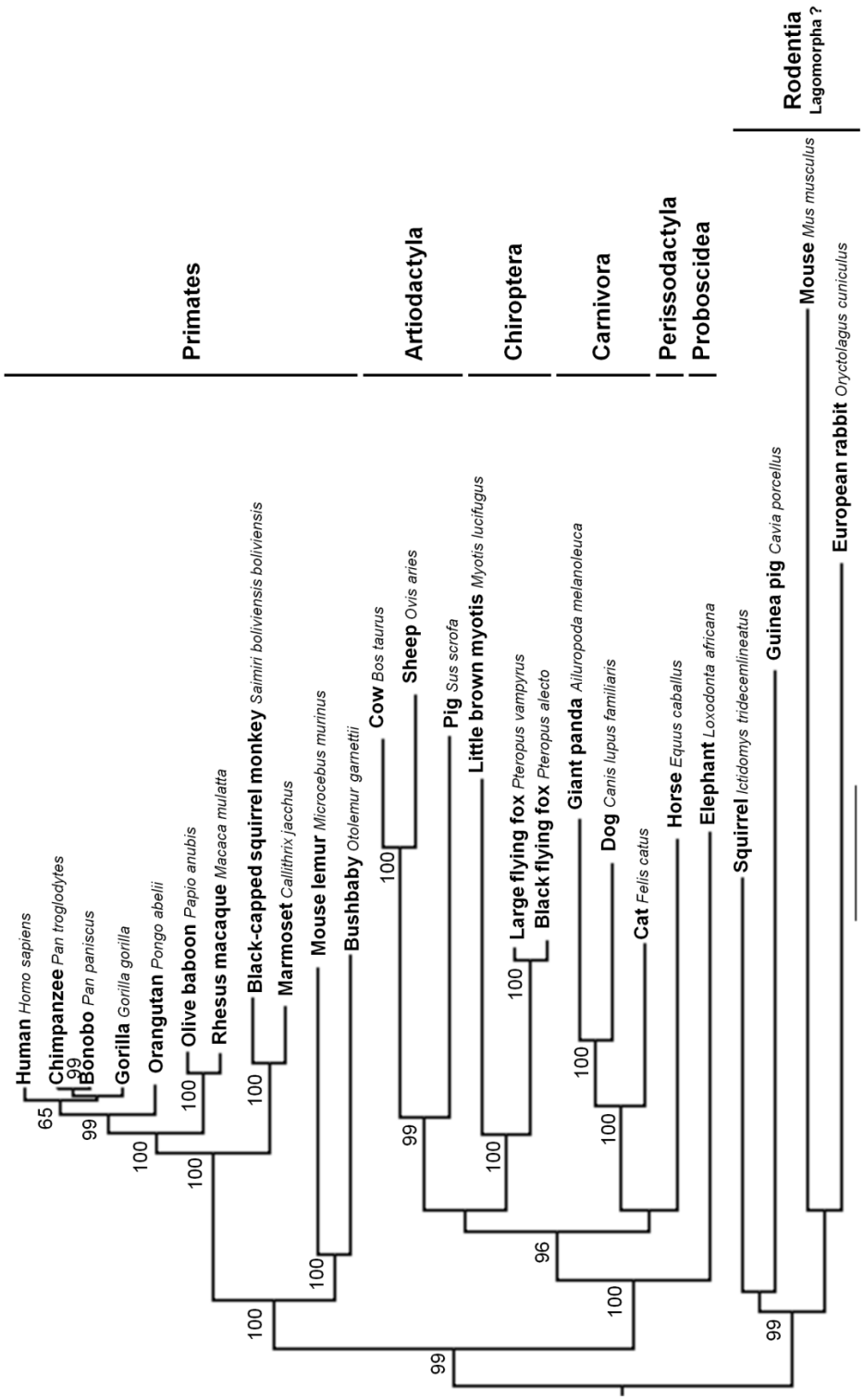


Figure 1 - Maximum likelihood (ML) phylogenetic tree of mammalian *RIG-I* gene used for codon-based ML analysis.
The GTR+G nucleotide substitution model was employed in mammalian *RIG-I* gene phylogenetic reconstruction. Bootstrap values >50 are indicated on the branches.

For the mammalian *MDA5* gene sequences alignment, a significant recombination breakpoint was detected (nucleotide 903; $p<0.01$). Therefore, two ML trees were reconstructed for the resulting segments, one for the first 903 nucleotides (Figure 2A) and another ML tree for the remaining 2211 nucleotides (Figure 2B). The gene phylogeny was also reconstructed for the whole alignment without testing recombination (Figure 2C) to compare its topology with the other two resulting trees. The monophyly of the eight taxonomic orders included in the *MDA5* alignment was roughly recovered, with the clear exception of Chiroptera in Figure 2A and Primates in Figure 2B.

Regarding the *LGP2* gene, no clear evidence of recombination was detected. The ML tree obtained (Figure 3) supported the monophyly of the nine mammalian orders collected for this gene.

4.3. Evidence of positive selection in the mammalian RIG-I-like receptors

All the molecular evolutionary analyses in this study were performed for both the complete nucleotide alignments (Supplementary Figure S1, Supplementary Figure S2 and Supplementary Figure S3) and for a trimmed version of the same genes to remove alignment gaps. Supplementary Figure S7 (*RIG-I* alignment), Supplementary Figure S8 (*MDA5* alignment) and Supplementary Figure S9 (*LGP2* alignment) correspond to the alignments where gaps present in all sequences, with the exception of one or two, have been removed, while gaps present in only one or two sequences were kept. We observed no significant differences in the results when using one or the other alignment for each gene (data not shown), but ultimately only the results from the trimmed version are presented here.

Evidence for positive selection on mammalian orthologous for *RIG-I* (Supplementary Figure S7), *MDA5* (Supplementary Figure S8) and *LGP2* (Supplementary Figure S9) genes was detected using PAML package [53, 54] site-specific models M1a versus M2a and M7 versus M8. These models test at the codon level whether a hypothesis that allows for positive selection (models M2a and M8) is a better fit to the data when compared to a null neutral hypothesis (models M1a and M7). Results on the likelihood ratio test (LRT) performed between the likelihood scores of the null neutral and alternative selection models for each gene is indicated in Table 1.

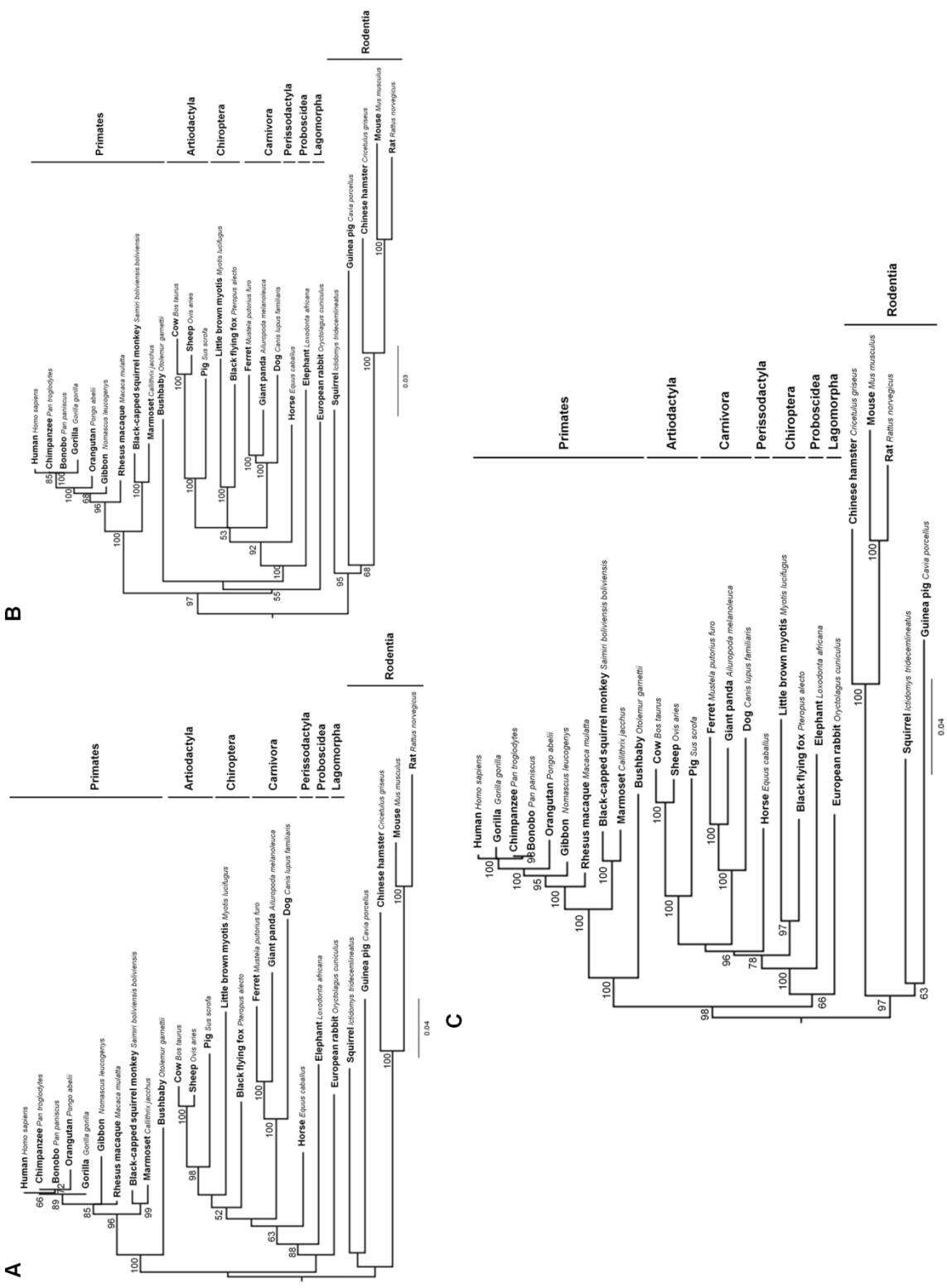

A
Human Homo sapiens
Chimpanzee Pan troglodytes
Bonobo Pan paniscus
Orangutan Pongo abelii
Gorilla Gorilla gorilla
Gibbon Nomascus leucogenys
Rhesus macaque Macaca mulatta
Black-capped squirrel monkey Saimiri boliviensis boliviensis
Marmoset Callithrix jacchus
Bushbaby Otolemur garnettii
Cow Bos taurus
Sheep Ovis aries
Pig Sus scrofa
Little brown myotis Myotis lucifugus
Black flying fox Pteropus alecto
Ferret Mustela putorius furo
Giant panda Ailuropoda melanoleuca
Dog Canis lupus familiaris
Horse Equus caballus
Elephant Loxodonta africana
European rabbit Oryctolagus cuniculus
Squirrel Ictidomys tridecemlineatus
Guinea pig Cavia porcellus
Chinese hamster Cricetulus griseus
Mouse Mus musculus
Rat Rattus norvegicus
Primates
Artiodactyla
Chiroptera
Carnivora
Perissodactyla
Proboscidea
Lagomorpha
Rodentia
0.04
B
Human Homo sapiens
Chimpanzee Pan troglodytes
Bonobo Pan paniscus
Gorilla Gorilla gorilla
Orangutan Pongo abelii
Gibbon Nomascus leucogenys
Rhesus macaque Macaca mulatta
Black-capped squirrel monkey Saimiri boliviensis boliviensis
Marmoset Callithrix jacchus
Bushbaby Otolemur garnettii
Cow Bos taurus
Sheep Ovis aries
Pig Sus scrofa
Little brown myotis Myotis lucifugus
Black flying fox Pteropus alecto
Ferret Mustela putorius furo
Giant panda Ailuropoda melanoleuca
Dog Canis lupus familiaris
Horse Equus caballus
Elephant Loxodonta africana
European rabbit Oryctolagus cuniculus
Squirrel Ictidomys tridecemlineatus
Guinea pig Cavia porcellus
Chinese hamster Cricetulus griseus
Mouse Mus musculus
Rat Rattus norvegicus
Primates
Artiodactyla
Chiroptera
Carnivora
Perissodactyla
Proboscidea
Lagomorpha
Rodentia
0.03
C
Human Homo sapiens
Gorilla Gorilla gorilla
Chimpanzee Pan troglodytes
Bonobo Pan paniscus
Orangutan Pongo abelii
Gibbon Nomascus leucogenys
Rhesus macaque Macaca mulatta
Black-capped squirrel monkey Saimiri boliviensis boliviensis
Marmoset Callithrix jacchus
Bushbaby Otolemur garnettii
Cow Bos taurus
Sheep Ovis aries
Pig Sus scrofa
Ferret Mustela putorius furo
Giant panda Ailuropoda melanoleuca
Dog Canis lupus familiaris
Horse Equus caballus
Little brown myotis Myotis lucifugus
Black flying fox Pteropus alecto
Elephant Loxodonta africana
European rabbit Oryctolagus cuniculus
Chinese hamster Cricetulus griseus
Mouse Mus musculus
Rat Rattus norvegicus
Squirrel Ictidomys tridecemlineatus
Guinea pig Cavia porcellus
Primates
Artiodactyla
Carnivora
Perissodactyla
Chiroptera
Proboscidea
Lagomorpha
Rodentia
0.04

Figure 2 - Maximum likelihood (ML) phylogenetic trees of mammalian *MDA5* gene used for codon-based ML analysis. When testing mammalian *MDA5* alignment for recombination, one significant breakpoint was detected at nucleotide position 903. (A) A phylogenetic tree was reconstructed for the first 903 nucleotides under the nucleotide substitution model TIM3+G. (B) A second ML tree was inferred for the remaining 2211 nucleotides and under the nucleotide substitution model TIM3+I+G. (C) A tree was also reconstructed for *MDA5* total alignment without recombination testing and under the nucleotide substitution model GTR+G. Bootstrap values >50 are indicated on the branches.

Models which allow for positive selection (M2a and M8) gave a significantly better fit to the data for both *RIG-I* and *LGP2* alignments, suggesting that at least some of the codons within each set of orthologous gene sequences are subject to positive selection [56]. Since a recombination breakpoint was detected on the *MDA5* alignment, each resulting segment (identified as 1st and 2nd segments) was tested individually for PAML package [53, 54] site-specific models. Although the comparison between the null neutral site model M1a and the selection site model M2a did not allow for rejection of the null hypothesis of neutral selection, the comparison between the more powerful pair of site-specific models M7 (neutral) and M8 (selection) yielded significant LRTs (Table 1).

The PARRIS method [58] implemented in the Datamonkey web server [49, 50] was also applied to each RLR trimmed gene alignment (Supplementary Figure S7, Supplementary Figure S8 and Supplementary Figure S9) to look for evidence of positive selection, but no selective pressures were detected in any of the three genes (Supplementary Table S2).

For each orthologous gene sequences alignment, the tree length parameter is indicated in Table 1. Higher values of tree length, i.e. the expected number of nucleotide substitutions per codon, correspond to higher sequence divergence [72, 73]. The tree length values registered for the three genes fell into an intermediate and realistic level of sequence divergence which confers power to the codon models indicated by the LRT scores and to the Bayes empirical Bayes (BEB) approach for site-specific inference of positive selection [72, 74].

Model M8 implemented in the PAML package [53, 54] and Datamonkey web server [49-50] SLAC, FEL, REL, MEME and FUBAR methods [62-64] were used to detect specific codons under selection in the three RLR genes. Based on the methodology adopted by other authors and in previous studies [43, 59, 60, 70], only codons identified by at least three of the six used methods are considered to be under positive selection (Supplementary Table S3). Since the breadth of species included in each alignment is wide, by applying several methods to detect codons under positive

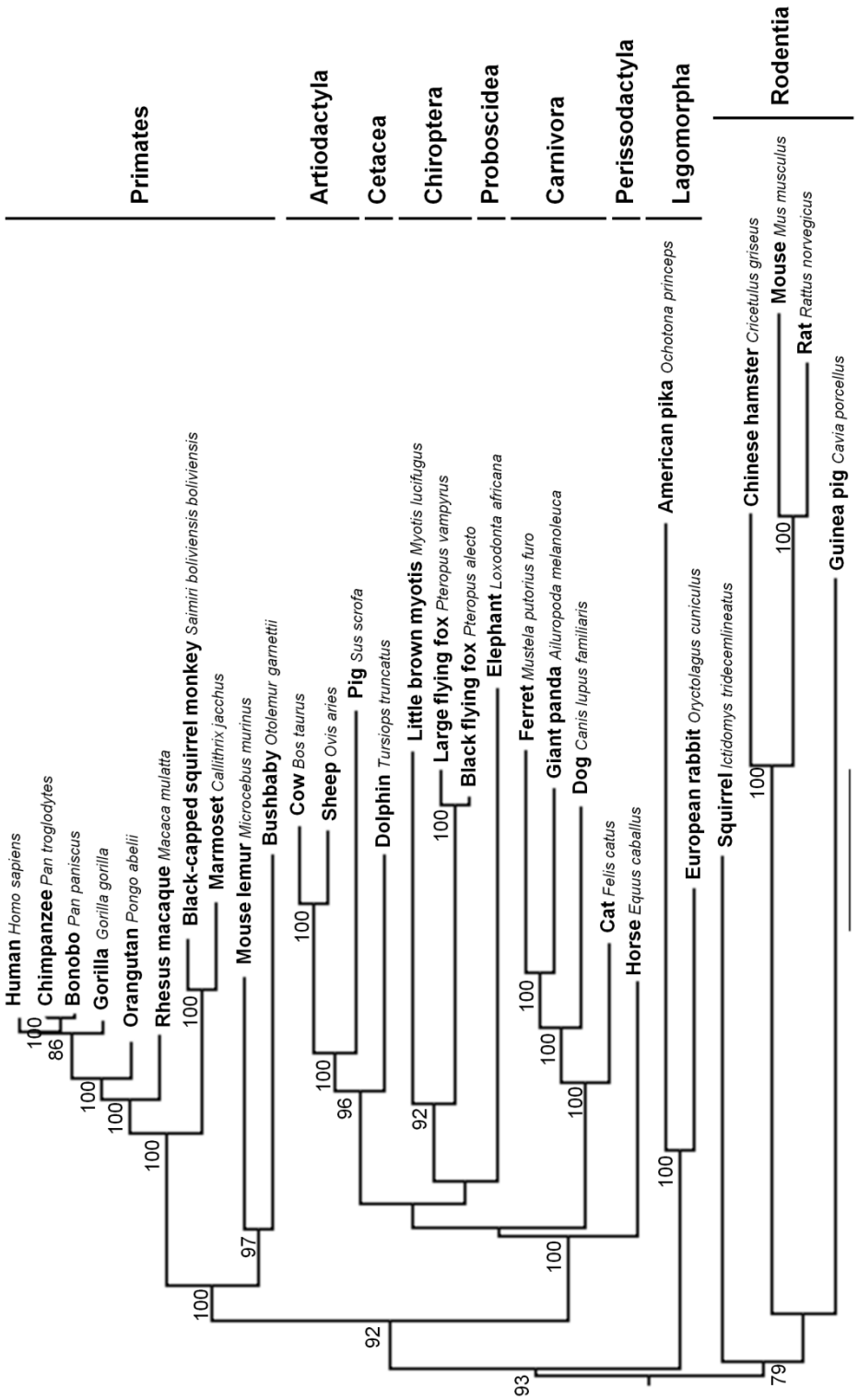


Figure 3 - Maximum likelihood (ML) phylogenetic tree of mammalian *LGP2* gene used for codon-based ML. The TPM2uf+I+G nucleotide substitution model was employed in mammalian *LGP2* gene phylogenetic reconstruction. Bootstrap values >50 are indicated on the branches.

selection and by overlapping the results, we should be decreasing the incidence of false positives. A total of sixteen codons for RIG-I (Figure 4), twenty for MDA5 (Figure 5) and ten for LGP2 (Figure 6) were identified as candidate codons under selective pressure. Regarding their location in each corresponding protein, the greatest number of these codons are located in protein functional domains, more specifically, eleven out of the sixteen RIG-I codons (~ 69%), ten out of the twenty MDA5 codons (50%) and seven out of the ten LGP2 codons (70%). To estimate the percentage of positively selected codons in the three proteins, we used human deduced protein sequences as a reference. Human LGP2 exhibited 1.47% (10/678) of codons under positive selection. Higher values were obtained for human MDA5 and RIG-I, 1.95% (20/1025) and 1.73% (16/925) of codons under selective pressure, respectively.

Table 1 - RIG-I-like receptors likelihood ratio tests (LRTs) for PAML M1a, M2a, M7 and M8 site models

PAML site models	lnL_{null}[a]	$lnL_{alternative}$[a]	$2\Delta lnL$[b]	*p*-Value	Tree length
RIG-I					4.60
M1a $_{(nearly\ neutral)}$ vs. M2a $_{(selection)}$	-21065.76	-21045.84	39.84	p<0.0001	
M7 $_{(neutral,\ beta)}$ vs. M8 $_{(selection,\ beta\ \&\ \omega)}$	-21053.14	-21016.64	73.00	p<0.0001	
MDA5_1stSegment					6.03
M1a $_{(nearly\ neutral)}$ vs. M2a $_{(selection)}$	-7981.98	-7981.98	0.00	n.s.	
M7 $_{(neutral,\ beta)}$ vs. M8 $_{(selection,\ beta\ \&\ \omega)}$	-7970.70	-7966.35	8.70	P<0.02	
MDA5_2ndSegment					3.80
M1a $_{(nearly\ neutral)}$ vs. M2a $_{(selection)}$	-14314.15	-14312.61	3.08	n.s.	
M7 $_{(neutral,\ beta)}$ vs. M8 $_{(selection,\ beta\ \&\ \omega)}$	-14284.53	-14272.94	23.18	p<0.0001	
LGP2					6.38
M1a $_{(nearly\ neutral)}$ vs. M2a $_{(selection)}$	-18838.31	-18830.88	14.86	p<0.001	
M7 $_{(neutral,\ beta)}$ vs. M8 $_{(selection,\ beta\ \&\ \omega)}$	-18693.92	-18674.22	39.40	p<0.0001	

[a] lnL: log-likelihood scores.
[b] 2ΔlnL: likelihood ratio test (LRT) to detect positive selection.
n.s. – non-significant.

To detect signatures of episodic positive selection in specific lineages of each RLR orthologous gene sequences alignment we performed two branch-site model analyses. These models allow the selective pressure indicated by the nonsynonymous to synonymous substitution rate ratio ω (d_N/d_S) to vary both across sites in the gene and across lineages on the tree [75]. Since no biological hypothesis existed to specify *a priori* branches to be examined for positive selection, the branch-site model A

implemented in the PAML package [53, 54, 65] was applied to all species branches on each RLR gene phylogenetic tree. The LRT performed for each branch was significant for 2ΔlnL > 3.84 [54, 65]. Our analyses suggest that nine species branches in *RIG-I* are under selective pressure (Table 2 and Figure 4B). Branch-site model A was applied to the two *MDA5* trees resultant from recombination testing and, for each tree, positive selection has operated only in two species branches (Table 3 and Figure 5B). For *LGP2*, a total of six species branches had significant LRTs corresponding to candidate lineages under selection (Table 4 and Figure 6B). Some of the species branches recognized by the branch-site model A were also identified by the branch-site REL method [66] (Table 5) available in the Datamonkey web server [49, 50]. For both *RIG-I* and *MDA5*, two species branches were simultaneously identified by the two methods, consisting in dog (Calu) and European rabbit (Orcu) branches for *RIG-I* (Figure 4B) and giant panda (Aime) and Guinea pig (Capo) branches for *MDA5* (Figure 5B). Only one *LGP2* species branch, corresponding to the giant panda (Aime), was simultaneously identified by the branch-site model A and the branch-site REL method (Figure 6B).

5. Discussion

In a human population genetics context, the first study on RLRs evolutionary history and selective footprints has been recently published [76]. Nevertheless, to the best of our knowledge, our study is the first that searches for selective pressure acting on mammalian orthologous of the three RLRs and, in fact, we provide strong evidence of positive selection as well as identify a significant number of codons under probable selective pressures for RIG-I, MDA5 and LGP2. Furthermore, our results on the RIG-I RD in specific hosts suggest that certain viruses might be exerting long-term selective pressures on this gene.

TLRs adaptive evolution has been the most extensively characterized of all the PRRs in several animal groups, such as echinoderms [77], birds [78] and different mammals [43, 59, 79-82]. Studies on known viral-recognition TLRs (TLR3, 7, 8 and 9) of closely related animal groups, like birds [78], or within species, like humans and chimpanzees [59], demonstrated that this class of PRRs exhibits a background of strong purifying selection to keep their functional integrity, albeit in the birds study [78] significant instances of positive selection acting on a few amino acid sites were identified. Nevertheless, when different ML codon-based methods were applied to detect evidence of acting positive selection in broader groups where a great number of species are included, like primates [59] and mammals [43], most of the viral TLRs exhibited strong evidence of positive selection and specific codons with a high

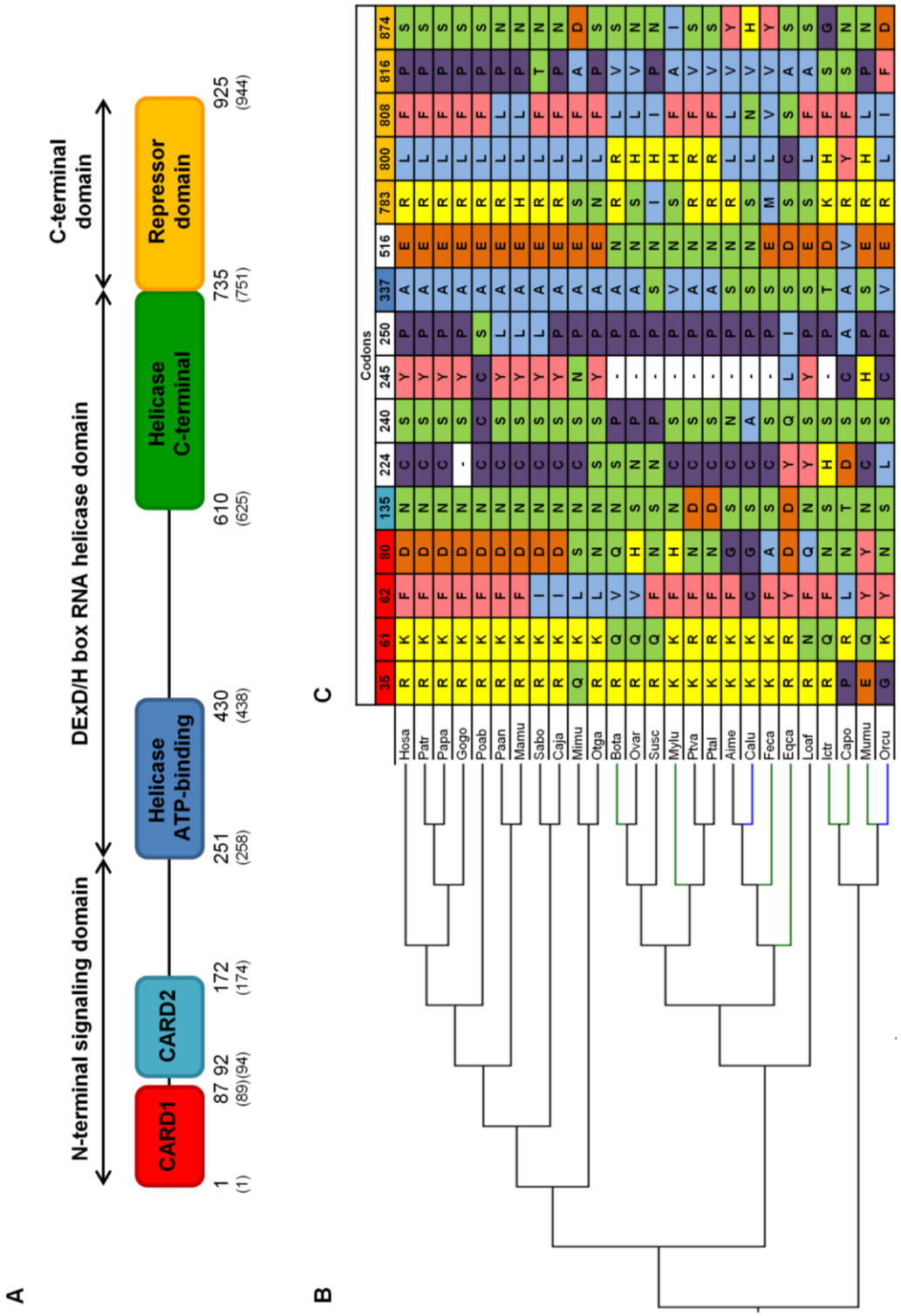

A
N-terminal signaling domain
DExD/H box RNA helicase domain
C-terminal domain
CARD1
CARD2
Helicase ATP-binding
Helicase C-terminal
Repressor domain
1 (1)
87 92 (89)(94)
172 (174)
251 (258)
430 (438)
610 (625)
735 (751)
925 (944)
B
C
Codons
35 61 62 80 135 224 240 245 250 337 516 783 800 808 816 874
Hosa R K F D N C S Y P A E R L F P S
Patr R K F D N C S Y P A E R L F P S
Papa R K F D N C S Y P A E R L F P S
Gogo R K F D N - S Y P A E R L F P S
Poab R K F D N C C C S A E R L F P S
Paan R K F D N C S Y L A E R L L P N
Mamu R K F D N C S Y L A E H L L P N
Sabo R K I D N C S Y P A E R L F T N
Caja R K I D N C S Y P A E R L F P N
Mimu Q K L S N C S N P A E S L F A D
Otga R K L N N S S Y P A E N L F P S
Bota R Q V Q S P P - P A N R L L V S
Ovar R Q V H S N P - P A N S H L V N
Susc R Q F N S N P - P S N I H I P N
Mylu K K F H N C S - P V N S H F A I
Ptva K R F N D C S - P A N R R F V S
Ptal K R F N D C S - P A N R R F V S
Aime K K F G S C N - P S N R R L V Y
Calu K K C G S C A - P S N S L N V H
Feca K K F A S C S - P S E M L V V Y
Eqca R R Y D D Y Q L I S D S C S A S
Loaf R N F Q N Y S Y P S E S L F A S
Ictr R Q F N S H S - P T D K H F S G
Capo P R L N T D S C A A V R Y F S N
Mumu E Q Y Y N C S H P S E R H L P N
Orcu G K Y N S L S C P V E R L I F D

Figure 4 - Structural representation and identification of positively-selected branches and codons in mammalian RIG-I. (A) Based on human protein structure, the key domains of RIG-I (http://www.uniprot.org/uniprot/O95786) and the corresponding boundaries are schematically represented. Also, the human domain boundaries while in the mammalian RIG-I deduced protein sequences alignment (Supplementary Figure S4) are shown in brackets. (B) Cladogram of 26 mammalian *RIG-I* genes collected from Ensembl and NCBI databases. Branch-site analyses were performed to identify specific branches under episodic positive selection. Branches with statistically significant likelihood ratio tests (LRTs) when performing PAML branch-site model A (Table 2) are colored in green; branches simultaneously identified by PAML branch-site model A and Hyphy branch-site REL method (Table 5) are colored in blue. (C) Positively-selected codons are exhibited in the table and numbered according to the mammalian RIG-I deduced protein sequences alignment (Supplementary Figure S4). Symbol "-" represents a deletion. Colors on the codon numbering row correspond to the RIG-I domain with the same color in the protein structural representation (A). The background colors on the identified sites match different amino acid properties: polar positive (yellow), polar negative (orange), polar neutral (green), non-polar neutral (purple), non-polar aliphatic (blue) and non-polar aromatic (pink). The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Paan – Olive baboon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mylu – Little brown myotis; Ptva – Large flying fox; Ptal – Black flying fox; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Loaf – Elephant; Ictr – Squirrel; Capo – Guinea pig; Mumu – Mouse; Orcu – European rabbit.

probability of being under selection were identified. Similarly, in our study the codon-based analyses strongly support that the three RLR genes, *RIG-I*, *MDA5* and *LGP2*, have all been subject to long-term selective pressures during mammalian evolution. Also, we applied several methods that identified specific RLR codons with a high probability of being under selection, which may directly perturb downstream immune responses in a particular host infected by a viral pathogen.

One of the major concerns when using large scale divergent species to infer positive selection acting on a set of orthologous genes and across lineages on the phylogenetic tree is the effect of saturation in synonymous substitutions, since they may saturate quickly as sequences diverge [83-84]. As codon models consider both synonymous and nonsynonymous substitutions, the saturation of the first could cloud the information provided by nonsynonymous substitutions. Nevertheless, the sequence divergence in our study, inferred through RLRs tree length values, fit into intermediate and realistic levels that should confer power to the LRT used to compare nested codon-models and robustness to the branch-site models, and to the BEB approach for codon-specific detection of positive selection [72-74]. Also, in this study the mammalian species collected for each of the three RLR genes were nearly the same, thus this host species spectrum should not influence the codon-based analyses and our observations when comparing the level of selective pressure between genes.

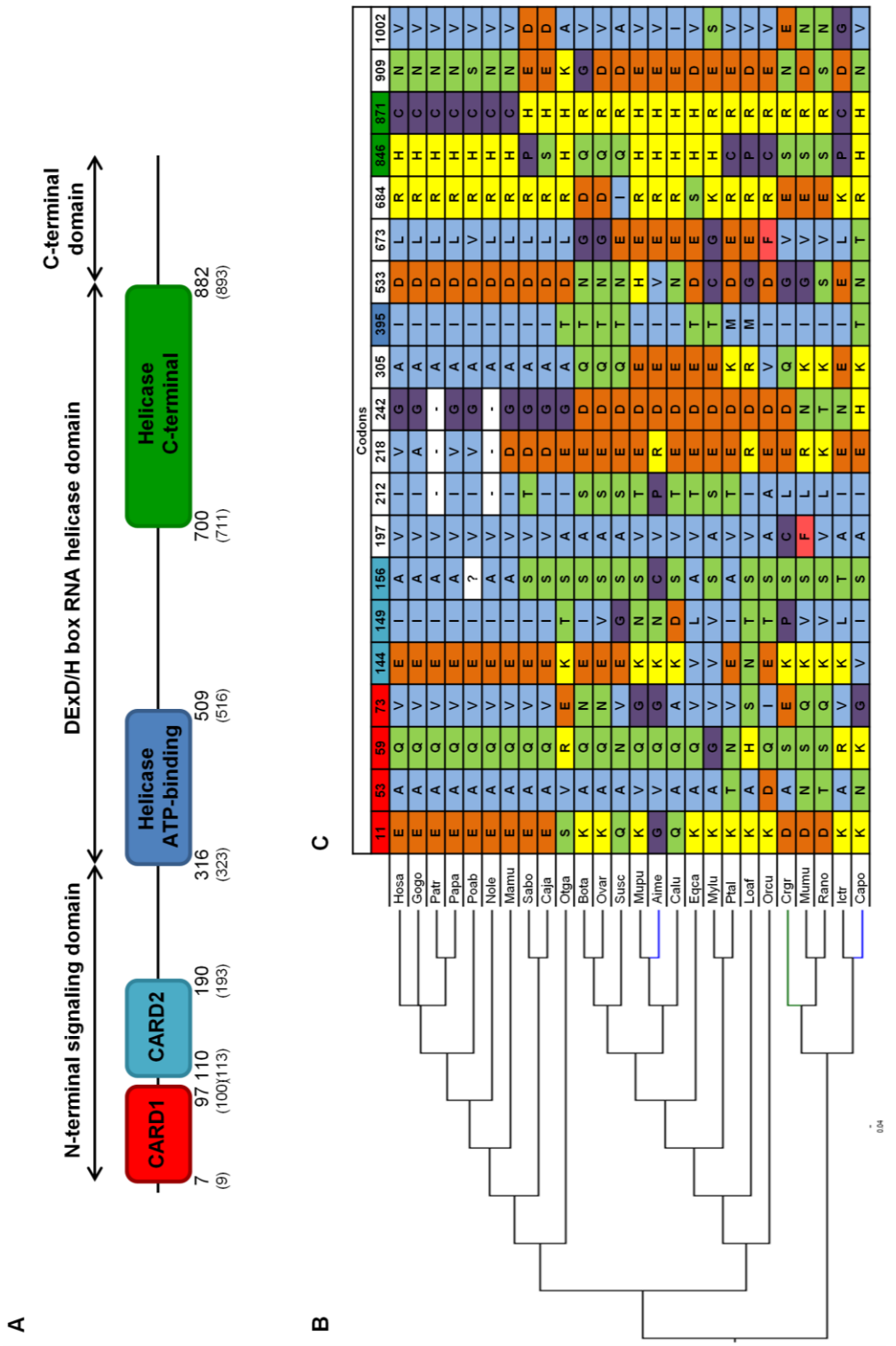

A
N-terminal signaling domain
DExD/H box RNA helicase domain
C-terminal domain
CARD1
CARD2
Helicase ATP-binding
Helicase C-terminal
7 (9)
97 110 (100)(113)
190 (193)
316 (323)
509 (516)
700 (711)
882 (893)
B
C
Codons
11 53 59 73 144 149 156 197 212 218 242 305 395 533 673 684 846 871 909 1002
Hosa
Gogo
Patr
Papa
Poab
Nole
Mamu
Sabo
Caja
Otga
Bota
Ovar
Susc
Mupu
Aime
Calu
Eqca
Mylu
Ptal
Loaf
Orcu
Crgr
Mumu
Rano
Ictr
Capo

Figure 5 - Structural representation and identification of positively-selected branches and codons in mammalian MDA5. (A) Based on the human protein structure, the key domains of MDA5 (http://www.uniprot.org/uniprot/Q9BYX4) and the corresponding boundaries are schematically represented. Also, the human domain boundaries while in the mammalian MDA5 deduced protein sequences alignment (Supplementary Figure S5) are shown in brackets. (B) Cladogram of 26 mammalian *MDA5* genes collected from Ensembl and NCBI databases. Branch-site analyses were performed to identify specific branches under episodic positive selection. Branches with statistically significant likelihood ratio tests (LRTs) when performing PAML branch-site model A (Table 3) are colored in green; branches simultaneously identified by PAML branch-site model A and Hyphy branch-site REL method (Table 5) are colored in blue. (C) Positively-selected codons are exhibited in the table and numbered according to the mammalian MDA5 deduced protein sequences alignment (Supplementary Figure S5). Symbol "?" represents an undetermined amino acid, while "-" symbolizes a deletion. Colors on the codon numbering row correspond to the MDA5 domain with the same color in the protein structural representation (A). The background colors on the identified sites match different amino acid properties: polar positive (yellow), polar negative (orange), polar neutral (green), non-polar neutral (purple), non-polar aliphatic (blue) and non-polar aromatic (pink). The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Gogo – Gorilla; Patr – Chimpanzee; Papa – Bonobo; Poab – Orangutan; Nole – Gibbon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Eqca – Horse; Mylu – Little brown myotis; Ptal – Black flying fox; Loaf – Elephant; Orcu – European rabbit; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Ictr – Squirrel; Capo – Guinea pig.

In our study, mammalian MDA5 showed the highest number and percentage of positively selected codons. Nonetheless, the percentage of MDA5 codons under selection located in the known protein functional domains was the lowest. This should reflect the imposition of functional and structural constraints in MDA5 defined domains. On the other hand, we observed that LGP2 is apparently less prone to evolutionary change with the lowest number and percentage of codons under selective pressures. For RIG-I, the greatest number of codons identified as candidates under selective pressures were located in known protein functional domains, which might reveal the pressure imposed by the great number of viruses recognized by this RLR [reviewed in 13, 14]. Vasseur and colleagues [76] came to different conclusions in their study, once they were focused on intra-species (human populations) polymorphisms and on the comparison of nonsynonymous to synonymous rates ratio ω (d_N/d_S) between human and chimpanzee lineages for the three RLR genes. RIG-I exhibited a stronger evolutionary constraint [76], as attested by its low levels of nucleotide diversity, population differentiation and low tolerance of amino acid-altering variation. It also exhibited a dramatic decay in the ω ratio when compared to the other two RLRs [76]. This is the expected outcome in evolutionary studies when using closely related species, or genetic information for population of the same species, which result in a

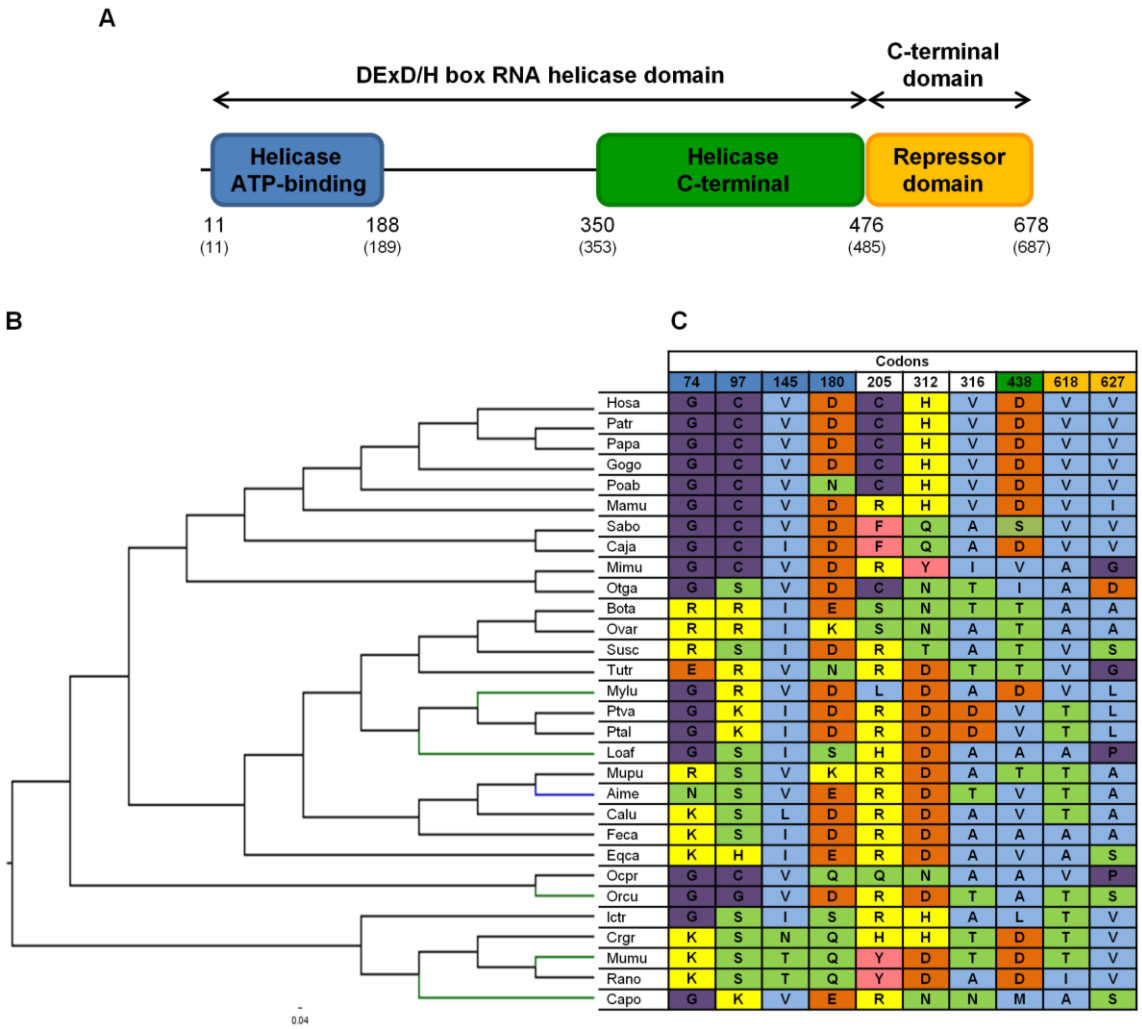


Figure 6 - Structural representation and identification of positively-selected branches and codons in mammalian LGP2. (A) Based on human protein structure, the key domains of LGP2 (http://www.uniprot.org/uniprot/Q96C10) and the corresponding boundaries are schematically represented. Also, the human domain boundaries while in the mammalian LGP2 deduced protein sequences alignment (Supplementary Figure S6) are shown in brackets. (B) Cladogram of 30 mammalian *LGP2* genes collected from Ensembl and NCBI databases. Branch-site analyses were performed to identify specific branches under episodic positive selection. Branches with statistically significant likelihood ratio tests (LRTs) when performing PAML branch-site model A (Table 4) are colored in green; branch colored in blue has been simultaneously identified by PAML branch-site model A and Hyphy branch-site REL method (Table 5). (C) Positively-selected codons are exhibited in the table and numbered according to the mammalian LGP2 deduced protein sequences alignment (Supplementary Figure S6). Colors on the codon numbering row correspond to the LGP2 domain with the same color in the protein structural representation (A). The background colors on the identified codons match different amino acid properties: polar positive (yellow), polar negative (orange), polar neutral (green), non-polar neutral (purple), non-polar aliphatic (blue) and non-polar aromatic (pink). The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Tutr – Dolphin; Mylu – Little brown myotis; Ptva – Large flying fox; Ptal – Black flying fox; Loaf – Elephant; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Ocpr – American pika; Orcu – European rabbit; Ictr – Squirrel; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Capo – Guinea pig.

Table 2 - PAML branch-site model A analysis to identify branches under episodic positive selection in *RIG-I* phylogenetic tree

Foreground branches[a]	Parameters under null model	lnL[b] (null)	Parameters under alternative model	lnL[b] (alternative)	2ΔlnL[c]	*p*-Value	Positively selected sites[d]
Bota	p_0= 0.536 p_1= 0.294 p_{2a}= 0.110 p_{2b}= 0.060 ω_0= 0.090 ω_1= 1 ω_2= 1	-21065.614	p_0= 0.637 p_1= 0.349 p_{2a}= 0.009 p_{2b}= 0.005 ω_0= 0.090 ω_1= 1 ω_2= **18.749**	-21063.247	**4.734**	<0.05	**655**F (0.907) **656**Q (0.996)
Calu	p_0= 0.631 p_1= 0.343 p_{2a}= 0.016 p_{2b}= 0.009 ω_0= 0.092 ω_1= 1 ω_2= 1	-21069.596	p_0= 0.645 p_1= 0.350 p_{2a}= 0.003 p_{2b}= 0.002 ω_0= 0.093 ω_1= 1 ω_2= **70.492**	-21065.810	**7.572**	<0.01	none
Capo	p_0= 0.617 p_1= 0.335 p_{2a}= 0.031 p_{2b}= 0.017 ω_0= 0.090 ω_1= 1 ω_2= 1	-21068.600	p_0= 0.641 p_1= 0.347 p_{2a}= 0.008 p_{2b}= 0.004 ω_0= 0.092 ω_1= 1 ω_2= [e]	-21065.132	**6.936**	<0.01	none
Eqca	p_0= 0.648 p_1= 0.352 p_{2a}= 0 p_{2b}= 0 ω_0= 0.093 ω_1= 1 ω_2= 1	-21069.711	p_0= 0.643 p_1= 0.351 p_{2a}= 0.004 p_{2b}= 0.002 ω_0= 0.092 ω_1= 1 ω_2= **47.106**	-21067.540	**4.342**	<0.05	none
Feca	p_0= 0.613 p_1= 0.333 p_{2a}= 0.035 p_{2b}= 0.019 ω_0= 0.092 ω_1= 1 ω_2= 1	-21068.674	p_0= 0.643 p_1= 0.351 p_{2a}= 0.004 p_{2b}= 0.002 ω_0= 0.092 ω_1= 1 ω_2= **265.180**	-21064.248	**8.852**	<0.005	**44**W (0.943)
Ictr	p_0= 0.611 p_1= 0.331 p_{2a}= 0.038 p_{2b}= 0.020 ω_0= 0.091 ω_1= 1 ω_2= 1	-21068.446	p_0= 0.644 p_1= 0.348 p_{2a}= 0.005 p_{2b}= 0.003 ω_0= 0.093 ω_1= 1 ω_2= **16.527**	-21066.281	**4.330**	<0.05	none
Mumu	p_0= 0.648 p_1= 0.352 p_{2a}= 0 p_{2b}= 0 ω_0= 0.093 ω_1= 1 ω_2= 1	-21069.711	p_0= 0.646 p_1= 0.346 p_{2a}= 0.005 p_{2b}= 0.003 ω_0= 0.093 ω_1= 1 ω_2= [f]	-21067.171	**5.080**	<0.025	none
Mylu	p_0= 0.648 p_1= 0.352 p_{2a}= 0 p_{2b}= 0 ω_0= 0.093 ω_1= 1 ω_2= 1	-21069.711	p_0= 0.645 p_1= 0.351 p_{2a}= 0.002 p_{2b}= 0.001 ω_0= 0.093 ω_1= 1 ω_2= **172.548**	-21066.025	**7.372**	<0.01	none
Orcu	p_0= 0.599 p_1= 0.320 p_{2a}= 0.053 p_{2b}= 0.028 ω_0= 0.089 ω_1= 1 ω_2= 1	-21064.127	p_0= 0.637 p_1= 0.336 p_{2a}= 0.017 p_{2b}= 0.009 ω_0= 0.092 ω_1= 1 ω_2= **38.747**	-21046.263	**35.728**	<0.001	**849**T (0.989) **851**C (0.997) **854**S (0.958) **857**H (0.988) **861**G (0.982) **898**V (0.989)

[a] Species names on the foreground branches: Bota – Cow; Calu – Dog; Capo – Guinea pig; Eqca – Horse; Feca – Cat; Ictr – Squirrel; Mumu – Mouse; Mylu – Little brown myotis; Orcu – European rabbit.
[b] lnL: log-likelihood scores.
[c] 2ΔlnL: likelihood ratio test (LRT) to detect positive selection.
[d] Positively selected sites: posterior probabilities >0.90 in the BEB (Bayes empirical Bayes) analyses.
[e] ω_2 parameter varied for different (ω) values: ω_2(2) = 242.957; ω_2(3) = 340.801; ω_2(4) = 982.380.
[f] ω_2 parameter varied for different (ω) values: ω_2(2) = 131.879; ω_2(3) = 246.814; ω_2(4) = 289.634.

Table 3 - PAML branch-site model A analysis to identify branches under episodic positive selection in *MDA5* phylogenetic trees

Foreground branches [a]	Parameters under null model	lnL[b] (null)	Parameters under alternative model	lnL[b] (alternative)	2ΔlnL[c]	*p*-Value	Positively selected sites[d]
1st Segment							
Aime	p_0= 0.506 p_1 = 0.377 p_{2a}= 0.067 p_{2b} = 0.050 ω_0= 0.086 ω_1 = 1 ω_2 = 1	-7981.227	p_0= 0.559 p_1 = 0.418 p_{2a} = 0.013 p_{2b} = 0.010 ω_0 = 0.086 ω_1 = 1 ω_2 = **27.801**	-7976.648	**9.158**	<0.005	**295**P (0.995)
Capo	p_0= 0.488 p_1 = 0.364 p_{2a} = 0.085 p_{2b} = 0.063 ω_0= 0.077 ω_1 = 1 ω_2 = 1	-7973.674	p_0= 0.543 p_1 = 0.409 p_{2a} = 0.027 p_{2b} = 0.021 ω_0 = 0.078 ω_1 = 1 ω_2 = **27.756**	-7964.534	**18.280**	<0.001	**281**K (0.972) **284**F (0.993) **291**P (0.996) **293**L (0.989) **297**I (0.998)
2nd Segment							
Capo	p_0= 0.727 p_1 = 0.205 p_{2a} = 0.053 p_{2b} = 0.015 ω_0= 0.070 ω_1 = 1 ω_2 = 1	-14272.950	p_0= 0.759 p_1 = 0.209 p_{2a} = 0.025 p_{2b} = 0.007 ω_0 = 0.073 ω_1 = 1 ω_2 = **998.992**	-14255.172	**35.556**	<0.001	**358**G (0.958) **593**S (1) **594**S (1) **595**L (0.990)
Crgr	p_0= 0.778 p_1= 0.222 p_{2a}= 0 p_{2b}= 0 ω_0= 0.074 ω_1= 1 ω_2= 1	-14281.640	p_0= 0.778 p_1 = 0.219 p_{2a} = 0.003 p_{2b} = 0.001 ω_0 = 0.075 ω_1 = 1 ω_2 = [e]	-14277.837	**7.606**	<0.01	none

[a] Species names on the foreground branches: Aime – Giant panda; Capo – Guinea pig; Crgr – Chinese hamster.
[b] lnL: log-likelihood scores.
[c] 2ΔlnL: likelihood ratio test (LRT) to detect positive selection.
[d] Positively selected sites: posterior probabilities >0.90 in the BEB (Bayes empirical Bayes) analyses.
[e] ω_2 parameter varied for different (ω) values: ω_2 (2) = 999.000; ω_2 (3) = 832.570; ω_2 (4) = 681.973.

Table 4 - PAML branch-site model A analysis to identify branches under episodic positive selection in *LGP2* phylogenetic tree

Foreground branches [a]	Parameters under null model	lnL[b] (null)	Parameters under alternative model	lnL[b] (alternative)	2ΔlnL[c]	*p*-Value	Positively selected sites[d]
Aime	p_0= 0.729 p_1 = 0.182 p_{2a}= 0.070 p_{2b} = 0.018 ω_0= 0.082 ω_1 = 1 ω_2 = 1	-18816.537	p_0= 0.794 p_1 = 0.199 p_{2a}= 0.005 p_{2b} = 0.001 ω_0= 0.082 ω_1 = 1 ω_2= **86.683**	-18810.272	**12.530**	<0.001	**227**L (0.913) **519**C (0.999) **520**N (0.995)
Capo	p_0= 0.794 p_1 = 0.198 p_{2a}= 0.006 p_{2b} = 0.002 ω_0= 0.083 ω_1 = 1 ω_2 = 1	-18821.176	p_0= 0.795 p_1 = 0.197 p_{2a}= 0.006 p_{2b} = 0.001 ω_0= 0.083 ω_1 = 1 ω_2= **12.697**	-18819.037	**4.278**	<0.05	**176**N (0.929) **267**S (0.979)
Loaf	p_0= 0.787 p_1 = 0.196 p_{2a}= 0.014 p_{2b} = 0.003 ω_0= 0.083 ω_1 = 1 ω_2 = 1	-18820.429	p_0= 0.792 p_1 = 0.196 p_{2a}= 0.009 p_{2b} = 0.002 ω_0= 0.083 ω_1 = 1 ω_2= **423.742**	-18812.759	**15.340**	<0.001	**421**S (0.996) **569**S (0.985)
Mimu	p_0= 0.801 p_1 = 0.199 p_{2a}= 0 p_{2b} = 0 ω_0= 0.083 ω_1 = 1 ω_2 = 1	-18821.530	p_0= 0.797 p_1 = 0.198 p_{2a}= 0.004 p_{2b} = 0.001 ω_0= 0.083 ω_1 = 1 ω_2= **775.269**	-18819.489	**4.082**	<0.05	none
Mylu	p_0= 0.769 p_1 = 0.190 p_{2a}= 0.033 p_{2b} = 0.008 ω_0= 0.082 ω_1 = 1 ω_2 = 1	-18819.788	p_0= 0.795 p_1 = 0.193 p_{2a}= 0.009 p_{2b} = 0.002 ω_0= 0.083 ω_1 = 1 ω_2= **11.149**	-18817.547	**4.482**	<0.05	**646**A (0.980)
Orcu	p_0= 0.779 p_1 = 0.193 p_{2a}= 0.022 p_{2b} = 0.006 ω_0= 0.083 ω_1 = 1 ω_2 = 1	-18820.688	p_0= 0.797 p_1 = 0.197 p_{2a}= 0.005 p_{2b} = 0.001 ω_0= 0.083 ω_1 = 1 ω_2= **63.328**	-18817.321	**6.734**	<0.01	**605**P (0.961)

[a] Species names on the foreground branches: Aime – Giant panda; Capo – Guinea pig; Loaf – Elephant; Mimu – Mouse lemur; Mylu – Little brown myotis; Orcu – European rabbit.
[b] lnL: log-likelihood scores.
[c] 2ΔlnL: likelihood ratio test (LRT) to detect positive selection.
[d] Positively selected sites: posterior probabilities >0.90 in the BEB (Bayes empirical Bayes) analyses.

background of strong purifying selection to keep the protein functional integrity. In the same study [76], the strongest signatures of positive selection were found in MDA5 and LGP2 by exhibiting higher *ω* ratios than RIG-I. Besides, MDA5 and LGP2 also appear to have evolved adaptively in specific human populations, presenting a great number of nonsynonymous mutations in both helicase and C-terminal domains [76].

Table 5 - Hyphy branch-site REL analysis to identify RIG-I-like receptor species branches subject to episodic diversifying selection

Branch[a]	ω^{+} [b]	*p*-Value
RIG-I		
Calu	57.89	0.032
Orcu	30.23	< 0.0001
MDA5		
Aime	11.46	0.004
Capo	3334.49	< 0.0001
LGP2		
Aime	196.66	0.043

[a] Species names on the branches: Aime – Giant panda; Calu – Dog; Capo – Guinea pig; Orcu – European rabbit.
[b] ω^{+}: strength and extent (proportion of sites) of selection along each branch.

RIG-I and MDA5 contain two N-terminal CARDs [10, 17]. The interaction of these domains with an adaptor protein named IPS-1 (also known as MAVS, VISA or CARDIF) is a crucial process to activate a wide range of downstream response factors, including type I IFNs and other essential anti-viral proteins to induce intracellular immune responses [85]. Interestingly, in our study, the CARDs of both RIG-I and MDA5 concentrated a large number of the deduced codons under selection. Some of these are radical in terms of their physicochemical properties changes across mammalian species (Figure 4 and Figure 5), strengthening the case for positive selection. Since the two CARDs are fundamental for downstream RIG-I and MDA5 signaling, which implies functional constraints, the observed variability across species can be perceived as a great structural plasticity for mammalian CARDs.

The helicase domain in the RLR family is generally described as exhibiting affinity for dsRNA [86]. The existence of six highly conserved sequence motifs within this domain is a characteristic of the helicase superfamily 2 which integrates DExD/H box RNA helicases. Also, different aspects of helicase functions have been assigned to specific motifs [87]. Bamming and Horvath [11] compared the amino acid sequences of the three human RLR helicase domains with the established consensus sequences of helicase families elements and, despite slight differences, the sequences in individual

motifs are highly conserved within RIG-I, MDA5 and LGP2. Indeed, in our study the six helicase motifs of the three proteins were evolutionary conserved (data not shown) in the mammalian species collected. Minor alterations occur in some species, but the extent of those differences concerning the involvement in substrate interaction, signal transduction and/or the whole anti-viral response profile, is difficult to predict.

RIG-I RD is responsible for recognizing and binding to its RNA substrates in a 5'-triphosphate (5'-ppp)-dependent manner. Besides, binding studies clearly established that the pppRNA binding site resides within the RD [14, 26, 88]. The function described for RIG-I RD makes our current results worthy of note, since the RD is the RIG-I domain that exhibits the strongest evidence of trans-acting selective pressures (Figure 4). Whether these differences play a role in RIG-I activation after binding to the RNAs from different viral pathogens that infect distinct mammalian hosts is a complex question. Nevertheless, we can assume that the RD variability in mammals is related to the fact that RIG-I senses a large variety of viruses [reviewed in 13, 14].

The performance of branch-site models in our study imposes a careful interpretation of data, since only one representative element of each species was included. Still, some branches of the three RLR phylogenetic trees exhibited evidence of positive selection. The two species under episodic positive selection on *RIG-I* phylogenetic tree, the domestic dog and the European rabbit, are susceptible hosts of two viruses recognized by RIG-I, rabies virus (Rhabdoviridae family) and myxoma virus (Poxviridae family), respectively [23, 39]. Such results suggest that these lethal pathogens, and possibly other re-occurring viral infections in these specific hosts, might be exerting long-term selective pressures on the susceptible host *RIG-I* gene. Therefore, the changes on RIG-I sequences across species, with special focus on the RD as suggested above, should be the result of a co-evolutionary process between species-specific infecting viruses and this host RNA sensor protein.

By detecting the extension of acting positive selection on mammalian RLRs, this study provides further insights into their biological functions in host defense against viral pathogens in general. Differences in these genes across mammalian species may consequently impact downstream immune responses and, as a result, contribute to the species-specific resistance/susceptibility profiles against many diverse viral pathogens.

6. Supplementary material

Supplementary material is appended to the present document by order of appearance in the main text.

Supplementary Table S1 - List of mammalian species and genes accession numbers used in this study.

Supplementary Figure S1 - Mammalian *RIG-I* nucleotide coding region sequences alignment.

RIG-I nucleotide coding region sequences for twenty-six mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol "." represents the same nucleotide as the reference sequence of human *RIG-I* gene, "?" symbolizes an undetermined nucleotide and "-" represents a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Paan – Olive baboon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mylu – Little brown myotis; Ptva – Large flying fox; Ptal – Black flying fox; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Loaf – Elephant; Ictr – Squirrel; Capo – Guinea pig; Mumu – Mouse; Orcu – European rabbit.

Supplementary Figure S2 - Mammalian *MDA5* nucleotide coding region sequences alignment.

MDA5 nucleotide coding region sequences for twenty-six mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol "." represents the same nucleotide as the reference sequence of human *MDA5* gene, "?" symbolizes an undetermined nucleotide and "-" represents a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Gogo – Gorilla; Patr – Chimpanzee; Papa – Bonobo; Poab – Orangutan; Nole – Gibbon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Eqca – Horse; Mylu – Little brown myotis; Ptal – Black flying fox; Loaf – Elephant; Orcu – European rabbit; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Ictr – Squirrel; Capo – Guinea pig.

Supplementary Figure S3 - Mammalian *LGP2* nucleotide coding region sequences alignment.

LGP2 nucleotide coding region sequences for thirty mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol "." represents the same nucleotide as the reference sequence of human *LGP2* gene and "-" symbolizes a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Tutr – Dolphin; Mylu – Little brown myotis; Ptva – Large flying fox; Ptal – Black flying fox; Loaf – Elephant; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Ocpr – American pika; Orcu – European rabbit; Ictr – Squirrel; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Capo – Guinea pig.

Supplementary Figure S4 - Mammalian RIG-I deduced protein sequences alignment.

RIG-I deduced protein sequences for twenty-six mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol "." represents the same codon as the reference sequence of human RIG-I protein, "?" symbolizes an undetermined codon and "-" represents a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Paan – Olive baboon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mylu – Little brown myotis; Ptva – Large flying fox; Ptal – Black flying fox; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Loaf – Elephant; Ictr – Squirrel; Capo – Guinea pig; Mumu – Mouse; Orcu – European rabbit.

Supplementary Figure S5 - Mammalian MDA5 deduced protein sequences alignment.

MDA5 deduced protein sequences for twenty-six mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol "." represents the same codon as the reference sequence of human MDA5 protein, "?" symbolizes an undetermined codon and "-" represents a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Gogo – Gorilla; Patr – Chimpanzee; Papa – Bonobo; Poab – Orangutan; Nole – Gibbon; Mamu – Rhesus macaque; Sabo – Black-capped

squirrel monkey; Caja – Marmoset; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Eqca – Horse; Mylu – Little brown myotis; Ptal – Black flying fox; Loaf – Elephant; Orcu – European rabbit; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Ictr – Squirrel; Capo – Guinea pig.

Supplementary Figure S6 - Mammalian LGP2 deduced protein sequences alignment.
LGP2 deduced protein sequences for thirty mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol "." represents the same codon as the reference sequence of human LGP2 protein and "-" symbolizes a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Tutr – Dolphin; Mylu – Little brown myotis; Ptva – Large flying fox; Ptal – Black flying fox; Loaf – Elephant; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Ocpr – American pika; Orcu – European rabbit; Ictr – Squirrel; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Capo – Guinea pig.

Supplementary Figure S7 - Mammalian *RIG-I* nucleotide trimmed sequences alignment.
RIG-I nucleotide trimmed sequences for twenty-six mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol "." represents the same nucleotide as the reference sequence of human *RIG-I* gene, "?" symbolizes an undetermined nucleotide and "-" represents a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Paan – Olive baboon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mylu – Little brown myotis; Ptva – Large flying fox; Ptal – Black flying fox; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Loaf – Elephant; Ictr – Squirrel; Capo – Guinea pig; Mumu – Mouse; Orcu – European rabbit.

Supplementary Figure S8 - Mammalian *MDA5* nucleotide trimmed sequences alignment.

MDA5 nucleotide trimmed sequences for twenty-six mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol "." represents the same nucleotide as the reference sequence of human *MDA5* gene, "?" symbolizes an undetermined nucleotide and "-" represents a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Gogo – Gorilla; Patr – Chimpanzee; Papa – Bonobo; Poab – Orangutan; Nole – Gibbon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Eqca – Horse; Mylu – Little brown myotis; Ptal – Black flying fox; Loaf – Elephant; Orcu – European rabbit; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Ictr – Squirrel; Capo – Guinea pig.

Supplementary Figure S9 - Mammalian *LGP2* nucleotide trimmed sequences alignment.

LGP2 nucleotide trimmed sequences for thirty mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol "." represents the same nucleotide as the reference sequence of human *LGP2* gene and "-" symbolizes a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Tutr – Dolphin; Mylu – Little brown myotis; Ptva – Large flying fox; Ptal – Black flying fox; Loaf – Elephant; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Ocpr – American pika; Orcu – European rabbit; Ictr – Squirrel; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Capo – Guinea pig.

Supplementary Table S2 - *RIG-I*, *MDA5* and *LGP2* likelihood ratio test (LRT) for PARRIS analysis from HyPhy software

Supplementary Table S3 - Positively-selected codon positions for RIG-I, MDA5 and LGP2 determined by six different methods.

7. References

1. Hoffmann JA: The immune response of *Drosophila*. *Nature* 2003, 426:33-38.
2. Akira S, Uematsu S, Takeuchi O: Pathogen recognition and innate immunity. *Cell* 2006, 124:783-801.
3. Kawai T, Akira S: The roles of TLRs, RLRs and NLRs in pathogen recognition. *International Immunology* 2009, 21:317-337.
4. Janeway CA, Jr.: Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harbor Symposia on Quantitative Biology* 1989, 54 Pt 1:1-13.
5. Medzhitov R, Janeway C, Jr.: Innate immunity. *The New England Journal of Medicine* 2000, 343:338-344.
6. Medzhitov R: Recognition of microorganisms and activation of the immune response. *Nature* 2007, 449:819-826.
7. van Vliet SJ, Garcia-Vallejo JJ, van Kooyk Y: Dendritic cells and C-type lectin receptors: coupling innate to adaptive immune responses. *Immunology & Cell Biology* 2008, 86:580-587.
8. Eisenacher K, Krug A: Regulation of RLR-mediated innate immune signaling-it is all about keeping the balance. *European Journal of Cell Biology* 2012, 91:36-47.
9. Kawai T, Akira S: Innate immune recognition of viral infection. *Nature Immunology* 2006, 7:131-137.
10. Yoneyama M, Kikuchi M, Matsumoto K, Imaizumi T, Miyagishi M, Taira K, Foy E, Loo YM, Gale M, Jr., Akira S, Yonehara S, Kato A, Fujita T: Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *The Journal of Immunology* 2005, 175:2851-2858.
11. Bamming D, Horvath CM: Regulation of signal transduction by enzymatically inactive antiviral RNA helicase proteins MDA5, RIG-I, and LGP2. *The Journal of Biological Chemistry* 2009, 284:9700-9712.
12. Kato H, Takahasi K, Fujita T: RIG-I-like receptors: cytoplasmic sensors for non-self RNA. *Immunological Reviews* 2011, 243:91-98.
13. Loo YM, Gale M, Jr.: Immune signaling by RIG-I-like receptors. *Immunity* 2011, 34:680-692.
14. Dixit E, Kagan JC: Intracellular pathogen detection by RIG-I-like receptors. *Advances in Immunology* 2013, 117:99-125.
15. Bruns AM, Horvath CM: Activation of RIG-I-like receptor signal transduction. *Critical Reviews in Biochemistry and Molecular Biology* 2012, 47:194-206.

16. Schmidt A, Rothenfusser S, Hopfner KP: Sensing of viral nucleic acids by RIG-I: from translocation to translation. *European Journal of Cell Biology* 2012, 91:78-85.
17. Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T, Miyagishi M, Taira K, Akira S, Fujita T: The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunology* 2004, 5:730-737.
18. Saito T, Hirai R, Loo YM, Owen D, Johnson CL, Sinha SC, Akira S, Fujita T, Gale M, Jr.: Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. *Proceedings of the National Academy of Sciences of the United States of America* 2007, 104:582-587.
19. Rothenfusser S, Goutagny N, DiPerna G, Gong M, Monks BG, Schoenemeyer A, Yamamoto M, Akira S, Fitzgerald KA: The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I. *The Journal of Immunology* 2005, 175:5260-5268.
20. Venkataraman T, Valdes M, Elsby R, Kakuta S, Caceres G, Saijo S, Iwakura Y, Barber GN: Loss of DExD/H box RNA helicase LGP2 manifests disparate antiviral responses. *The Journal of Immunology* 2007, 178:6444-6455.
21. Satoh T, Kato H, Kumagai Y, Yoneyama M, Sato S, Matsushita K, Tsujimura T, Fujita T, Akira S, Takeuchi O: LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107:1512-1517.
22. Kato H, Takeuchi O, Sato S, Yoneyama M, Yamamoto M, Matsui K, Uematsu S, Jung A, Kawai T, Ishii KJ, Yamaguchi O, Otsu K, Tsujimura T, Koh CS, Reis e Sousa C, Matsuura Y, Fujita T, Akira S: Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* 2006, 441:101-105.
23. Hornung V, Ellegast J, Kim S, Brzozka K, Jung A, Kato H, Poeck H, Akira S, Conzelmann KK, Schlee M, Endres S, Hartmann G: 5'-Triphosphate RNA is the ligand for RIG-I. *Science* 2006, 314:994-997.
24. Pichlmair A, Schulz O, Tan CP, Naslund TI, Liljestrom P, Weber F, Reis e Sousa C: RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* 2006, 314:997-1001.
25. Schlee M, Roth A, Hornung V, Hagmann CA, Wimmenauer V, Barchet W, Coch C, Janke M, Mihailovic A, Wardle G, Juranek S, Kato H, Kawai T, Poeck H, Fitzgerald KA, Takeuchi O, Akira S, Tuschl T, Latz E, Ludwig J, Hartmann G: Recognition of 5' triphosphate by RIG-I helicase requires short blunt double-

stranded RNA as contained in panhandle of negative-strand virus. *Immunity* 2009, 31:25-34.

26. Baum A, Sachidanandam R, Garcia-Sastre A: Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107:16303-16308.
27. Berke IC, Li Y, Modis Y: Structural basis of innate immune recognition of viral RNA. *Cellular Microbiology* 2013, 15:386-394.
28. Habjan M, Andersson I, Klingstrom J, Schumann M, Martin A, Zimmermann P, Wagner V, Pichlmair A, Schneider U, Muhlberger E, Mirazimi A, Weber F: Processing of genome 5' termini as a strategy of negative-strand RNA viruses to avoid RIG-I-dependent interferon induction. *PLoS One* 2008, 3:e2032.
29. Sumpter R, Jr., Loo YM, Foy E, Li K, Yoneyama M, Fujita T, Lemon SM, Gale M, Jr.: Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I. *Journal of Virology* 2005, 79:2689-2699.
30. Loo YM, Fornek J, Crochet N, Bajwa G, Perwitasari O, Martinez-Sobrido L, Akira S, Gill MA, Garcia-Sastre A, Katze MG, Gale M, Jr.: Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity. *Journal of Virology* 2008, 82:335-345.
31. Plumet S, Herschke F, Bourhis JM, Valentin H, Longhi S, Gerlier D: Cytosolic 5'-triphosphate ended viral leader transcript of measles virus as activator of the RIG I-mediated interferon response. *PLoS One* 2007, 2:e279.
32. Gitlin L, Barchet W, Gilfillan S, Cella M, Beutler B, Flavell RA, Diamond MS, Colonna M: Essential role of mda-5 in type I IFN responses to polyriboinosinic:polyribocytidylic acid and encephalomyocarditis picornavirus. *Proceedings of the National Academy of Sciences of the United States of America* 2006, 103:8459-8464.
33. McCartney SA, Thackray LB, Gitlin L, Gilfillan S, Virgin HW, Colonna M: MDA-5 recognition of a murine norovirus. *PLoS Pathogens* 2008, 4:e1000108.
34. Roth-Cross JK, Bender SJ, Weiss SR: Murine coronavirus mouse hepatitis virus is recognized by MDA5 and induces type I interferon in brain macrophages/ microglia. *Journal of Virology* 2008, 82:9829-9838.
35. Fredericksen BL, Keller BC, Fornek J, Katze MG, Gale M, Jr.: Establishment and maintenance of the innate antiviral response to West Nile Virus involves both RIG-I and MDA5 signaling through IPS-1. *Journal of Virology* 2008, 82:609-616.

36. Ablasser A, Bauernfeind F, Hartmann G, Latz E, Fitzgerald KA, Hornung V: RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nature Immunology* 2009, 10:1065-1072.
37. Chiu YH, Macmillan JB, Chen ZJ: RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell* 2009, 138:576-591.
38. Samanta M, Iwakiri D, Kanda T, Imaizumi T, Takada K: EB virus-encoded RNAs are recognized by RIG-I and activate signaling to induce type I IFN. *The EMBO Journal* 2006, 25:4207-4214.
39. Wang F, Gao X, Barrett JW, Shao Q, Bartee E, Mohamed MR, Rahman M, Werden S, Irvine T, Cao J, Dekaban GA, McFadden G: RIG-I mediates the co-induction of tumor necrosis factor and type I interferon elicited by myxoma virus in primary human macrophages. *PLoS Pathogens* 2008, 4:e1000099.
40. Pichlmair A, Schulz O, Tan CP, Rehwinkel J, Kato H, Takeuchi O, Akira S, Way M, Schiavo G, Reis e Sousa C: Activation of MDA5 requires higher-order RNA structures generated during virus infection. *Journal of Virology* 2009, 83:10761-10769.
41. Jiggins FM, Hurst GD: The evolution of parasite recognition genes in the innate immune system: purifying selection on *Drosophila melanogaster* peptidoglycan recognition proteins. *Journal of Molecular Evolution* 2003, 57:598-605.
42. Meyerson NR, Sawyer SL: Two-stepping through time: mammals and viruses. *Trends in Microbiology* 2011, 19:286-294.
43. Areal H, Abrantes J, Esteves PJ: Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evolutionary Biology* 2011, 11:368.
44. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 1994, 22:4673-4680.
45. Hall T: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 1999, 41:95-98.
46. Posada D, Crandall KA: The effect of recombination on the accuracy of phylogeny estimation. *Journal of Molecular Evolution* 2002, 54:396-402.
47. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD: GARD: a genetic algorithm for recombination detection. *Bioinformatics* 2006, 22:3096-3098.

48. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD: Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular Biology and Evolution* 2006, 23:1891-1901.
49. Pond SL, Frost SD: Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 2005, 21:2531-2533.
50. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL: Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 2010, 26:2455-2457.
51. Posada D: jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 2008, 25:1253-1256.
52. Zwickl DJ: Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. *PhD Thesis.* University of Texas, 2006.
53. Yang Z: PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* 1997, 13:555-556.
54. Yang Z: PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 2007, 24:1586-1591.
55. Nielsen R, Yang Z: Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 1998, 148:929-936.
56. Yang Z: Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution* 1998, 15:568-573.
57. Yang Z, Nielsen R, Goldman N, Pedersen AM: Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 2000, 155:431-449.
58. Scheffler K, Martin DP, Seoighe C: Robust inference of positive selection from recombining coding sequences. *Bioinformatics* 2006, 22:2493-2499.
59. Wlasiuk G, Nachman MW: Adaptation and constraint at Toll-like receptors in primates. *Molecular Biology and Evolution* 2010, 27:2172-2186.
60. Lemos de Matos A, Liu J, McFadden G, Esteves P: Evolution and divergence of the mammalian SAMD9/SAMD9L gene family. *BMC Evolutionary Biology* 2013, 13:121.
61. Yang Z, Wong WS, Nielsen R: Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* 2005, 22:1107-1118.

62. Kosakovsky Pond SL, Frost SD: Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* 2005, 22:1208-1222.
63. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL: Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics* 2012, 8:e1002764.
64. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K: FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution* 2013, 30:1196-1205.
65. Zhang J, Nielsen R, Yang Z: Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution* 2005, 22:2472-2479.
66. Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K: A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution* 2011.
67. Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW: Adaptive evolution of young gene duplicates in mammals. *Genome Research* 2009, 19:859-867.
68. Machado J, Johnson W, O'Brien S, Vasconcelos V, Antunes A: Adaptive evolution of the matrix extracellular phosphoglycoprotein in mammals. *BMC Evolutionary Biology* 2011, 11:342.
69. Grayson P, Civetta A: Positive selection and the evolution of izumo genes in mammals. *International Journal of Evolutionary Biology* 2012, 2012:7.
70. Neves F, Abrantes J, Steinke JW, Esteves PJ: Maximum-likelihood approaches reveal signatures of positive selection in IL genes in mammals. *Innate Immunity* 2014, 20:184-191.
71. Pinheiro A, Woof JM, Abi-Rached L, Parham P, Esteves PJ: Computational analyses of an evolutionary arms race between mammalian immunity mediated by immunoglobulin A and its subversion by bacterial pathogens. *PLoS One* 2013, 8:e73934.
72. Anisimova M, Bielawski JP, Yang Z: Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution* 2001, 18:1585-1592.
73. Gharib WH, Robinson-Rechavi M: The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Molecular Biology and Evolution* 2013, 30:1675-1686.

74. Scheffler K, Seoighe C: A Bayesian model comparison approach to inferring positive selection. *Molecular Biology and Evolution* 2005, 22:2531-2540.
75. Anisimova M, Yang Z: Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Molecular Biology and Evolution* 2007, 24:1219-1228.
76. Vasseur E, Patin E, Laval G, Pajon S, Fornarino S, Crouau-Roy B, Quintana-Murci L: The selective footprints of viral pressures at the human RIG-I-like receptor family. *Human Molecular Genetics* 2011, 20:4462-4474.
77. Buckley KM, Rast JP: Dynamic evolution of toll-like receptor multigene families in echinoderms. *Frontiers in Immunology* 2012, 3:136.
78. Alcaide M, Edwards SV: Molecular evolution of the toll-like receptor multigene family in birds. *Molecular Biology and Evolution* 2011, 28:1703-1715.
79. Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, Bouchier C, Tichit M, Neyrolles O, Gicquel B, Kidd JR, Kidd KK, Alcais A, Ragimbeau J, Pellegrini S, Abel L, Casanova JL, Quintana-Murci L: Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genetics* 2009, 5:e1000562.
80. Jann OC, Werling D, Chang JS, Haig D, Glass EJ: Molecular evolution of bovine Toll-like receptor 2 suggests substitutions of functional relevance. *BMC Evolutionary Biology* 2008, 8:288.
81. Nakajima T, Ohtani H, Satta Y, Uno Y, Akari H, Ishida T, Kimura A: Natural selection in the TLR-related genes in the course of primate evolution. *Immunogenetics* 2008, 60:727-735.
82. Wlasiuk G, Khan S, Switzer WM, Nachman MW: A history of recurrent positive selection at the toll-like receptor 5 in primates. *Molecular Biology and Evolution* 2009, 26:937-949.
83. Smith JM, Smith NH: Synonymous nucleotide divergence: what is "saturation"? *Genetics* 1996, 142:1033-1036.
84. Seo TK, Kishino H: Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins. *Systematic Biology* 2008, 57:367-377.
85. Kawai T, Takahashi K, Sato S, Coban C, Kumar H, Kato H, Ishii KJ, Takeuchi O, Akira S: IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nature Immunology* 2005, 6:981-988.
86. Takahasi K, Yoneyama M, Nishihori T, Hirai R, Kumeta H, Narita R, Gale M, Jr., Inagaki F, Fujita T: Nonself RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. *Molecular Cell* 2008, 29:428-440.

87. Fairman-Williams ME, Guenther UP, Jankowsky E: SF1 and SF2 helicases: family matters. *Current Opinion in Structural Biology* 2010, 20:313-324.

88. Cui S, Eisenacher K, Kirchhofer A, Brzozka K, Lammens A, Lammens K, Fujita T, Conzelmann KK, Krug A, Hopfner KP: The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I. *Molecular Cell* 2008, 29:169-179.

Supplementary Table S1 - List of mammalian species and genes accession numbers used in this study

Order	Family	Species common name (*Scientific name*)	Gene Abbreviation	Database ID
Artiodactyla	Bovidae	Cow (*Bos taurus*)	LGP2_Bota	ENSBTAT00000065738 [a]
			MDA5_Bota	ENSBTAG00000008142 [a]
			RIG-I_Bota	ENSBTAG00000003366 [a]
		Sheep (*Ovis aries*)	LGP2_Ovar	XM_004012918.1 [b]
			MDA5_Ovar	XM_004004655.1 [b]
			RIG-I_Ovar	XM_004005323.1 [b]
	Suidae	Pig (*Sus scrofa*)	LGP2_Susc	ENSSSCT00000018960 [a]
			MDA5_Susc	NM_001100194.1 [b]
			RIG-I_Susc	ENSSSCG00000030408 [a]
Carnivora	Canidae	Dog (*Canis lupus familiaris*)	LGP2_Calu	ENSCAFG00000015720 [a]
			MDA5_Calu	ENSCAFG00000010438 [a]
			RIG-I_Calu	ENSCAFG00000001807 [a]
	Felidae	Cat (*Felis catus*)	LGP2_Feca	XM_003996876.1 [b]
			RIG-I_Feca	XM_003995540.1 [b]
	Mustelidae	Ferret (*Mustela putorius furo*)	LGP2_Mupu	ENSMPUG00000010568 [a]
			MDA5_Mupu	ENSMPUG00000000564 [a]
	Ursidae	Giant panda	LGP2_Aime	ENSAMEG00000005941 [a]
		(*Ailuropoda melanoleuca*)	MDA5_Aime	ENSAMEG00000005698 [a]
			RIG-I_Aime	ENSAMEG00000003766 [a]
Cetacea	Delphinidae	Dolphin (*Tursiops truncatus*)	LGP2_Tutr	ENSTTRG00000006650 [a]
Chiroptera	Pteropodidae	Black flying fox (*Pteropus alecto*)	LGP2_Ptal	JN031516.1 [b]
			MDA5_Ptal	JN031515.1 [b]
			RIG-I_Ptal	JN031514.1 [b]
		Large flying fox	LGP2_Ptva	ENSPVAG00000003855 [a]
		(*Pteropus vampyrus*)	RIG-I_Ptva	ENSPVAG00000009207 [a]
	Vespertilionidae	Little brown myotis	LGP2_Mylu	ENSMLUG00000012001 [a]
		(*Myotis lucifugus*)	MDA5_Mylu	ENSMLUG00000009505 [a]
			RIG-I_Mylu	ENSMLUG00000003041 [a]
Perissodactyla	Equidae	Horse (*Equus caballus*)	LGP2_Eqca	XM_001495212.2 [b]
			MDA5_Eqca	ENSECAT00000008541 [a]
			RIG-I_Eqca	ENSECAG00000021989 [a]
Proboscidea	Elephantidae	Elephant (*Loxodonta africana*)	LGP2_Loaf	ENSLAFG00000027301 [a]
			MDA5_Loaf	XM_003405827.1 [b]
			RIG-I_Loaf	ENSLAFG00000005416 [a]
Lagomorpha	Leporidae	European rabbit	LGP2_Orcu	ENSOCUG00000013278 [a]
		(*Oryctolagus cuniculus*)	MDA5_Orcu	ENSOCUG00000002863 [a]
			RIG-I_Orcu	ENSOCUT00000029633 [a]
	Ochotonidae	American pika (*Ochotona princeps*)	LGP2_Ocpr	ENSOPRG00000000753 [a]
Primates	Cebidae	Black-capped squirrel monkey	LGP2_Sabo	XM_003942770.1 [b]
		(*Saimiri boliviensis boliviensis*)	MDA5_Sabo	XM_003921929.1 [b]
			RIG-I_Sabo	XM_003939729.1 [b]
		Marmoset (*Callithrix jacchus*)	LGP2_Caja	ENSCJAT00000028952 [a]
			MDA5_Caja	ENSCJAT00000011621 [a]
			RIG-I_Caja	ENSCJAT00000015302 [a]
	Cercopithecidae	Olive baboon (*Papio anubis*)	RIG-I_Paan	XM_003911615.1 [b]
		Rhesus macaque (*Macaca mulatta*)	LGP2_Mamu	ENSMMUG00000014755 [a]
			MDA5_Mamu	ENSMMUG00000003202 [a]
			RIG-I_Mamu	ENSMMUT00000017446 [a]
	Cheirogaleidae	Mouse lemur (*Microcebus murinus*)	LGP2_Mimu	ENSMICG00000011933 [a]
			RIG-I_Mimu	ENSMICG00000006555 [a]
	Galagidae	Bushbaby (*Otolemur garnettii*)	LGP2_Otga	ENSOGAG00000025321 [a]
			MDA5_Otga	ENSOGAG00000012611 [a]
			RIG-I_Otga	ENSOGAG00000010536 [a]

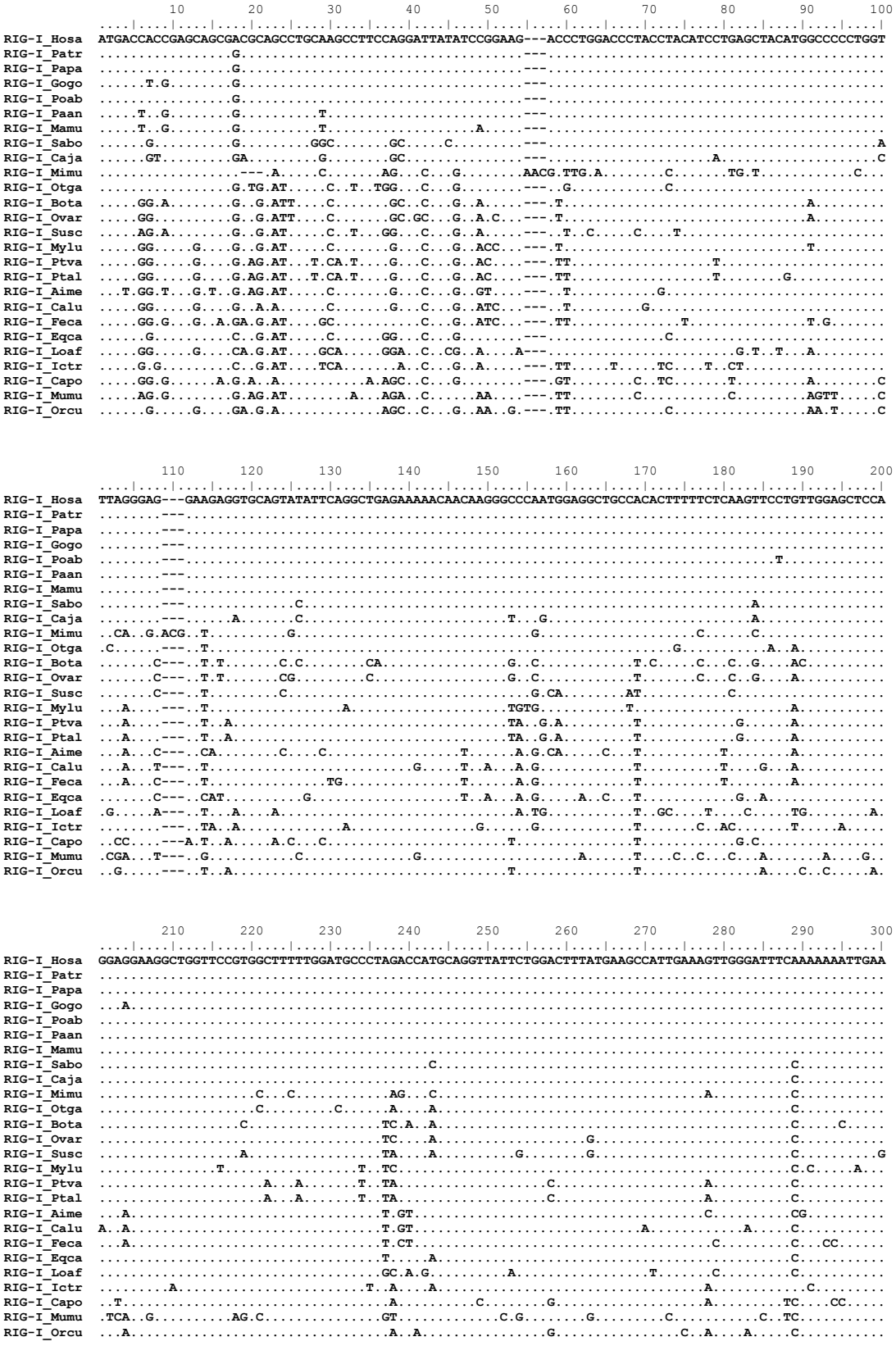
(Continues next page).

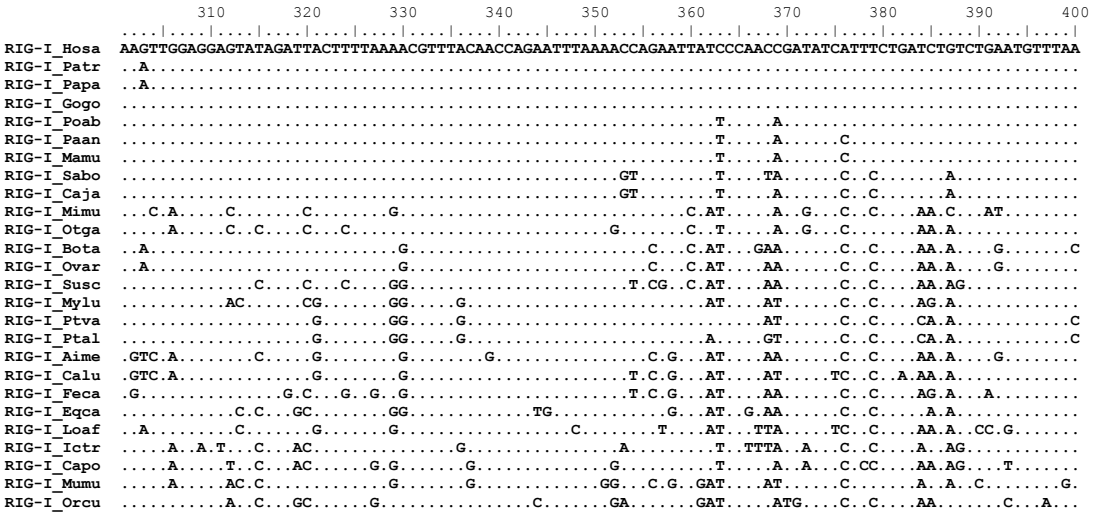
Supplementary Table S1 (continuation)

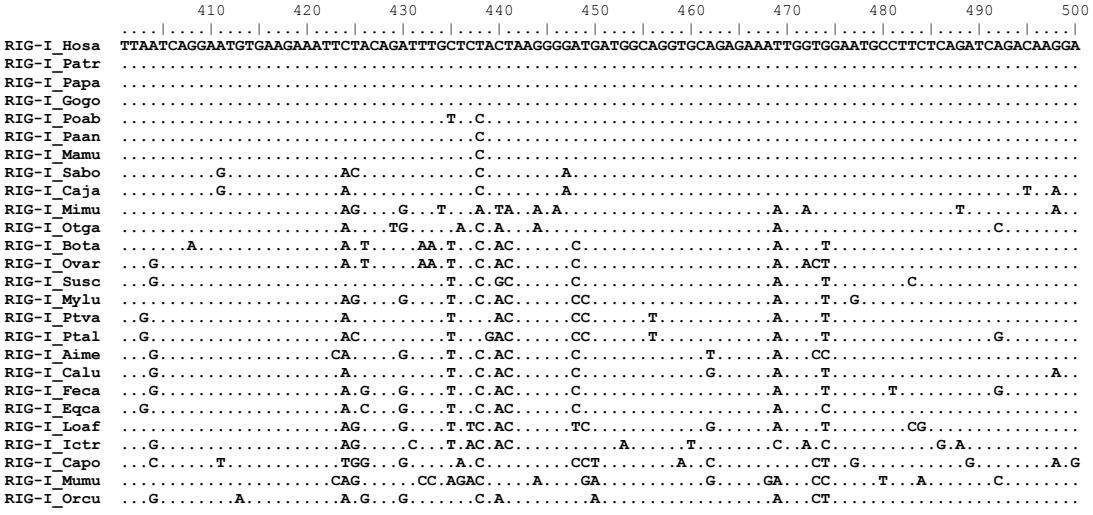
Order	Family	Species common name (*Scientific name*)	Gene Abbreviation	Database ID
Primates	Hominidae	Bonobo (*Pan paniscus*)	LGP2_Papa	XM_003813868.1 [b]
			MDA5_Papa	XM_003820935.1 [b]
			RIG-I_Papa	XM_003830100.1 [b]
		Chimpanzee (*Pan troglodytes*)	LGP2_Patr	ENSPTRG00000009191 [a]
			MDA5_Patr	ENSPTRG00000012582 [a]
			RIG-I_Patr	ENSPTRG00000020844 [a]
		Human (*Homo sapiens*)	LGP2_Hosa	ENST00000251642 [a]
			MDA5_Hosa	ENST00000263642 [a]
			RIG-I_Hosa	ENST00000379883 [a]
		Orangutan (*Pongo abelii*)	LGP2_Poab	NM_001131127.1 [b]
			MDA5_Poab	ENSPPYG00000012887 [a]
			RIG-I_Poab	XM_002819715.1 [b]
		Gorilla (*Gorilla gorilla*)	LGP2_Gogo	ENSGGOG00000005458 [a]
			MDA5_Gogo	ENSGGOT00000015950 [a]
			RIG-I_Gogo	ENSGGOT00000032102 [a]
	Hylobatidae	Gibbon (*Nomascus leucogenys*)	MDA5_Nole	XM_003266320.1 [b]
Rodentia	Caviidae	Guinea pig (*Cavia porcellus*)	LGP2_Capo	ENSCPOG00000023721 [a]
			MDA5_Capo	ENSCPOG00000007154 [a]
			RIG-I_Capo	ENSCPOG00000001598 [a]
	Cricetidae	Chinese hamster (*Cricetulus griseus*)	LGP2_Crgr	XM_003504640.1 [b]
			MDA5_Crgr	XM_003508631.1 [b]
	Muridae	Rat (*Rattus norvegicus*)	LGP2_Rano	ENSRNOG00000018247 [a]
			MDA5_Rano	ENSRNOG00000006227 [a]
		Mouse (*Mus musculus*)	LGP2_Mumu	ENSMUST00000017974 [a]
			MDA5_Mumu	ENSMUST00000028259 [a]
			RIG-I_Mumu	ENSMUST00000037907 [a]
	Sciuridae	Squirrel (*Ictidomys tridecemlineatus*)	LGP2_Ictr	ENSSTOG00000006524 [a]
			MDA5_Ictr	ENSSTOG00000004953 [a]
			RIG-I_Ictr	ENSSTOG00000006994 [a]

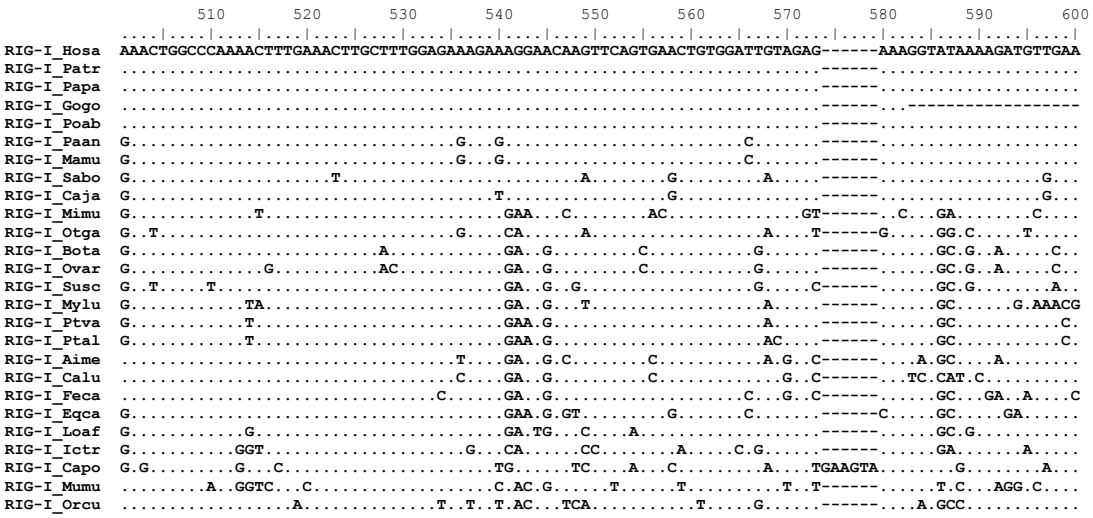
Database ID: [a]Ensembl; [b]NCBI GenBank.

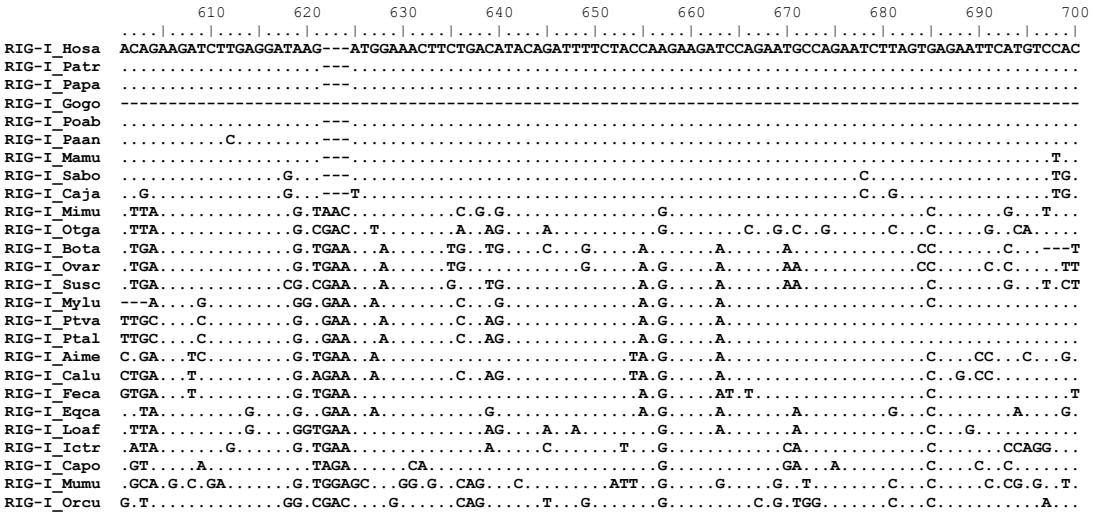
Supplementary Figure S1

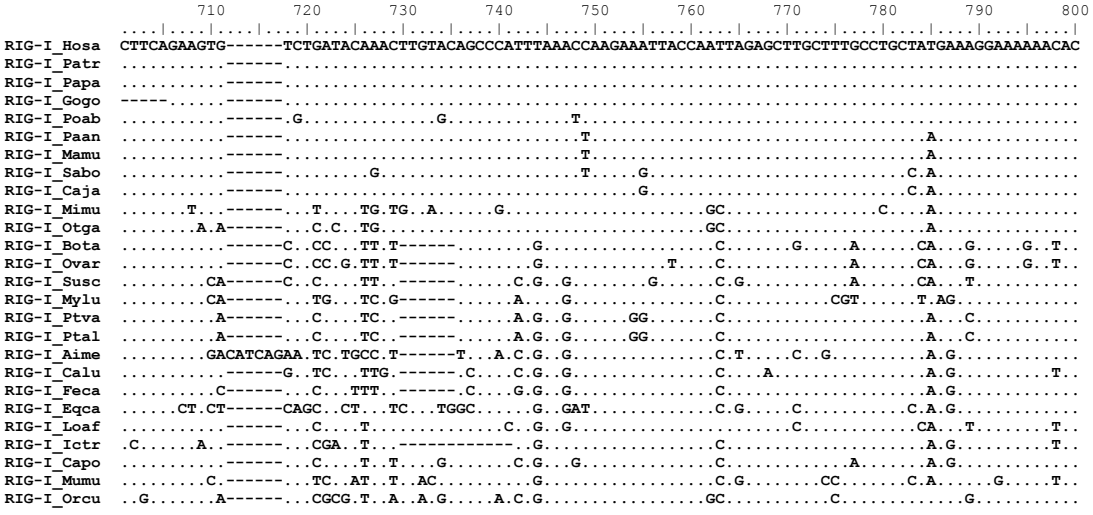


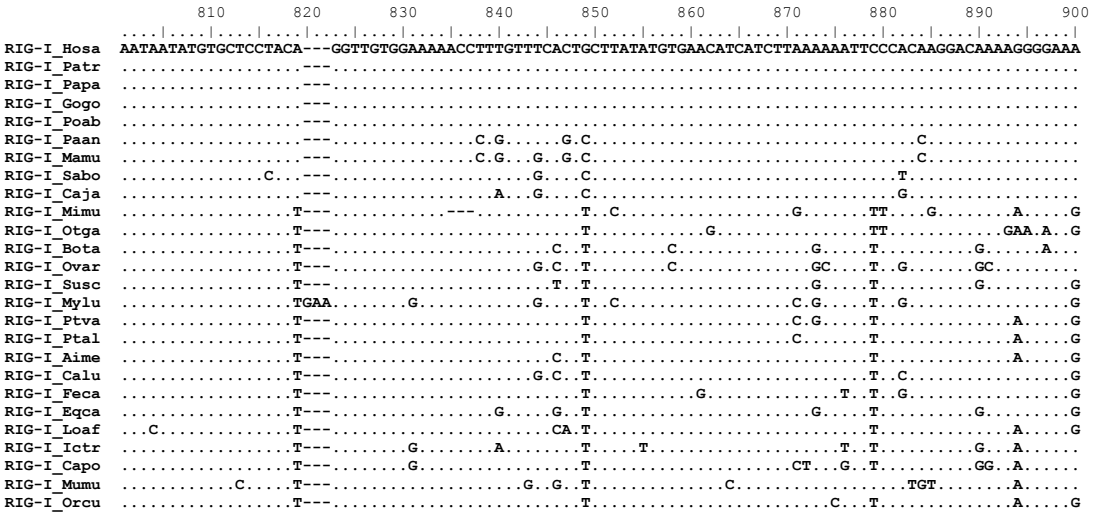

310 320 330 340 350 360 370 380 390 400
RIG-I_Hosa AAGTTGGAGGAGTATAGATTACTTTTAAAACGTTTACAACCAGAATTTAAAACCAGAATTATCCCAACCGATATCATTTCTGATCTGTCTGAATGTTTAA
RIG-I_Patr ..A.....
RIG-I_Papa ..A.....
RIG-I_Gogo .....
RIG-I_Poab .....T.....A.....
RIG-I_Paan .....T.....A.....C.....
RIG-I_Mamu .....T.....A.....C.....
RIG-I_Sabo .....GT.....T....TA.....C..C.....A.....
RIG-I_Caja .....GT.....T.....A.....C..C.....A.....
RIG-I_Mimu ...C.A.....C.....C.....G.....C.AT.....A..G...C..C....AA.C...AT.....
RIG-I_Otga .....A.....C..C....C...C.....G.....C..T.....A..G...C.....AA.A.....
RIG-I_Bota ..A.....G.....C...C.AT...GAA.....C..C....AA.A....G.....C
RIG-I_Ovar ..A.....G.....C...C.AT....AA.....C..C....AA.A....G.....
RIG-I_Susc .....C....C...C....GG.....T.CG..C.AT....AA.....C..C....AA.AG.....
RIG-I_Mylu .....AC.....CG.....GG....G.....AT....AT.....C..C....AG.A.....
RIG-I_Ptva .....G.....GG....G.....AT.....C..C....CA.A.....C
RIG-I_Ptal .....G.....GG....G.....A.....GT.....C..C....CA.A.....C
RIG-I_Aime .GTC.A.....C.....G.....G.....G.....C.G...AT....AA.....C..C....AA.A....G.....
RIG-I_Calu .GTC.A.....G.....G.....T.C.G...AT....AT....TC..C..A.AA.A.....
RIG-I_Feca .G.....G.C...G..G..G.....T.C.G...AT....AA.....C..C....AG.A...A.....
RIG-I_Eqca .....C.C...GC.....GG.....TG.....G...AT..G.AA.....C..C....A.A.....
RIG-I_Loaf ..A.....C.....G.....G.....C.....T....AT...TTA.....TC..C....AA.A..CC.G.....
RIG-I_Ictr .....A..A.T...C...AC.....G.....A.....T..TTTA..A...C..C....A..AG.....
RIG-I_Capo .....A.....T..C...AC.....G.G.....G.....G.....T.....A..A...C.CC....AA.AG....T.....
RIG-I_Mumu .....A.....AC.C.....G.....G.....GG...C.G..GAT....AT.....C.....A..A..C.....G.
RIG-I_Orcu .....A..C...GC.....G.....C.....GA.....GAT.....ATG....C..C....AA.....C...A...

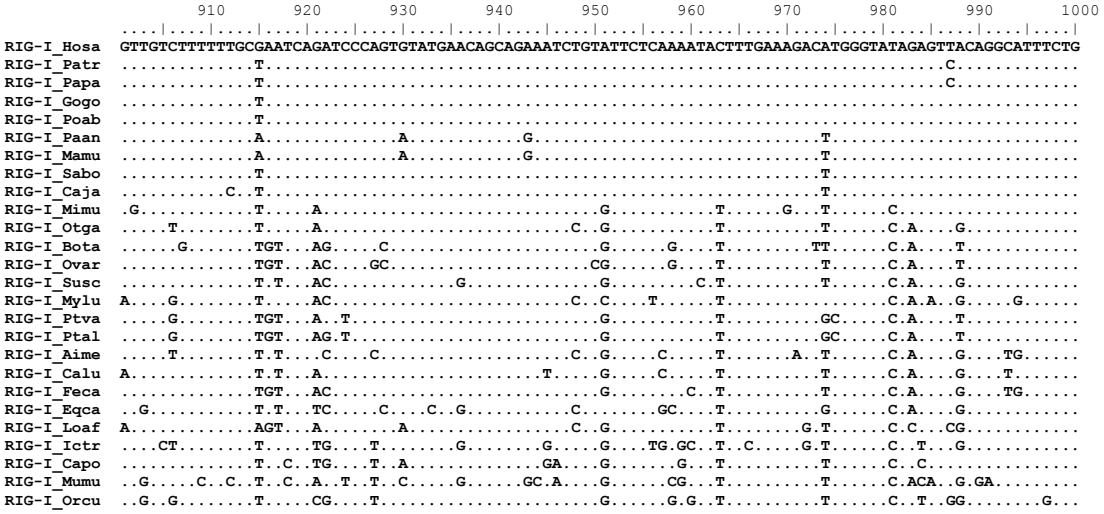

410 420 430 440 450 460 470 480 490 500
RIG-I_Hosa TTAATCAGGAATGTGAAGAAATTCTACAGATTTGCTCTACTAAGGGGATGATGGCAGGTGCAGAGAAATTGGTGGAATGCCTTCTCAGATCAGACAAGGA
RIG-I_Patr .....
RIG-I_Papa .....
RIG-I_Gogo .....
RIG-I_Poab .....T..C.....
RIG-I_Paan .....C.....
RIG-I_Mamu .....C.....
RIG-I_Sabo .....G.....AC.....C.....A.....
RIG-I_Caja .....G.....A.....C.....A.....T..A..
RIG-I_Mimu .....AG....G...T...A.TA..A.A.....A..A.....T.....A..
RIG-I_Otga .....A....TG.....A.C.A...A.....A.....C.....
RIG-I_Bota .....A.....A.T.....AA.T..C.AC.....C.....A....T.....
RIG-I_Ovar ...G.....A.T.....AA.T..C.AC.....C.....A..ACT.....
RIG-I_Susc ...G.....T..C.GC.....C.....A....T.....C.....
RIG-I_Mylu .....AG....G....T..C.AC.....CC.....A....T..G.....
RIG-I_Ptva ..G.....A.....T....AC.....CC.....T.....A....T.....
RIG-I_Ptal ..G.....AC.....T...GAC.....CC.....T.....A....T.....G.....
RIG-I_Aime ...G.....CA.....G....T..C.AC.....C.....T.....A...CC.....
RIG-I_Calu ...G.....A.....T..C.AC.....C.....G.....A....T.....A..
RIG-I_Feca ...G.....A.G...G....T..C.AC.....C.....A....T.....T.....G.....
RIG-I_Eqca ..G.....A.C...G....T..C.AC.....C.....A....C.....
RIG-I_Loaf .....AG....G....T.TC.AC.....TC.....G.....A....T.....CG.....
RIG-I_Ictr ...G.....AG.....C...T.AC.AC.....A.....T.....C..A.C.....G.A.....
RIG-I_Capo ...C.....T.....TGG...G.....A.C.....CCT.....A..C.....CT..G.....G.....A.G
RIG-I_Mumu .....CAG.....CC.AGAC.....A....GA.....G.....GA...CC.....T...A.....C.....
RIG-I_Orcu ...G.....A.....A.G...G.....C.A.....A.....A...CT.....

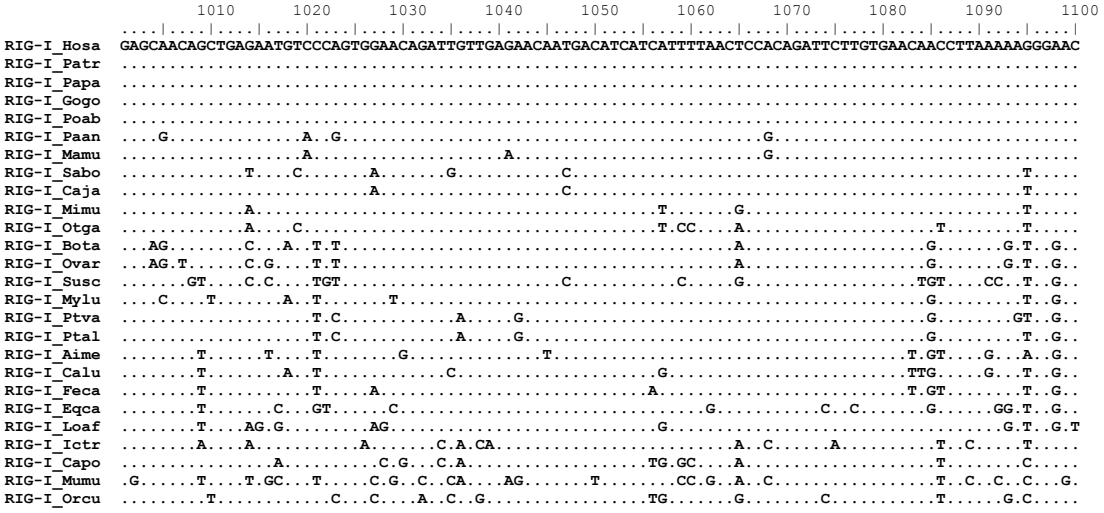

510 520 530 540 550 560 570 580 590 600
RIG-I_Hosa AAACTGGCCCAAAACTTTGAAACTTGCTTTGGAGAAAGAAAGGAACAAGTTCAGTGAACTGTGGATTGTAGAG------AAAGGTATAAAAGATGTTGAA
RIG-I_Patr .....------.....
RIG-I_Papa .....------.....
RIG-I_Gogo .....------...----------
RIG-I_Poab .....------.....
RIG-I_Paan G.....G...G.....C.....------.....
RIG-I_Mamu G.....G...G.....C.....------.....
RIG-I_Sabo G.....T.....A.....G.....A.....------.....G...
RIG-I_Caja G.....T.....G.....------.....G...
RIG-I_Mimu G.....T.....GAA...C.....AC.....GT------..C...GA.....C....
RIG-I_Otga G..T.....G....CA.....A.....A....T------G.....GG.C.....T.....
RIG-I_Bota G.....A.....GA..G.....C.....G.....------.....GC.G..A.....C..
RIG-I_Ovar G.....G.....AC.....GA..G.....C.....G.....------.....GC.G..A.....C..
RIG-I_Susc G..T.....T.....GA..G..G.....G.....C------.....GC.G.....A..
RIG-I_Mylu G.....TA.....GA..G...T.....A.....------.....GC.....G.AAACG
RIG-I_Ptva G.....T.....GAA.G.....A.....------.....GC.....C.
RIG-I_Ptal G.....T.....GAA.G.....AC.....------.....GC.....C.
RIG-I_Aime .....T....GA..G.C.....C.....A.G..C------....A.GC....A.....
RIG-I_Calu .....C....GA..G.....C.....G..C------...TC.CAT.C.....
RIG-I_Feca .....C.....GA..G.....C...G..C------.....GC...GA..A....C
RIG-I_Eqca G.....GAA.G.GT.....G.....C.....------C.....GC.....GA.....
RIG-I_Loaf G.....GA.TG...C....A.....------.....GC.G.....
RIG-I_Ictr G.....GGT.....G...CA.....CC.....A.....C.G.....------.....GA.....A.....
RIG-I_Capo G.G.....G...C.....TG.....TC....A...C.....A....TGAAGTA.....G.....A...
RIG-I_Mumu .....A..GGTC...C.....C.AC.G.....T.....T.....T..T------.....T.C...AGG.C....
RIG-I_Orcu .....A.....T..T..T.AC...TCA.....T.....G.....------....A.GCC.....

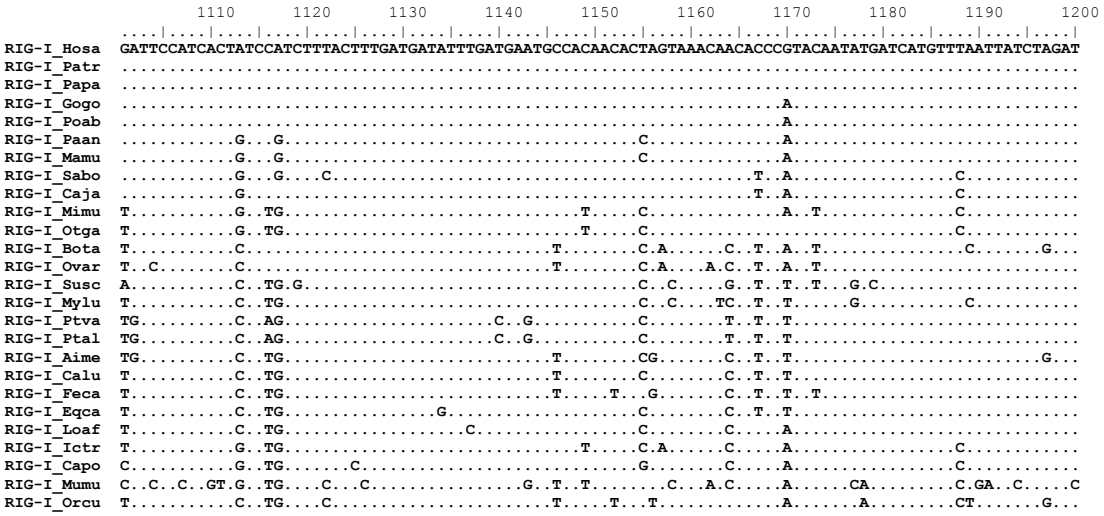

610 620 630 640 650 660 670 680 690 700
RIG-I_Hosa ACAGAAGATCTTGAGGATAAG---ATGGAAACTTCTGACATACAGATTTTCTACCAAGAAGATCCAGAATGCCAGAATCTTAGTGAGAATTCATGTCCAC
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

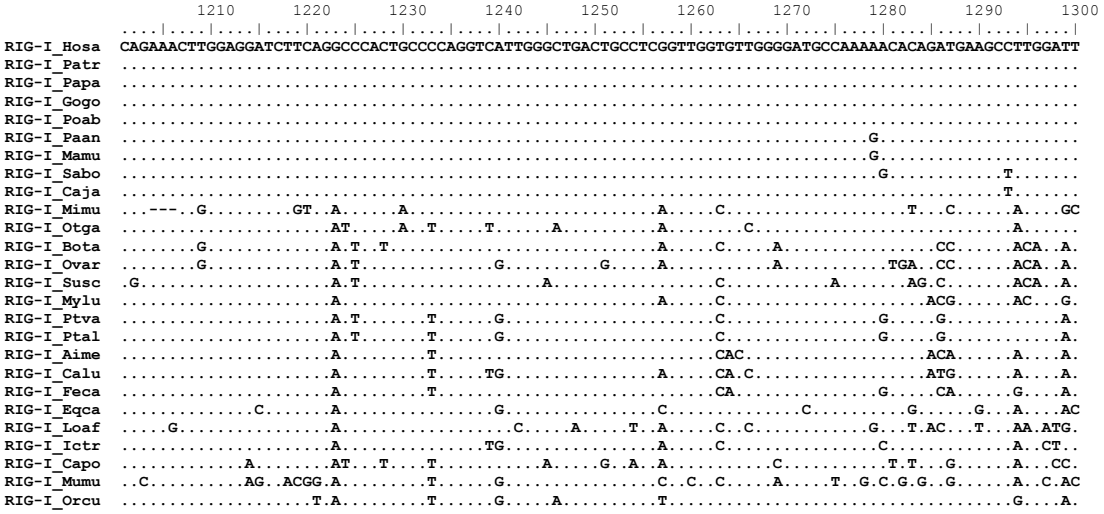

710 720 730 740 750 760 770 780 790 800
RIG-I_Hosa CTTCAGAAGTG------TCTGATACAAACTTGTACAGCCCATTTAAACCAAGAAATTACCAATTAGAGCTTGCTTTGCCTGCTATGAAAGGAAAAAACAC
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

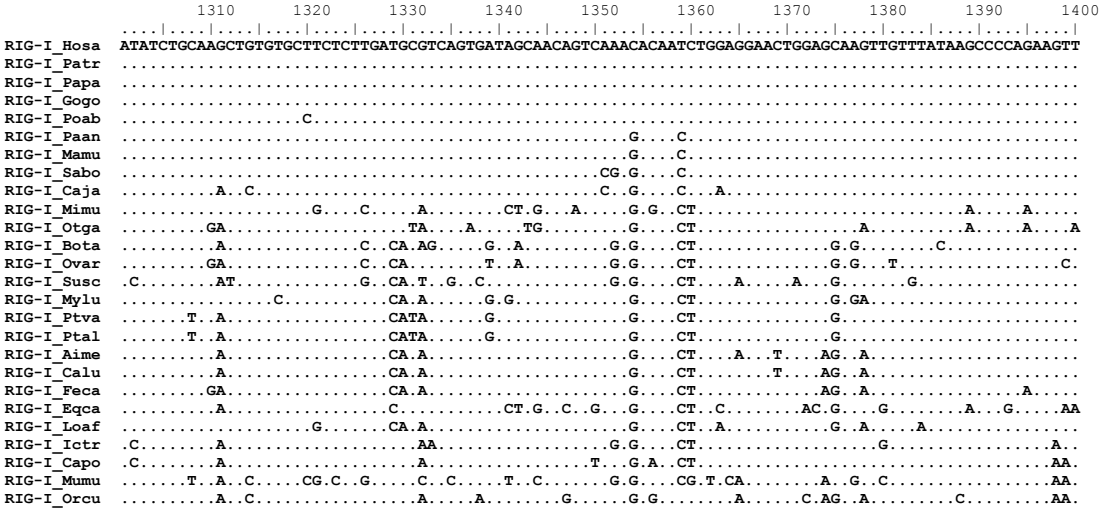

810 820 830 840 850 860 870 880 890 900
RIG-I_Hosa AATAATATGTGCTCCTACA---GGTTGTGGAAAAACCTTTGTTTCACTGCTTATATGTGAACATCATCTTAAAAAATTCCCACAAGGACAAAAGGGGAAA
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

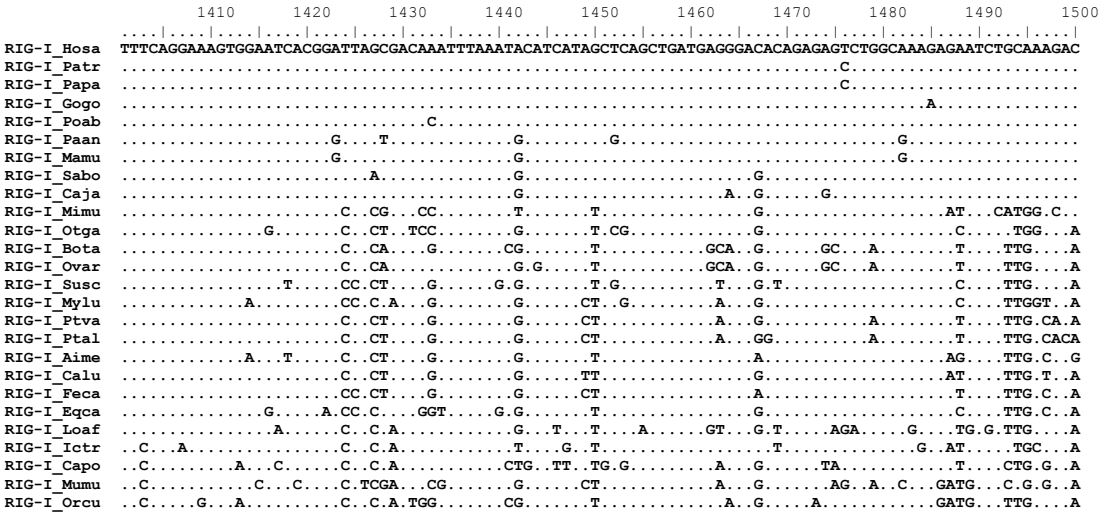

910 920 930 940 950 960 970 980 990 1000
RIG-I_Hosa GTTGTCTTTTTTGCGAATCAGATCCCAGTGTATGAACAGCAGAAATCTGTATTCTCAAAATACTTTGAAAGACATGGGTATAGAGTTACAGGCATTTCTG
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

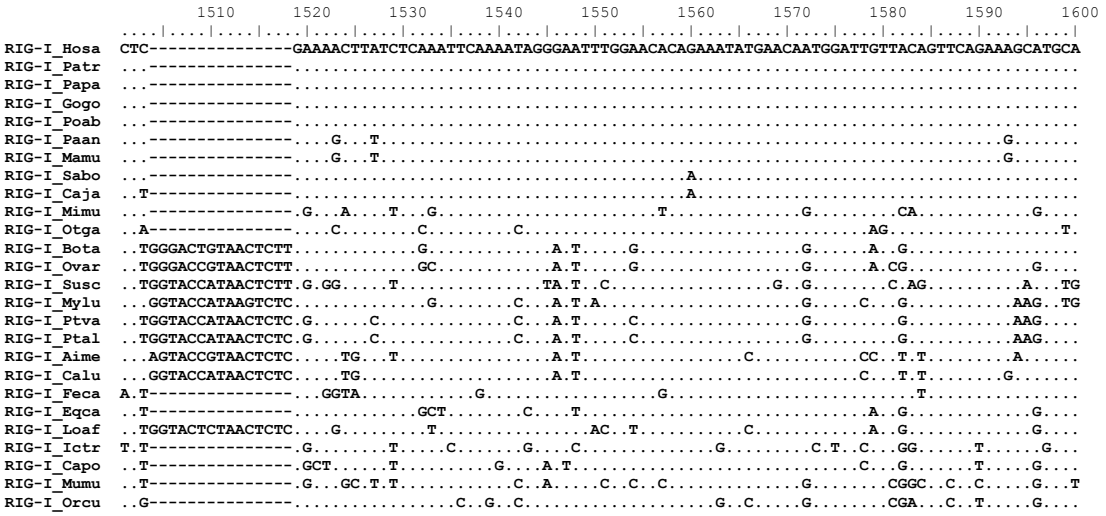

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
RIG-I_Hosa GAGCAACAGCTGAGAATGTCCCAGTGGAACAGATTGTTGAGAACAATGACATCATCATTTTAACTCCACAGATTCTTGTGAACAACCTTAAAAAGGGAAC
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

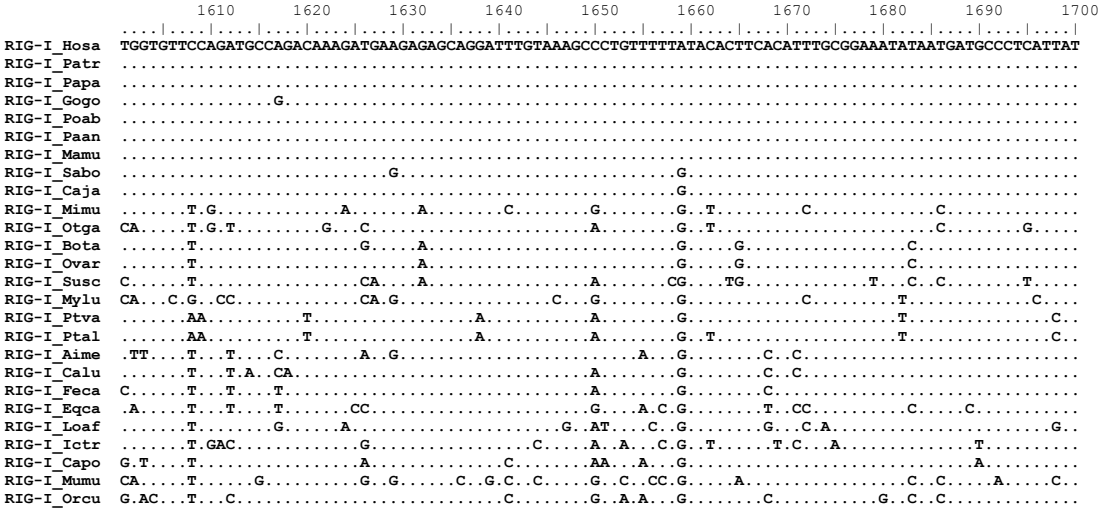

1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
RIG-I_Hosa GATTCCATCACTATCCATCTTTACTTTGATGATATTTGATGAATGCCACAACACTAGTAAACAACACCCGTACAATATGATCATGTTTAATTATCTAGAT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

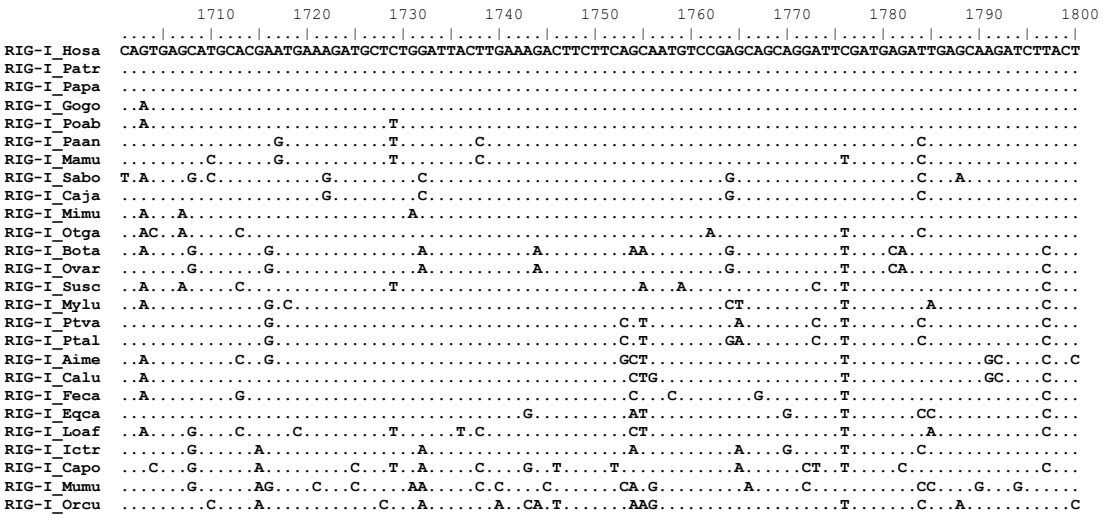

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
RIG-I_Hosa CAGAAACTTGGAGGATCTTCAGGCCCACTGCCCCAGGTCATTGGGCTGACTGCCTCGGTTGGTGTTGGGGATGCCAAAAACACAGATGAAGCCTTGGATT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

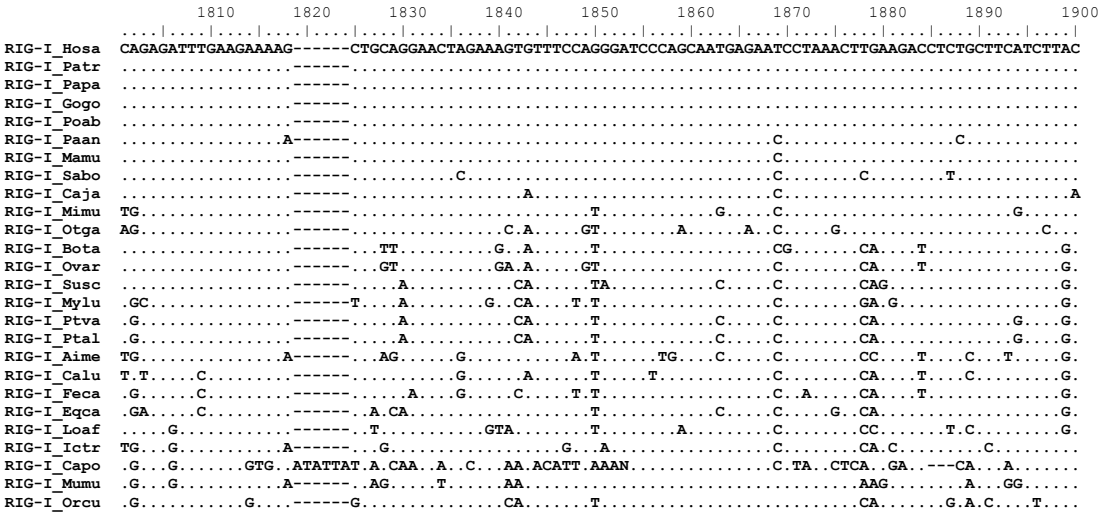

1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
RIG-I_Hosa ATATCTGCAAGCTGTGTGCTTCTCTTGATGCGTCAGTGATAGCAACAGTCAAACACAATCTGGAGGAACTGGAGCAAGTTGTTTATAAGCCCCAGAAGTT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

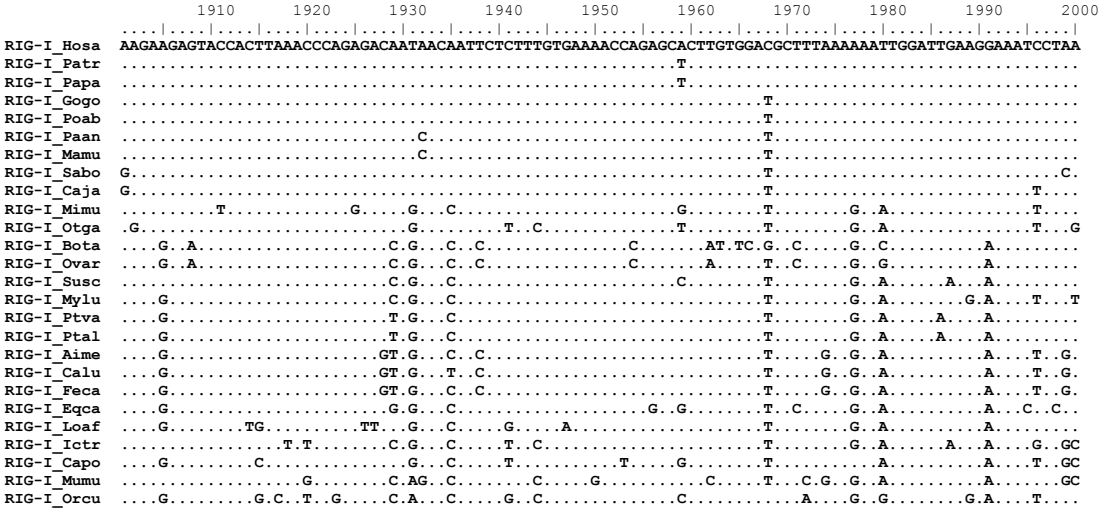

1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
RIG-I_Hosa TTTCAGGAAAGTGGAATCACGGATTAGCGACAAATTTAAATACATCATAGCTCAGCTGATGAGGGACACAGAGAGTCTGGCAAAGAGAATCTGCAAAGAC
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

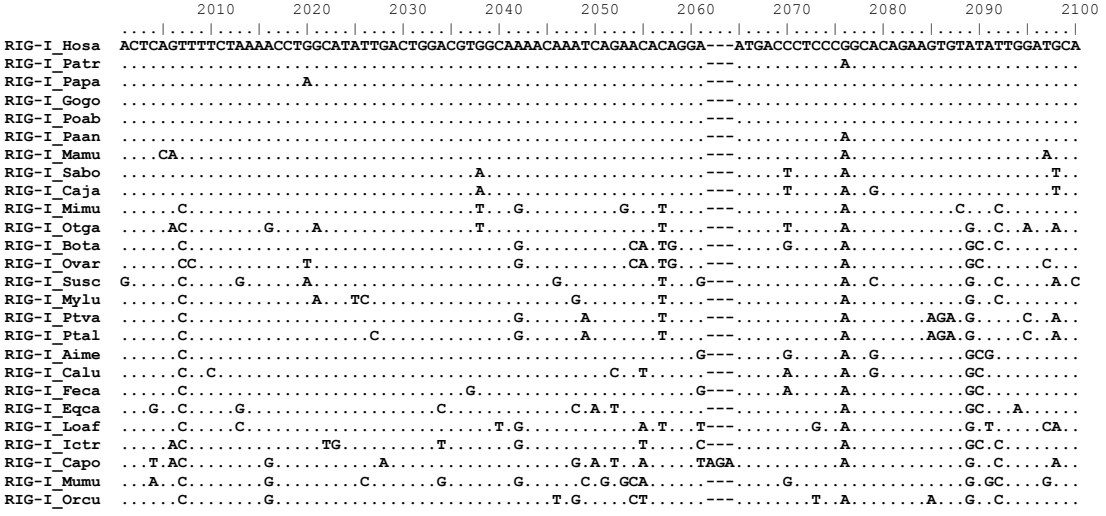

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
RIG-I_Hosa CTC----------GAAAACTTATCTCAAATTCAAAATAGGGAATTTGGAACACAGAAATATGAACAATGGATTGTTACAGTTCAGAAAGCATGCA
RIG-I_Patr ...----------
RIG-I_Papa ...----------
RIG-I_Gogo ...----------
RIG-I_Poab ...----------
RIG-I_Paan ...----------....G...T.....G.....
RIG-I_Mamu ...----------....G...T.....G.....
RIG-I_Sabo ...----------.....A.....
RIG-I_Caja ..T----------.....A.....
RIG-I_Mimu ...----------.G...A....T...G.....T.....G.....CA.....G....
RIG-I_Otga ..A----------....C.....C.....C.....AG.....T.
RIG-I_Bota ..TGGGACTGTAACTCTT.....G.....A.T.....G.....G.....A..G.....
RIG-I_Ovar ..TGGGACCGTAACTCTT.....GC.....A.T.....G.....G.....A.CG.....G....
RIG-I_Susc ..TGGTACCATAACTCTT.G.GG.....T.....TA.T..C.....G..G.....C.AG.....A...TG
RIG-I_Mylu ...GGTACCATAAGTCTC.....G.....C...A.T.A.....G.....C...G.....AAG..TG
RIG-I_Ptva ..TGGTACCATAACTCTC.G.....C.....C...A.T.....C.....G.....G.....AAG....
RIG-I_Ptal ..TGGTACCATAACTCTC.G.....C.....C...A.T.....C.....G.....G.....AAG....
RIG-I_Aime ...AGTACCGTAACTCTC.....TG...T.....A.T.....C.....CC..T.T.....A.....
RIG-I_Calu ...GGTACCATAACTCTC.....TG.....A.T.....C...T.T.....G.....
RIG-I_Feca A.T----------...GGTA.....G.....G.....T.....
RIG-I_Eqca ..T----------.....GCT.....C....T.....A..G.....G....
RIG-I_Loaf ..TGGTACTCTAACTCTC....G.....T.....AC..T.....C.....A..G.....G....
RIG-I_Ictr T.T----------.G.....T.....C.....G....C.....G.....C.T..C...GG.....T.....G...
RIG-I_Capo ..T----------.GCT.....T.....G....A.T.....C...G.....T.....G....
RIG-I_Mumu ..T----------.G...GC.T.T.....C..A.....C..C..C.....G.....CGGC..C..C.....G...T
RIG-I_Orcu ..G----------.....C..G..C.....G..C.....G.....CGA...C..T.....G....

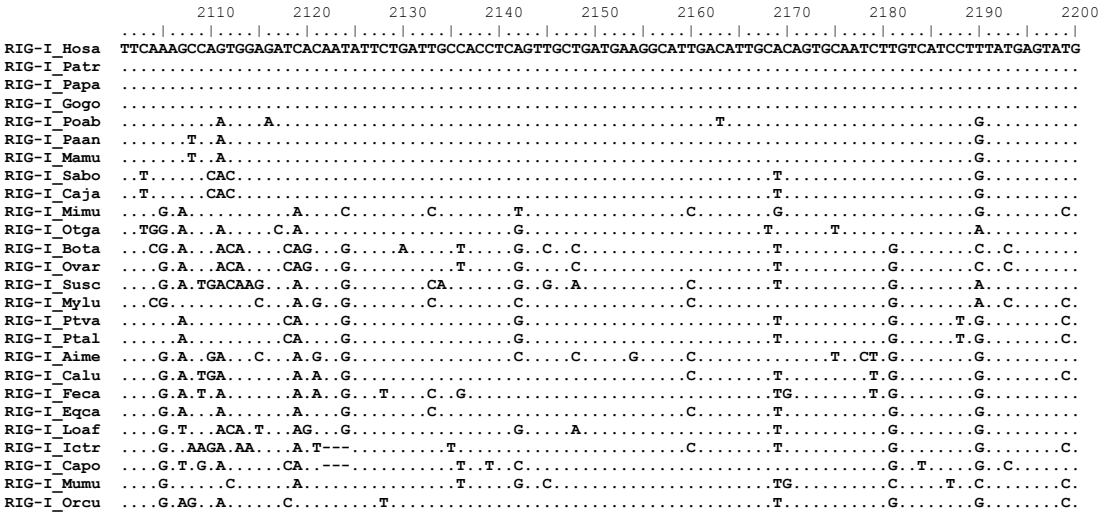

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
RIG-I_Hosa TGGTGTTCCAGATGCCAGACAAAGATGAAGAGAGCAGGATTTGTAAAGCCCTGTTTTTATACACTTCACATTTGCGGAAATATAATGATGCCCTCATTAT
RIG-I_Patr .....
RIG-I_Papa .....
RIG-I_Gogo .....G.....
RIG-I_Poab .....
RIG-I_Paan .....
RIG-I_Mamu .....
RIG-I_Sabo .....G.....G.....
RIG-I_Caja .....G.....
RIG-I_Mimu .....T.G.....A.....A.....C.....G.....G..T.....C.....C.....
RIG-I_Otga CA.....T.G.T.....G...C.....A.....G..T.....C.....G.....
RIG-I_Bota .....T.....G.....A.....G.....G.....C.....
RIG-I_Ovar .....T.....A.....G.....G.....C.....
RIG-I_Susc C.....T.....CA....A.....A.....CG...TG.....T...C..C.....T.....
RIG-I_Mylu CA...C.G..CC.....CA.G.....C...G.....G.....C.....T.....C....
RIG-I_Ptva .....AA.....T.....A.....A.....G.....T.....C..
RIG-I_Ptal .....AA.....T.....A.....A.....G..T.....T.....C..
RIG-I_Aime .TT....T...T....C.....A..G.....A...G.....C..C.....
RIG-I_Calu .....T...T.A..CA.....A.....G.....C..C.....
RIG-I_Feca C.....T...T....T.....A.....G.....C.....
RIG-I_Eqca .A.....T...T....T.....CC.....G....A.C.G.....T..CC.....C.....C.....
RIG-I_Loaf .....T.....G.....A.....G..AT....C..G.....G...C.A.....G..
RIG-I_Ictr .....T.GAC.....G.....C.....A..A...C.G..T.....T.C...A.....T.....
RIG-I_Capo G.T....T.....A.....C.....AA...A...G.....A.....
RIG-I_Mumu CA.....T.....G.....G..G.....C..G.C..C.....G..C..CC.G.....A.....C...C.....A.....C..
RIG-I_Orcu G.AC...T...C.....C.....G..A.A...G.....C.....G..C..C.....

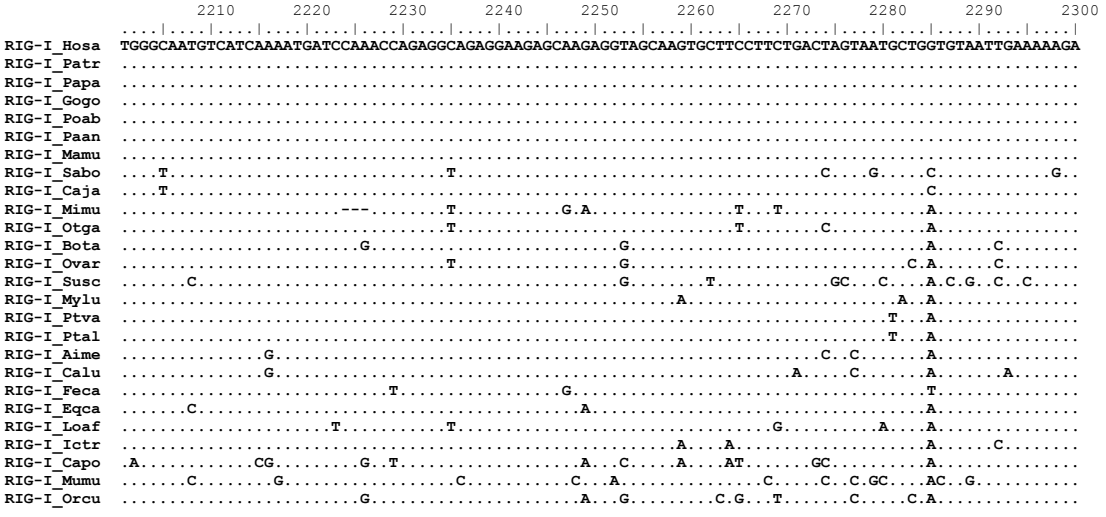

1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
RIG-I_Hosa CAGTGAGCATGCACGAATGAAAGATGCTCTGGATTACTTGAAAGACTTCTTCAGCAATGTCCGAGCAGCAGGATTCGATGAGATTGAGCAAGATCTTACT
RIG-I_Patr .....
RIG-I_Papa .....
RIG-I_Gogo ..A.....
RIG-I_Poab ..A.....T.....
RIG-I_Paan .....G.....T.....C.....C.....
RIG-I_Mamu .....C.....G.....T.....C.....T.....C.....
RIG-I_Sabo T.A....G.C.....G.....C.....G.....C...A.....
RIG-I_Caja .....G.....C.....G.....C.....
RIG-I_Mimu ..A...A.....A.....
RIG-I_Otga ..AC..A.....C.....A.....T.....C.....
RIG-I_Bota ..A....G.....G.....A.....A.....AA.....G.....T....CA.....C...
RIG-I_Ovar .....G.....G.....A.....A.....G.....T....CA.....C...
RIG-I_Susc ..A...A.....C.....T.....A...A.....C..T.....C.....
RIG-I_Mylu ..A.....G.C.....CT.....T.....A.....C...
RIG-I_Ptva .....G.....C.T.....A.....C..T.....C.....C...
RIG-I_Ptal .....G.....C.T.....GA.....C..T.....C.....C...
RIG-I_Aime ..A.....C..G.....GCT.....T.....GC....C..C
RIG-I_Calu ..A.....CTG.....T.....GC....C...
RIG-I_Feca ..A.....G.....C...C.....G.....T.....C...
RIG-I_Eqca .....G.....AT.....G.....T.....CC.....C...
RIG-I_Loaf ..A....G....C.....C.....T.....T.C.....CT.....T.....A.....C...
RIG-I_Ictr .....G.....A.....A.....A.....A....G.....T.....C.....
RIG-I_Capo ...C...G.....A.....C...T..A.....C....G..T.....T.....A.....CT..T.....C.....C...
RIG-I_Mumu .....G.....AG....C...C.....AA.....C.C....C.....CA.G.....A.....C.....CC....G...G.....
RIG-I_Orcu .....C....A.....C...A.....A..CA.T.....AAG.....T.....C...A.....C

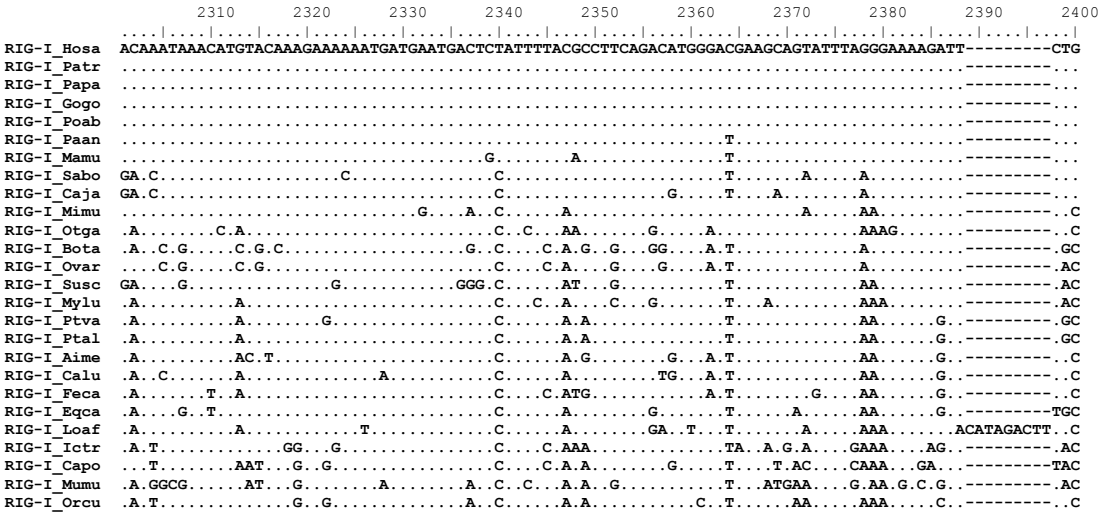

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
RIG-I_Hosa CAGAGATTTGAAGAAAAG------CTGCAGGAACTAGAAAGTGTTTCCAGGGATCCCAGCAATGAGAATCCTAAACTTGAAGACCTCTGCTTCATCTTAC
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

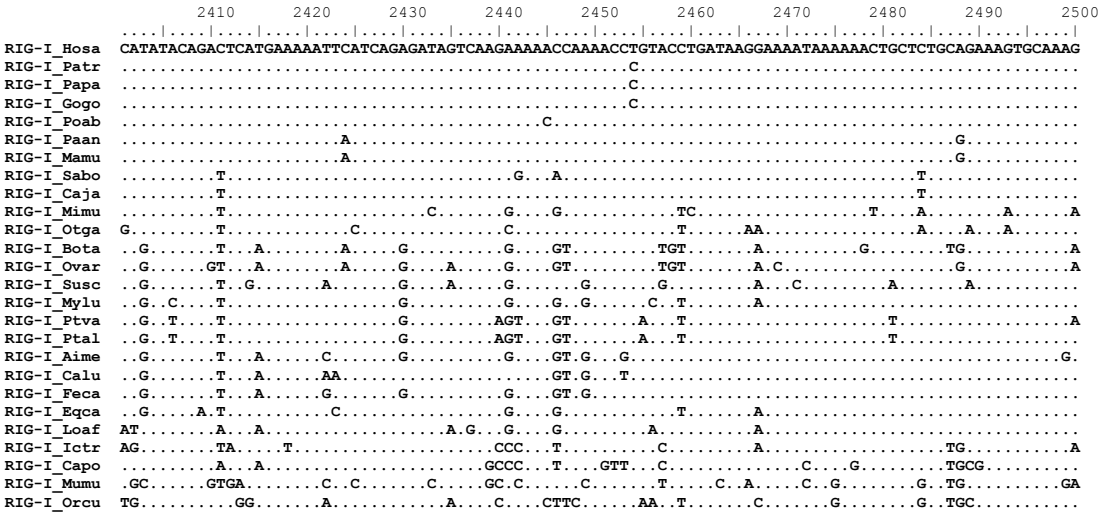

1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
RIG-I_Hosa AAGAAGAGTACCACTTAAACCCAGAGACAATAACAATTCTCTTTGTGAAAACCAGAGCACTTGTGGACGCTTTAAAAAATTGGATTGAAGGAAATCCTAA
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

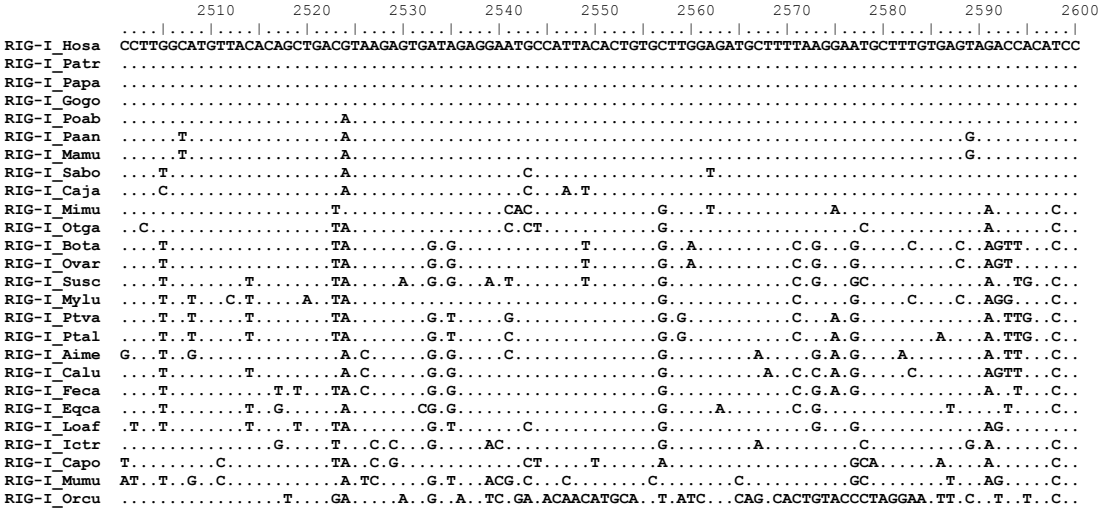

2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
RIG-I_Hosa ACTCAGTTTTCTAAAACCTGGCATATTGACTGGACGTGGCAAAACAAATCAGAACACAGGA---ATGACCCTCCCGGCACAGAAGTGTATATTGGATGCA
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

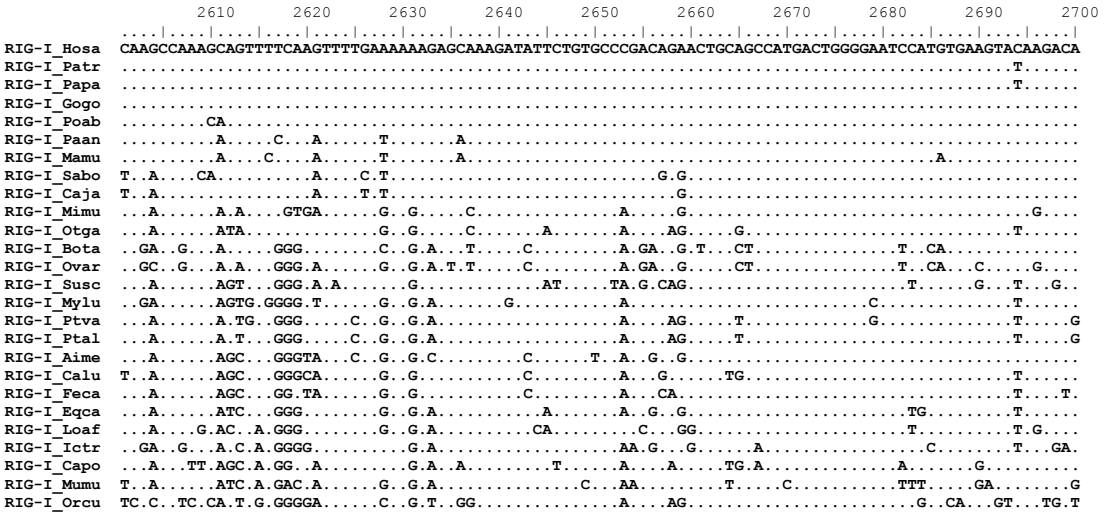

2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
RIG-I_Hosa TTCAAAGCCAGTGGAGATCACAATATTCTGATTGCCACCTCAGTTGCTGATGAAGGCATTGACATTGCACAGTGCAATCTTGTCATCCTTTATGAGTATG
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

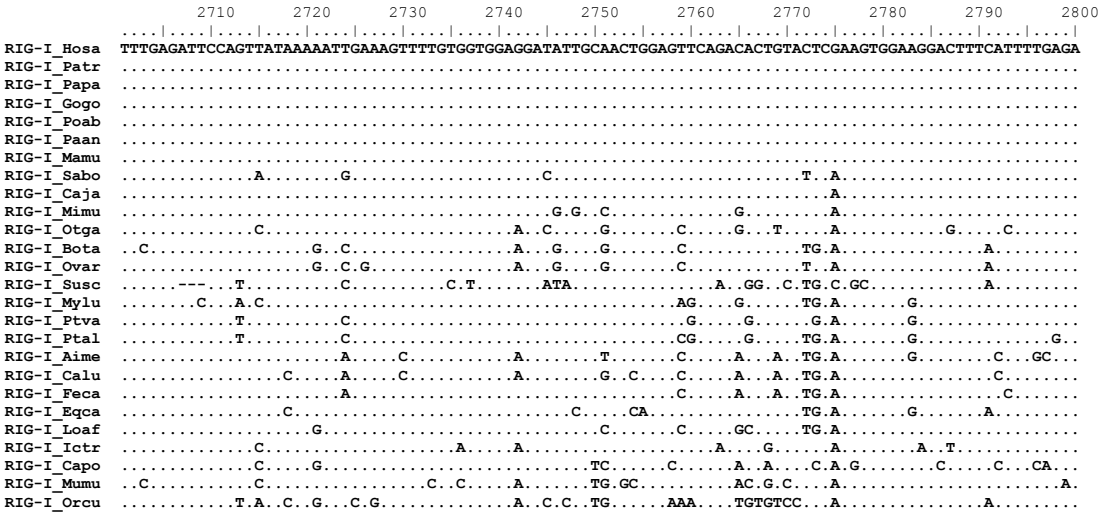

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
RIG-I_Hosa TGGGCAATGTCATCAAAATGATCCAAACCAGAGGCAGAGGAAGAGCAAGAGGTAGCAAGTGCTTCCTTCTGACTAGTAATGCTGGTGTAATTGAAAAAGA
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu


2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
RIG-I_Hosa ACAAATAAACATGTACAAAGAAAAAATGATGAATGACTCTATTTTACGCCTTCAGACATGGGACGAAGCAGTATTTAGGGAAAAGATT---------CTG
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

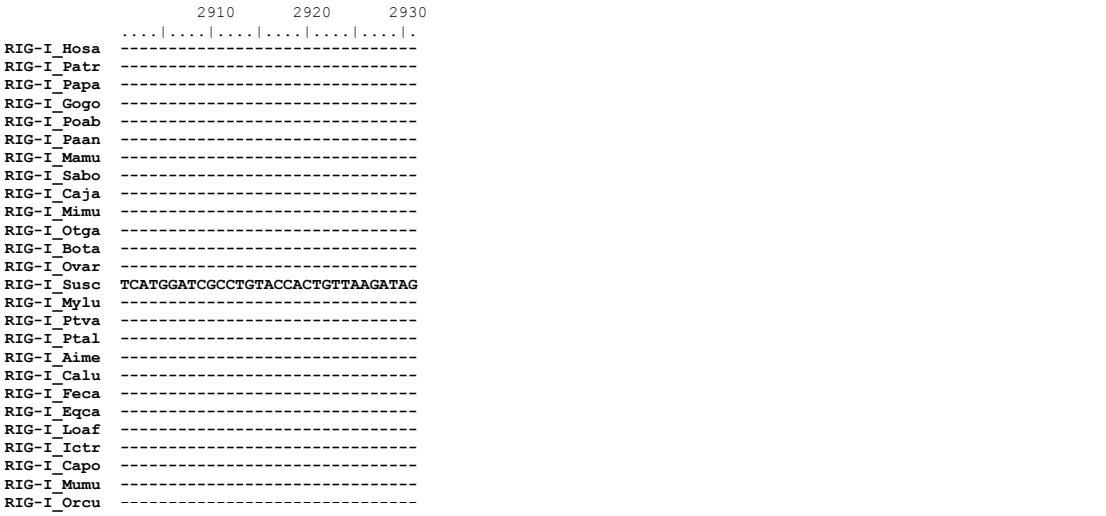

2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
RIG-I_Hosa CATATACAGACTCATGAAAAATTCATCAGAGATAGTCAAGAAAAACCAAAACCTGTACCTGATAAGGAAAATAAAAAACTGCTCTGCAGAAAGTGCAAAG
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu


2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
RIG-I_Hosa CCTTGGCATGTTACACAGCTGACGTAAGAGTGATAGAGGAATGCCATTACACTGTGCTTGGAGATGCTTTTAAGGAATGCTTTGTGAGTAGACCACATCC
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

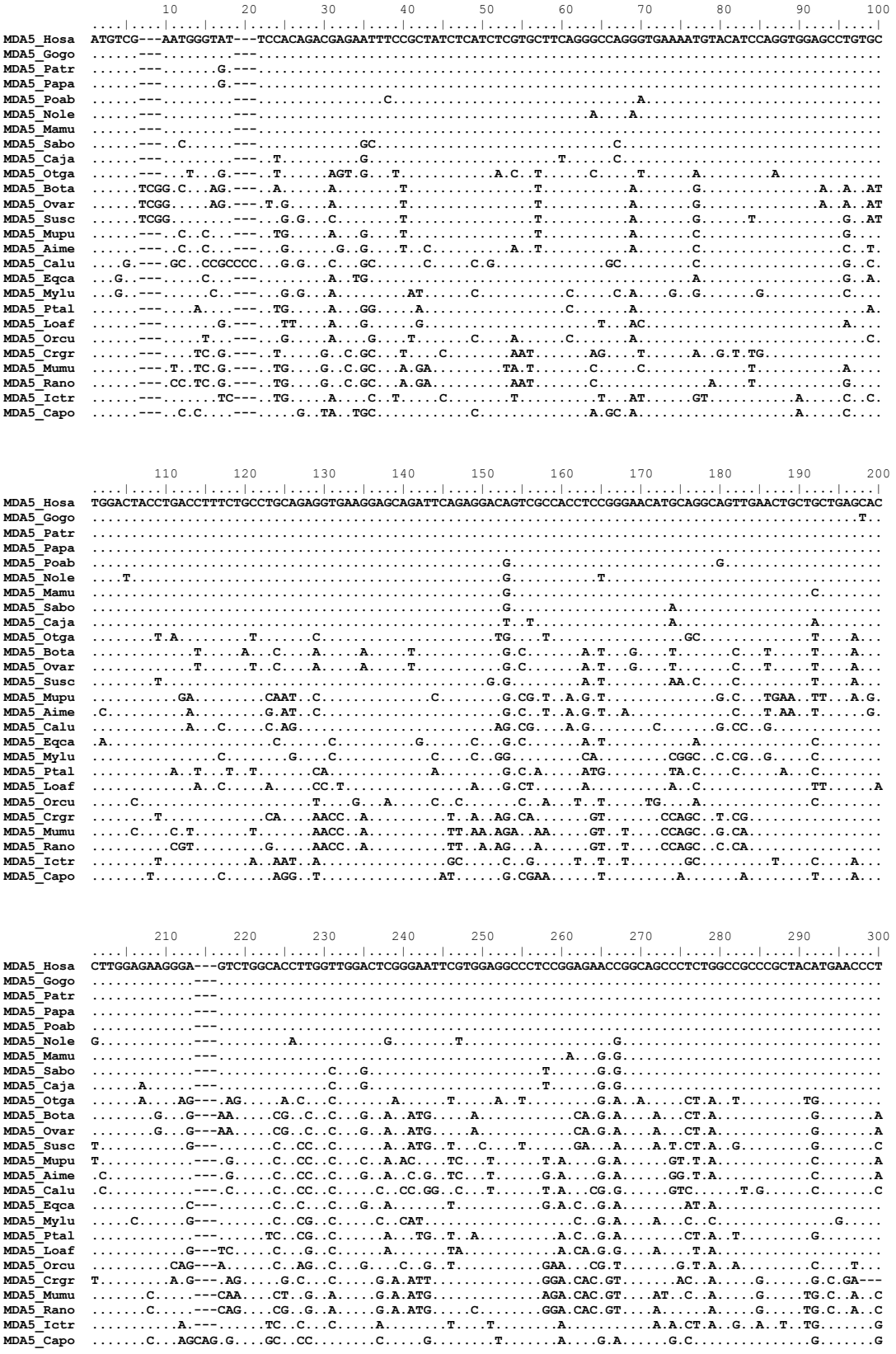

2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
RIG-I_Hosa CAAGCCAAAGCAGTTTTCAAGTTTTGAAAAAAGAGCAAAGATATTCTGTGCCCGACAGAACTGCAGCCATGACTGGGGAATCCATGTGAAGTACAAGACA
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

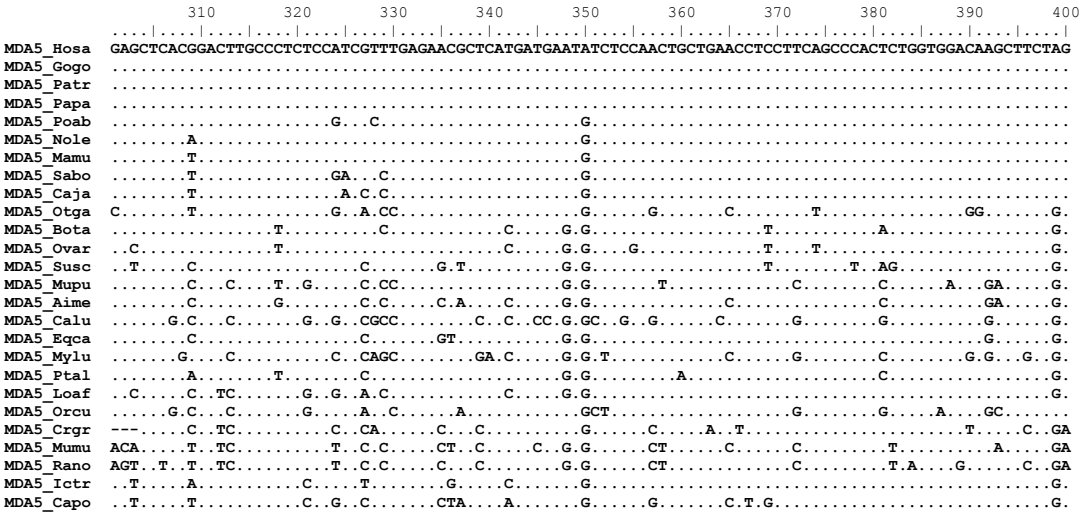

2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
RIG-I_Hosa TTTGAGATTCCAGTTATAAAAATTGAAAGTTTTGTGGTGGAGGATATTGCAACTGGAGTTCAGACACTGTACTCGAAGTGGAAGGACTTTCATTTTGAGA
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

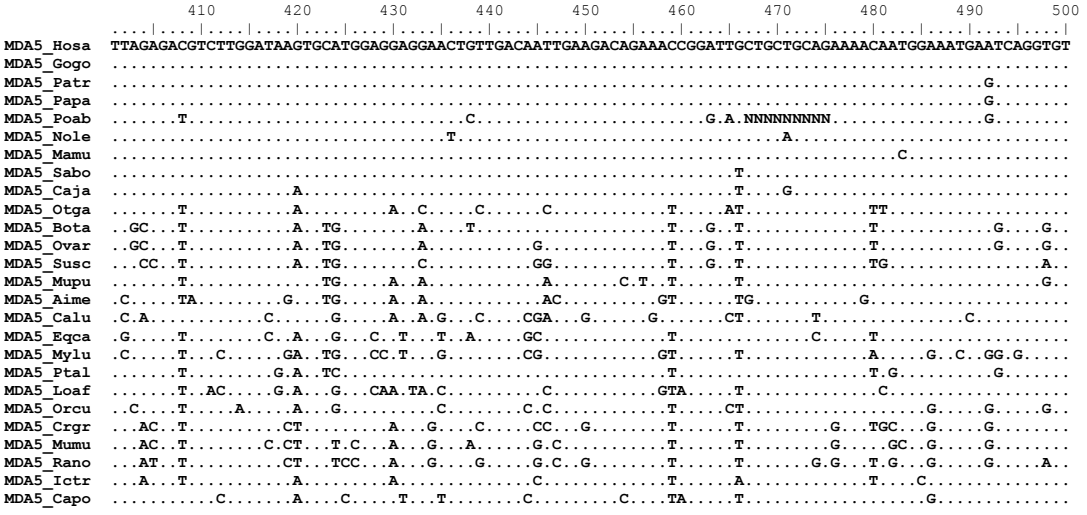

2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
RIG-I_Hosa AGATACCATTTGATCCAGCAGAAATGTCCAAATGA
RIG-I_Sabo CGAT.TCAGAGCCTCAACCTTCAGCTTCCGGGAATGAGTAA
RIG-I_Bota CCT..GGCTCAGGACCTCAATCTGCAGGGAGTGGATGGCCTTGAATGA
RIG-I_Ovar GCTG.TGCTCAGGACCTCAATCTGCAGGCAATGAATGGCCTTGAATGA
RIG-I_Susc ..C.CT.....G.T.....G..GGTG..GCTCAGGACATGGGTCTTCAGGGATGGGCAACCTTGAGTGAAGGAGAAATTGGACTGGGGCTAAA
RIG-I_Mylu ..C.....T.
RIG-I_Ptva .AC.C.....G.T.....C.
RIG-I_Ptal .AC.C.....G.T.....C.
RIG-I_Aime .....C.....G.T.
RIG-I_Calu .....C....G.T.TC.....AC..G.
RIG-I_Feca .....G.TAAG.....C.
RIG-I_Eqca .....G.T.....C...
RIG-I_Loaf .....G.T.....A.
RIG-I_Ictr .....AG.....G.T.....TATC.A..TGACTTCAGGGCCTGA
RIG-I_Capo .....GA.....G.TA.....T.
RIG-I_Mumu G.....AG..C.....T.....GT.
RIG-I_Orcu .....A...T.A...C....TG.

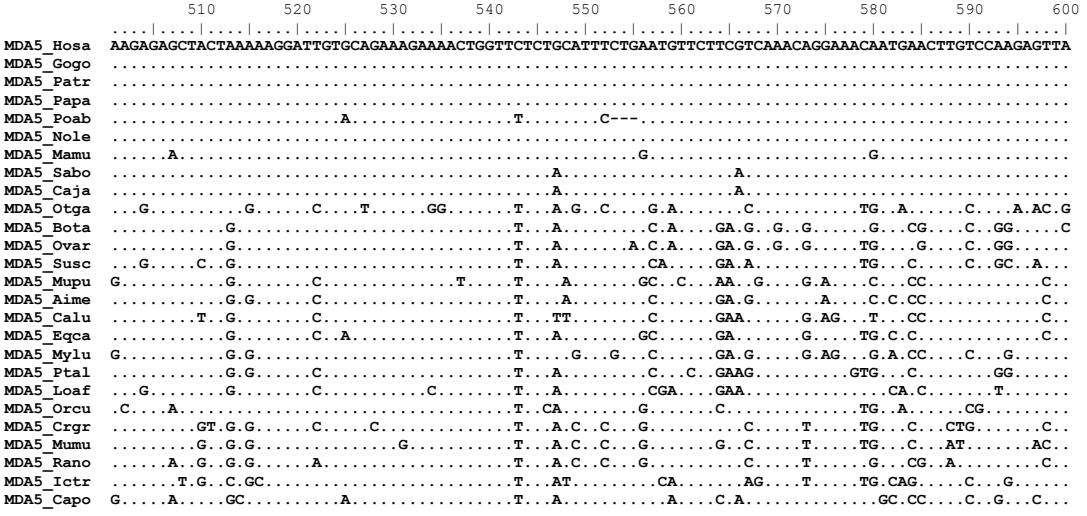

2910 2920 2930
RIG-I_Hosa
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc TCATGGATCGCCTGTACCACTGTTAAGATAG
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

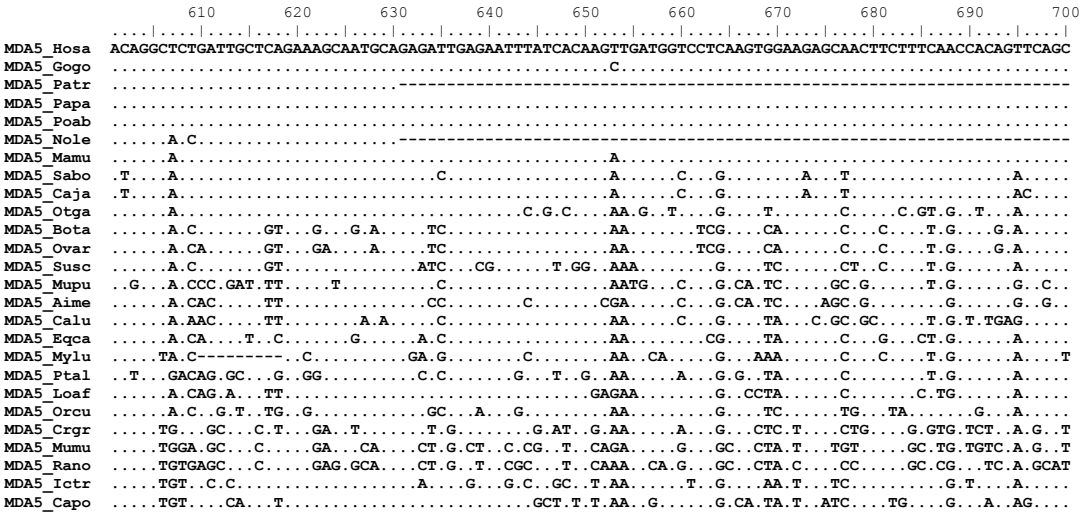
Supplementary Figure S2

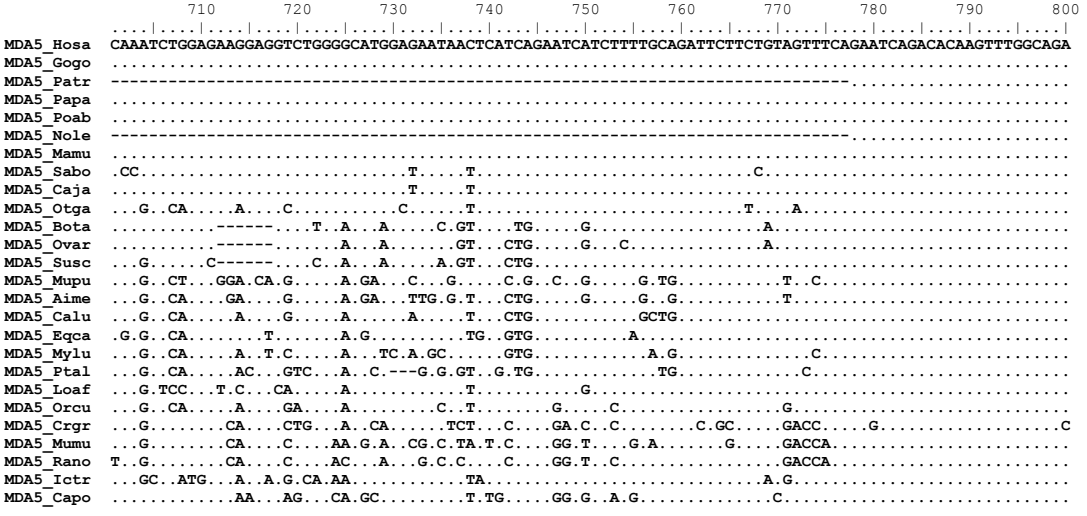


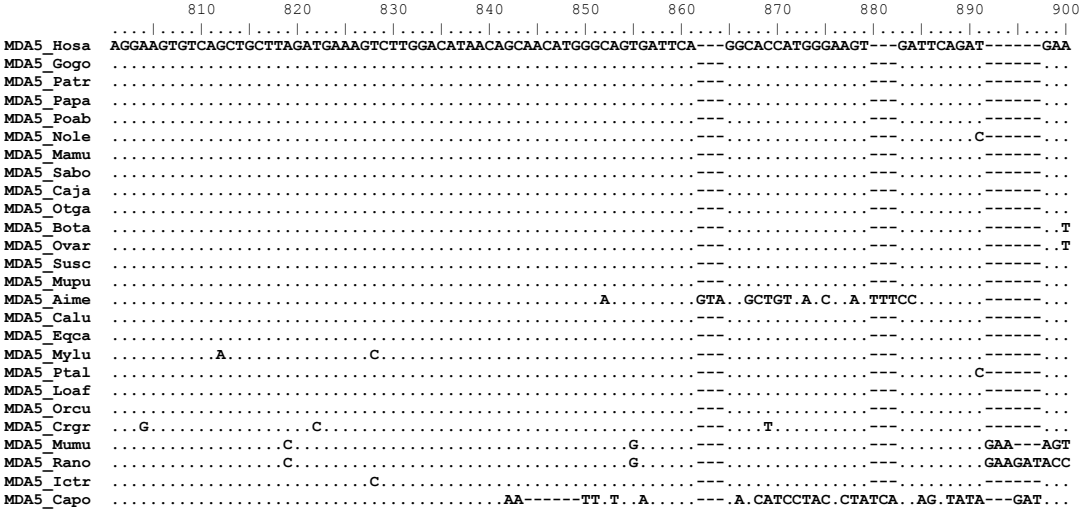

310 320 330 340 350 360 370 380 390 400
MDA5_Hosa GAGCTCACGGACTTGCCCTCTCCATCGTTTGAGAACGCTCATGATGAATATCTCCAACTGCTGAACCTCCTTCAGCCCACTCTGGTGGACAAGCTTCTAG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

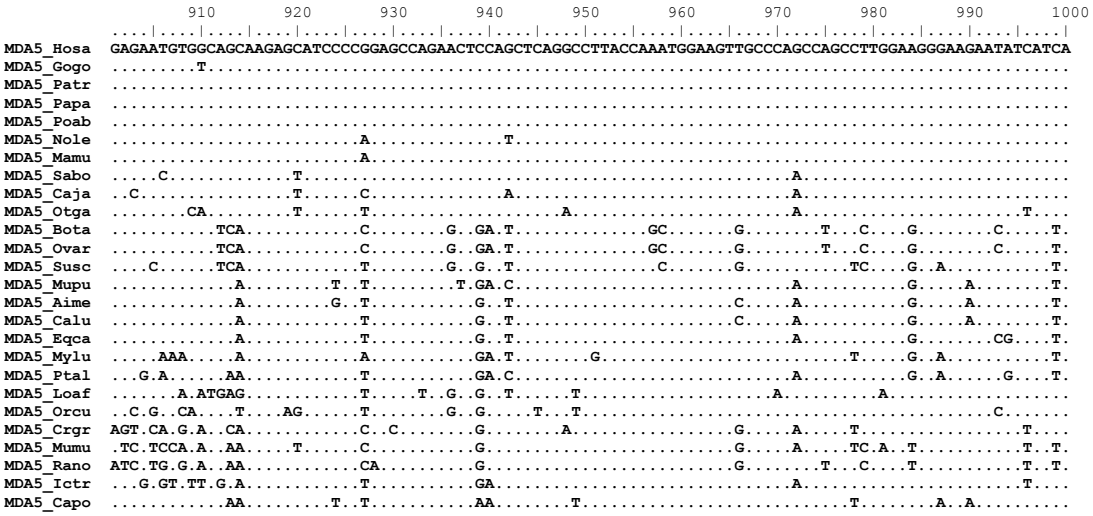

410 420 430 440 450 460 470 480 490 500
MDA5_Hosa TTAGAGACGTCTTGGATAAGTGCATGGAGGAGGAACTGTTGACAATTGAAGACAGAAACCGGATTGCTGCTGCAGAAAACAATGGAAATGAATCAGGTGT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

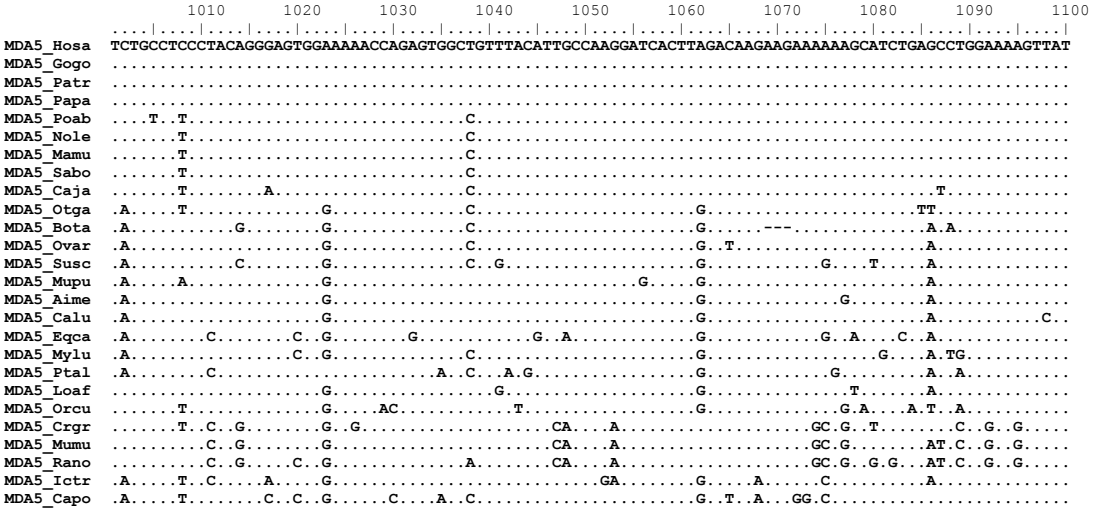

510 520 530 540 550 560 570 580 590 600
MDA5_Hosa AAGAGAGCTACTAAAAAGGATTGTGCAGAAAGAAAACTGGTTCTCTGCATTTCTGAATGTTCTTCGTCAAACAGGAAACAATGAACTTGTCCAAGAGTTA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

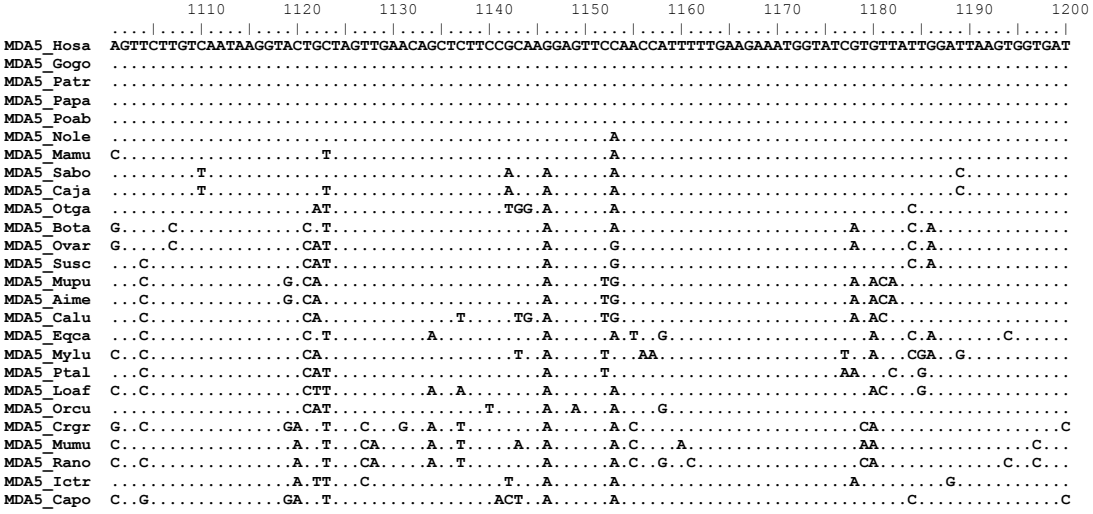

610 620 630 640 650 660 670 680 690 700
MDA5_Hosa ACAGGCTCTGATTGCTCAGAAAGCAATGCAGAGATTGAGAATTTATCACAAGTTGATGGTCCTCAAGTGGAAGAGCAACTTCTTTCAACCACAGTTCAGC
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

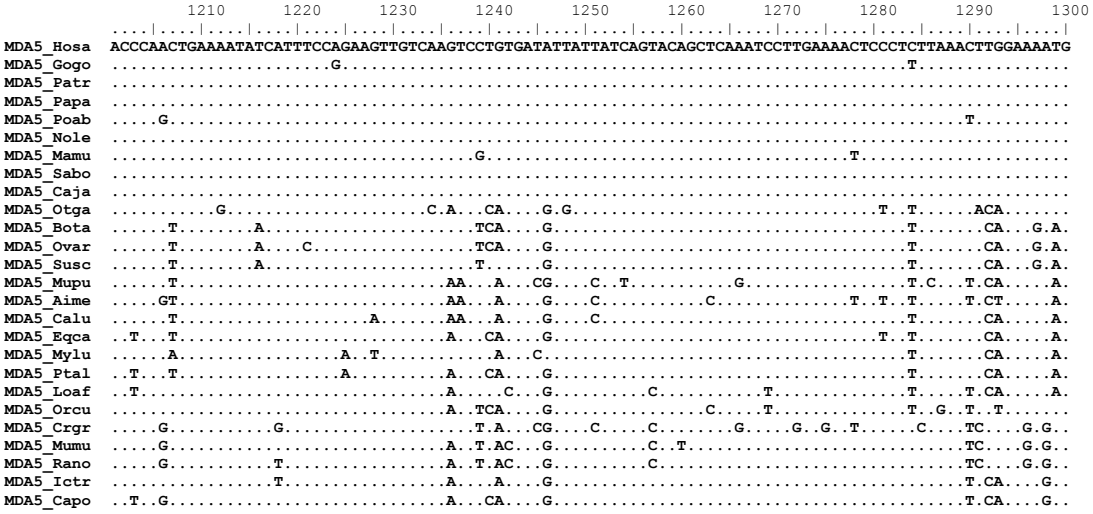

710 720 730 740 750 760 770 780 790 800
MDA5_Hosa CAAATCTGGAGAAGGAGGTCTGGGGCATGGAGAATAACTCATCAGAATCATCTTTTGCAGATTCTTCTGTAGTTTCAGAATCAGACACAAGTTTGGCAGA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

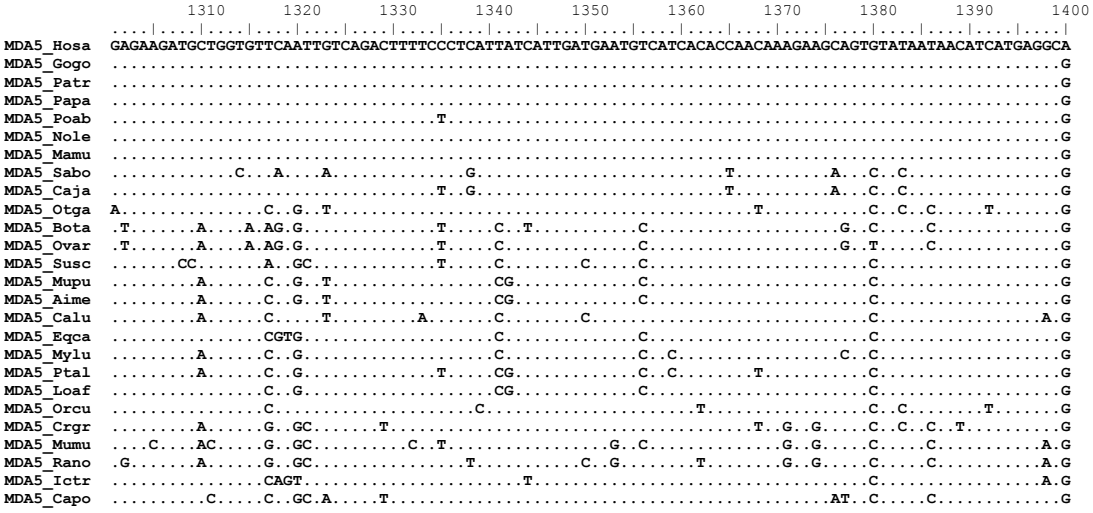

810 820 830 840 850 860 870 880 890 900
MDA5_Hosa AGGAAGTGTCAGCTGCTTAGATGAAAGTCTTGGACATAACAGCAACATGGGCAGTGATTCA---GGCACCATGGGAAGT---GATTCAGAT------GAA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

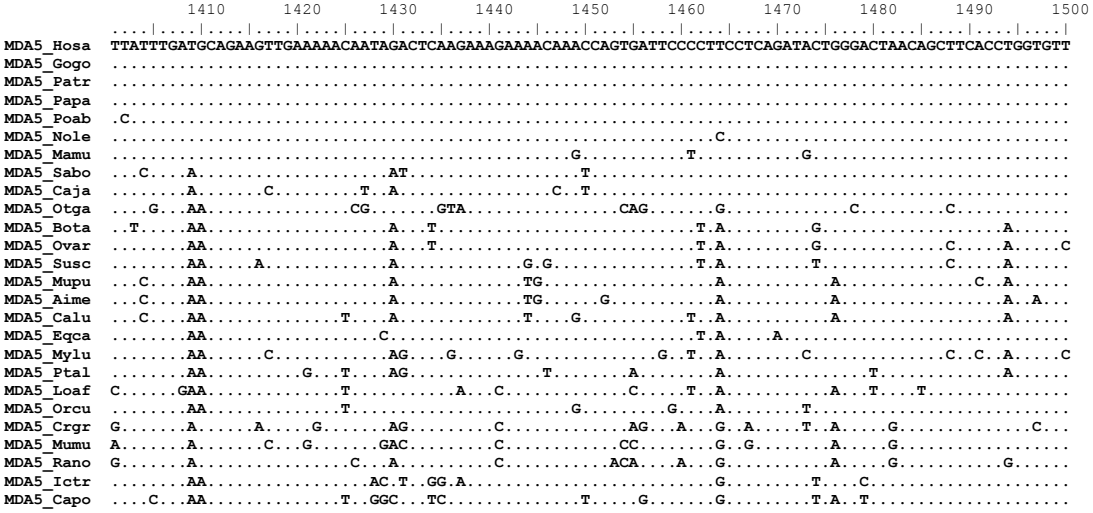

910 920 930 940 950 960 970 980 990 1000
MDA5_Hosa GAGAATGTGGCAGCAAGAGCATCCCCGGAGCCAGAACTCCAGCTCAGGCCTTACCAAATGGAAGTTGCCCAGCCAGCCTTGGAAGGGAAGAATATCATCA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

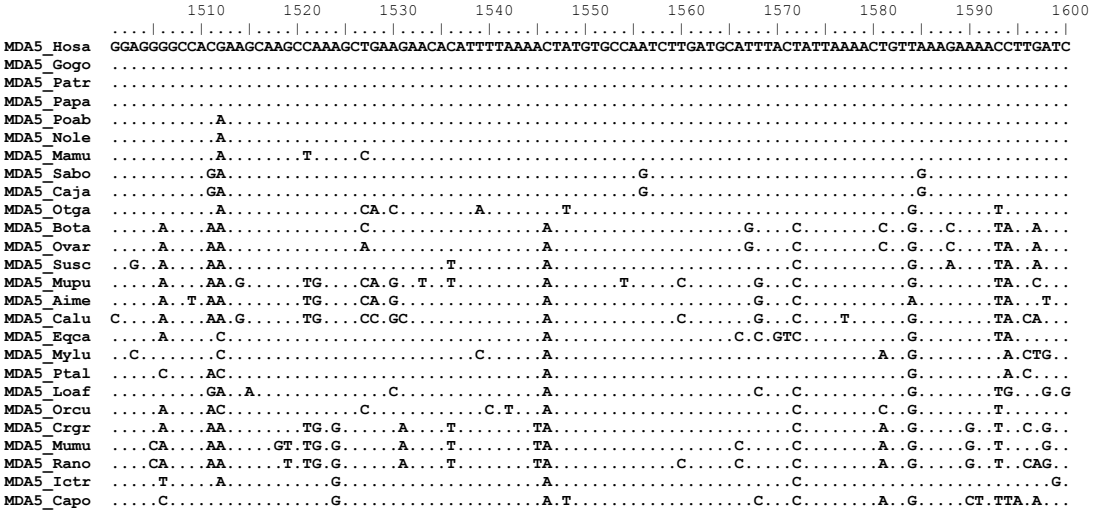

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
MDA5_Hosa TCTGCCTCCCTACAGGGAGTGGAAAAACCAGAGTGGCTGTTTACATTGCCAAGGATCACTTAGACAAGAAGAAAAAAGCATCTGAGCCTGGAAAAGTTAT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

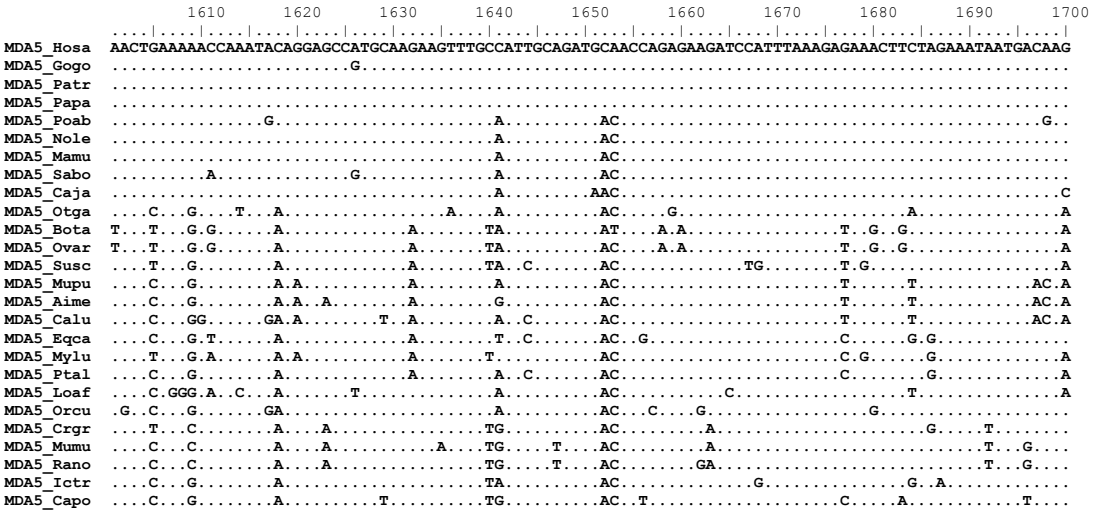

1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
MDA5_Hosa AGTTCTTGTCAATAAGGTACTGCTAGTTGAACAGCTCTTCCGCAAGGAGTTCCAACCATTTTTGAAGAAATGGTATCGTGTTATTGGATTAAGTGGTGAT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

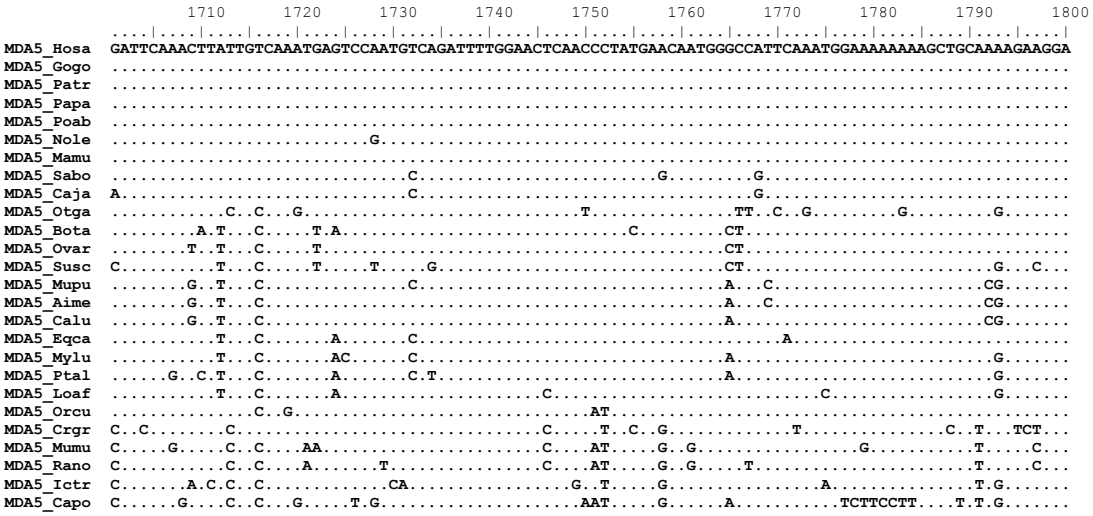

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
MDA5_Hosa ACCCAACTGAAAATATCATTTCCAGAAGTTGTCAAGTCCTGTGATATTATTATCAGTACAGCTCAAATCCTTGAAAACTCCCTCTTAAACTTGGAAAATG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

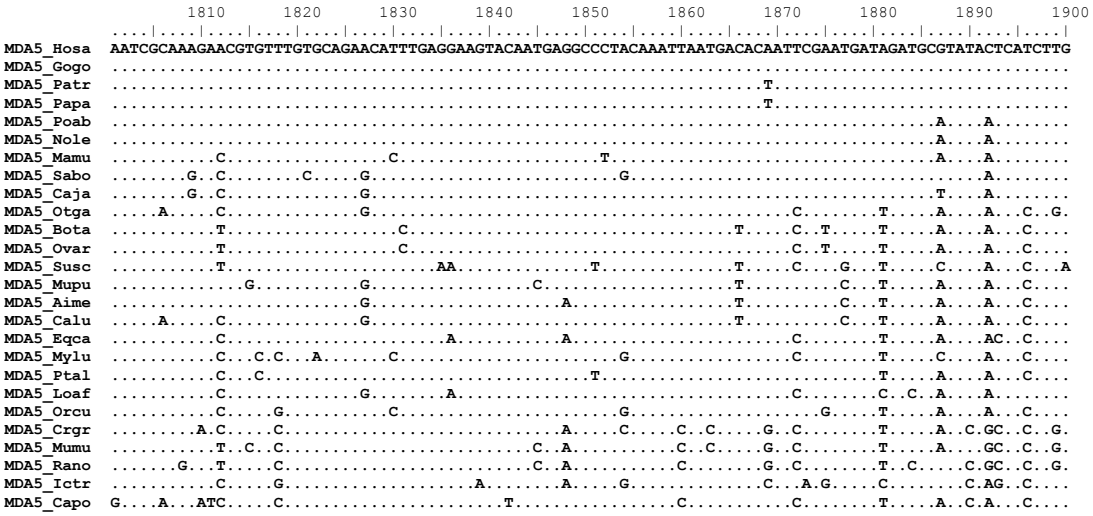

1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
MDA5_Hosa GAGAAGATGCTGGTGTTCAATTGTCAGACTTTTCCCTCATTATCATTGATGAATGTCATCACACCAACAAAGAAGCAGTGTATAATAACATCATGAGGCA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

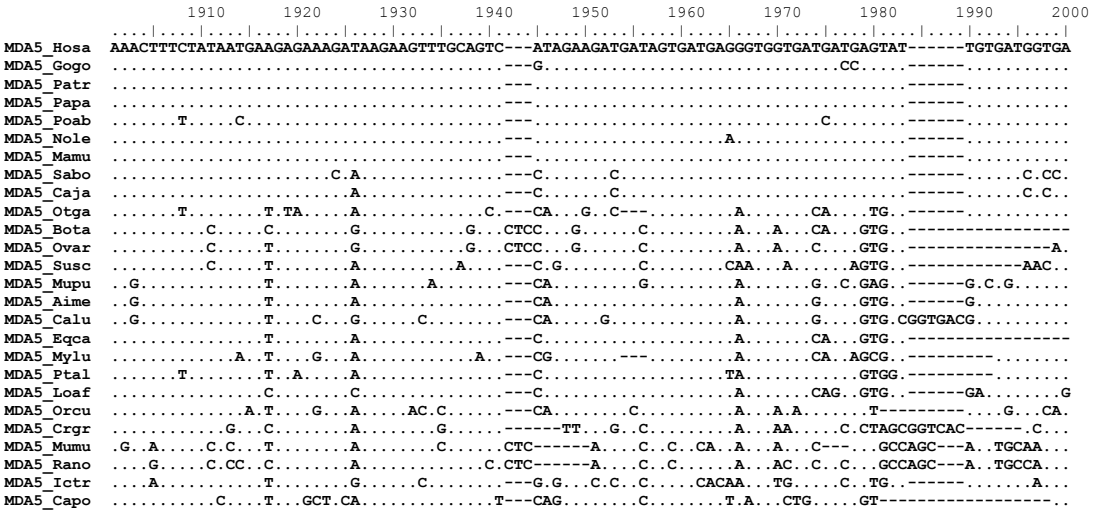

1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
MDA5_Hosa TTATTTGATGCAGAAGTTGAAAAACAATAGACTCAAGAAAGAAAACAAACCAGTGATTCCCCTTCCTCAGATACTGGGACTAACAGCTTCACCTGGTGTT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

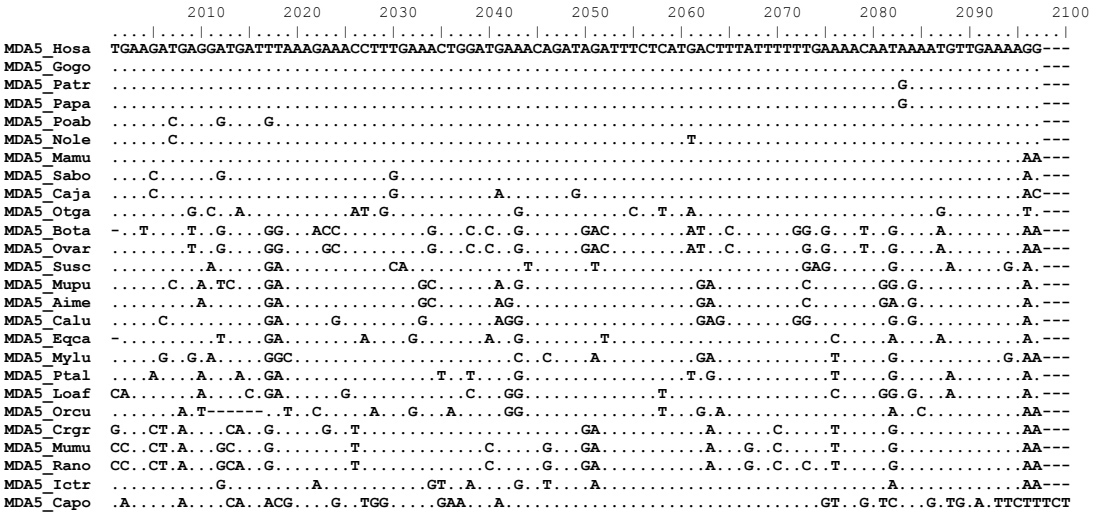

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
MDA5_Hosa GGAGGGGCCACGAAGCAAGCCAAAGCTGAAGAACACATTTTAAAACTATGTGCCAATCTTGATGCATTTACTATTAAAACTGTTAAAGAAAACCTTGATC
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

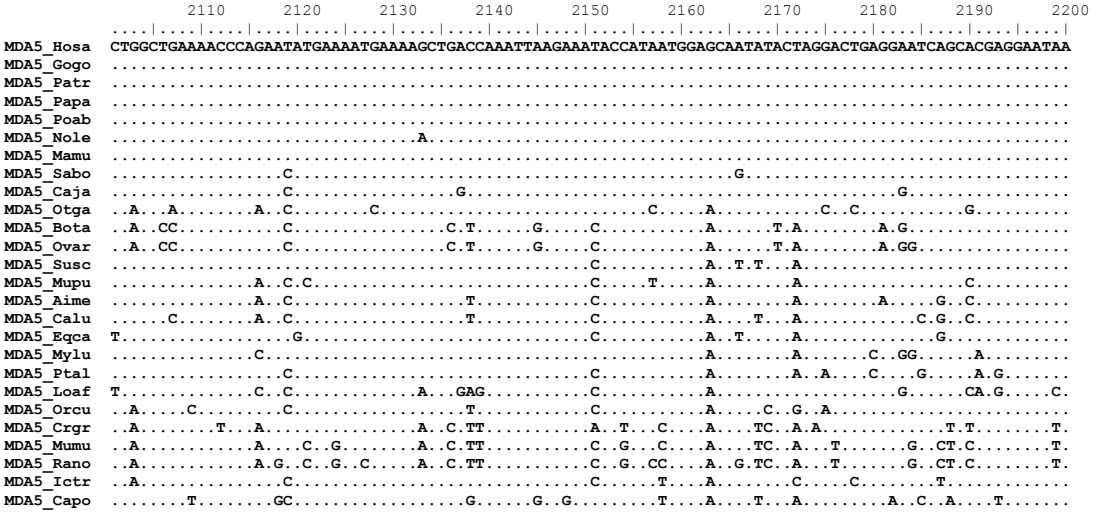

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
MDA5_Hosa AACTGAAAAACCAAATACAGGAGCCATGCAAGAAGTTTGCCATTGCAGATGCAACCAGAGAAGATCCATTTAAAGAGAAACTTCTAGAAATAATGACAAG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

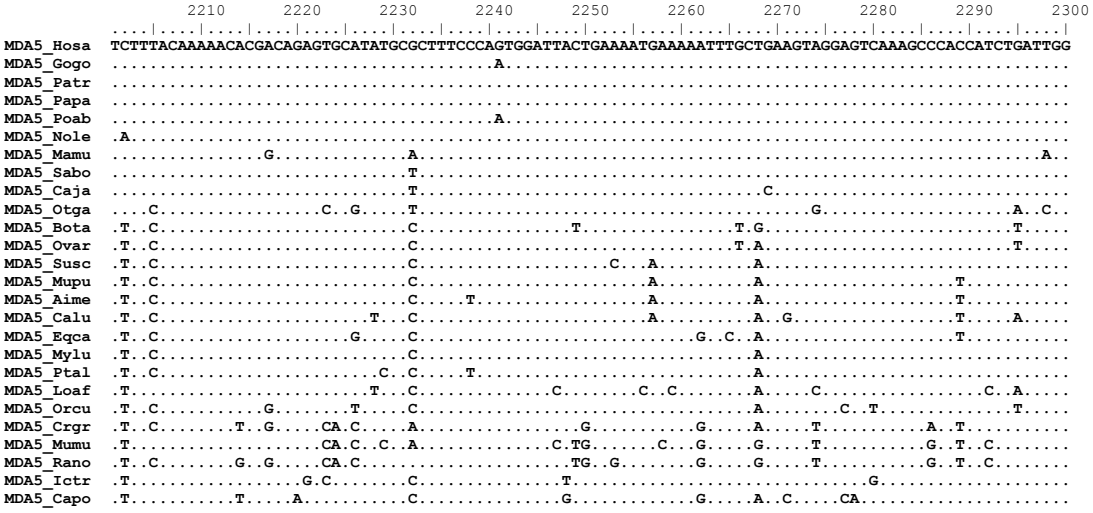

1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
MDA5_Hosa GATTCAAACTTATTGTCAAATGAGTCCAATGTCAGATTTTGGAACTCAACCCTATGAACAATGGGCCATTCAAATGGAAAAAAAAGCTGCAAAAGAAGGA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

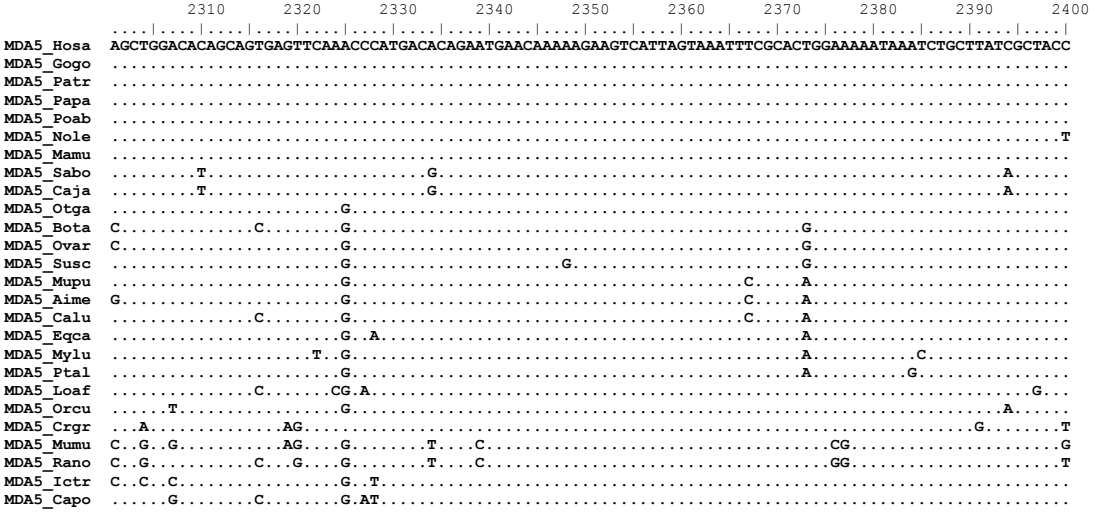

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
MDA5_Hosa AATCGCAAAGAACGTGTTTGTGCAGAACATTTGAGGAAGTACAATGAGGCCCTACAAATTAATGACACAATTCGAATGATAGATGCGTATACTCATCTTG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

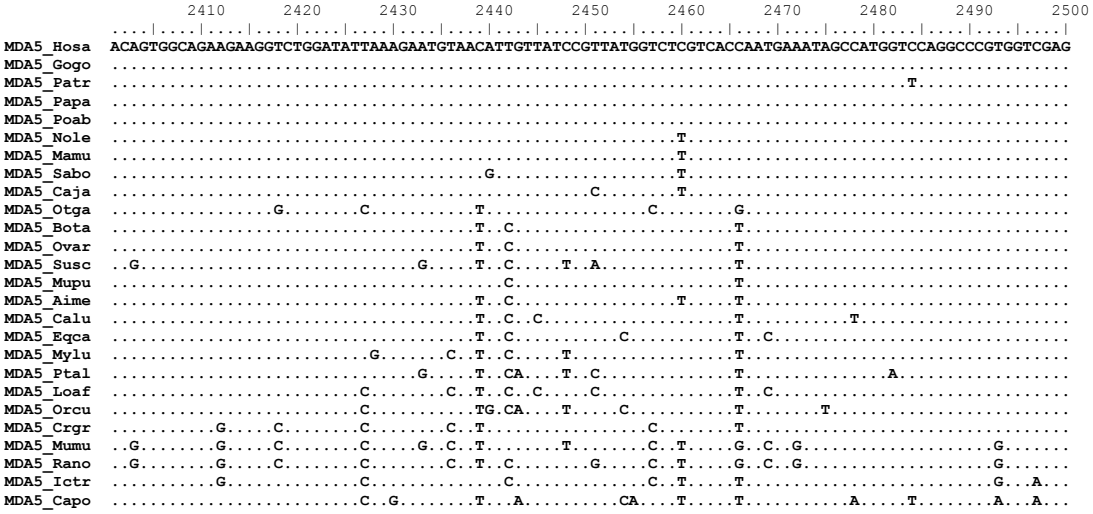

1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
MDA5_Hosa AAACTTTCTATAATGAAGAGAAAGATAAGAAGTTTGCAGTC---ATAGAAGATGATAGTGATGAGGGTGGTGATGATGAGTAT------TGTGATGGTGA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

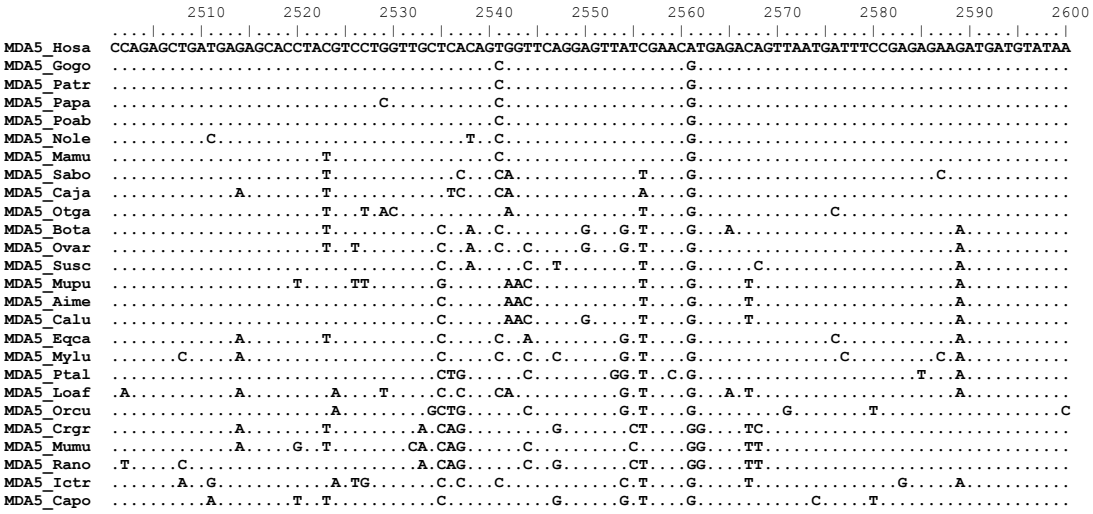

2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
MDA5_Hosa TGAAGATGAGGATGATTTAAAGAAACCTTTGAAACTGGATGAAACAGATAGATTTCTCATGACTTTATTTTTTGAAAACAATAAAATGTTGAAAAGG---
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

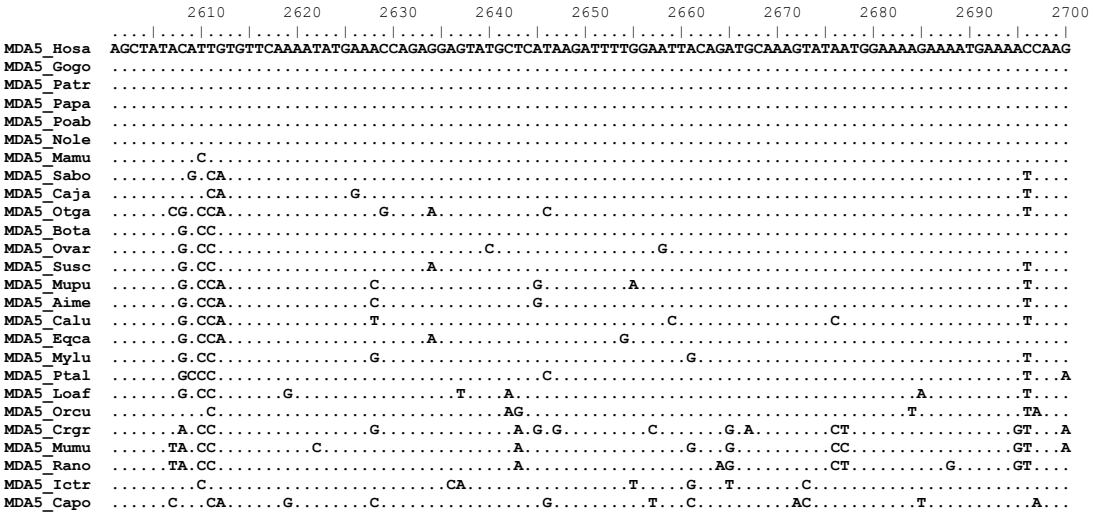

2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
MDA5_Hosa CTGGCTGAAAACCCAGAATATGAAAATGAAAAGCTGACCAAATTAAGAAATACCATAATGGAGCAATATACTAGGACTGAGGAATCAGCACGAGGAATAA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

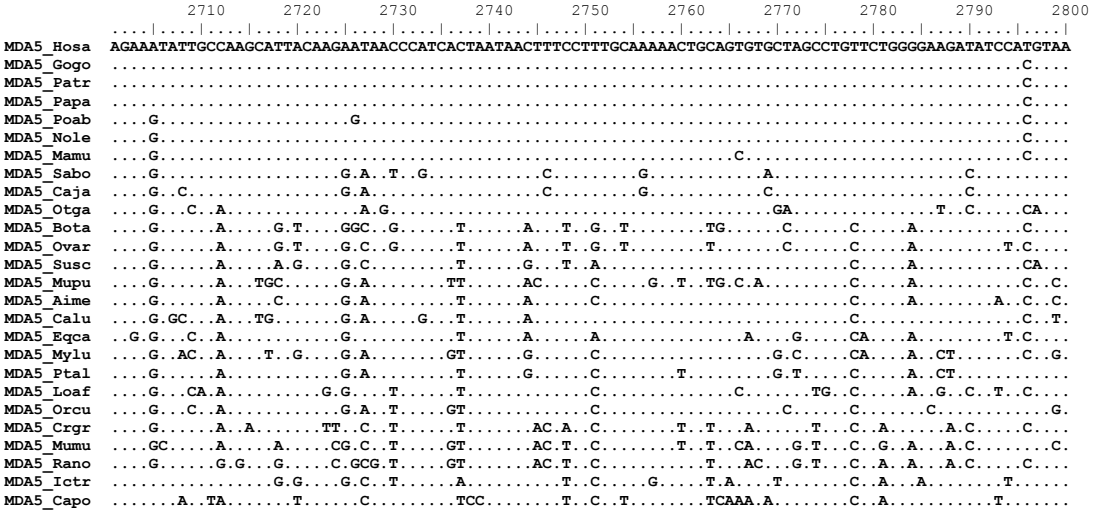

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
MDA5_Hosa TCTTTACAAAAACACGACAGAGTGCATATGCGCTTTCCCAGTGGATTACTGAAAATGAAAAATTTGCTGAAGTAGGAGTCAAAGCCCACCATCTGATTGG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

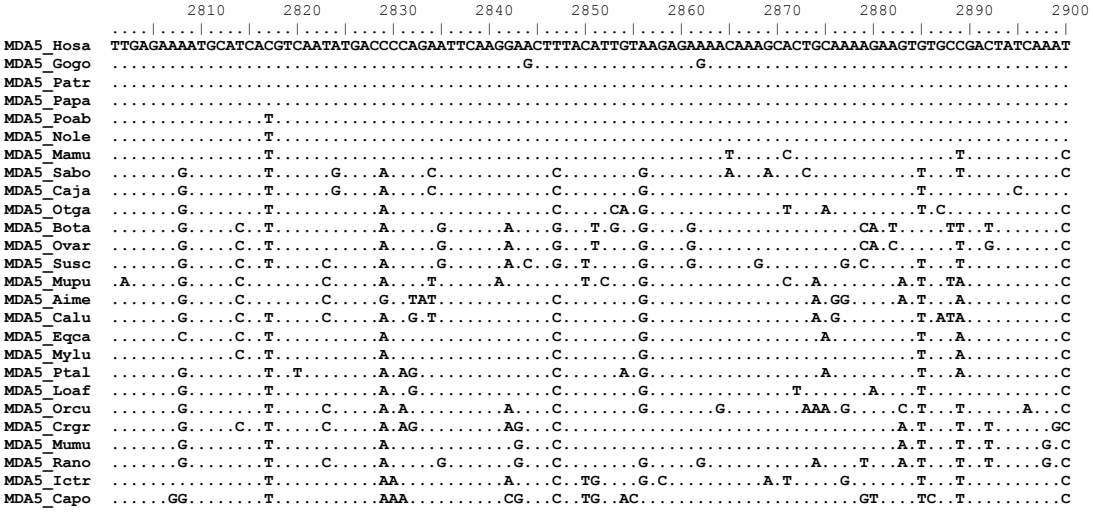

2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
MDA5_Hosa AGCTGGACACAGCAGTGAGTTCAAACCCATGACACAGAATGAACAAAAAGAAGTCATTAGTAAATTTCGCACTGGAAAAATAAATCTGCTTATCGCTACC
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

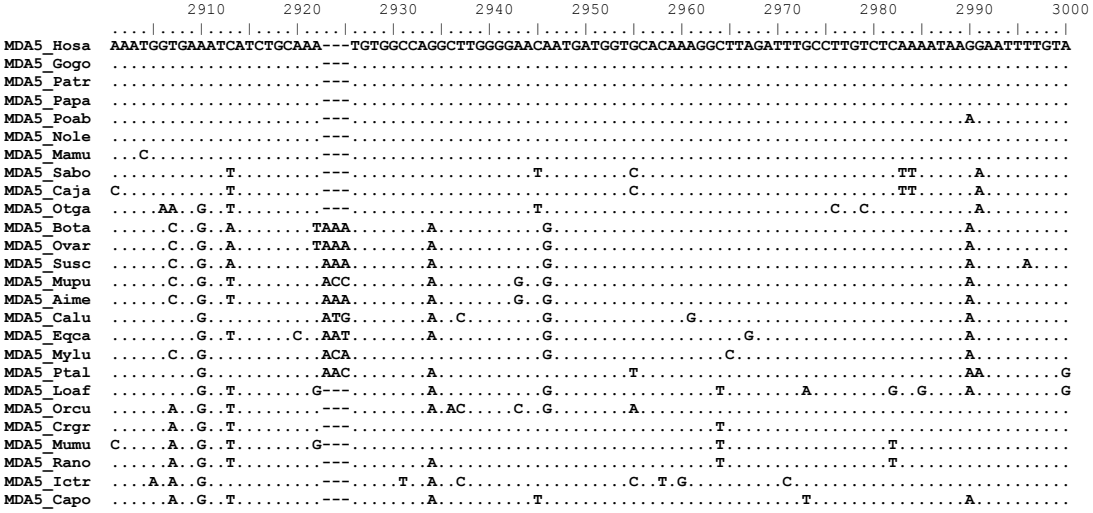

2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
MDA5_Hosa ACAGTGGCAGAAGAAGGTCTGGATATTAAAGAATGTAACATTGTTATCCGTTATGGTCTCGTCACCAATGAAATAGCCATGGTCCAGGCCCGTGGTCGAG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

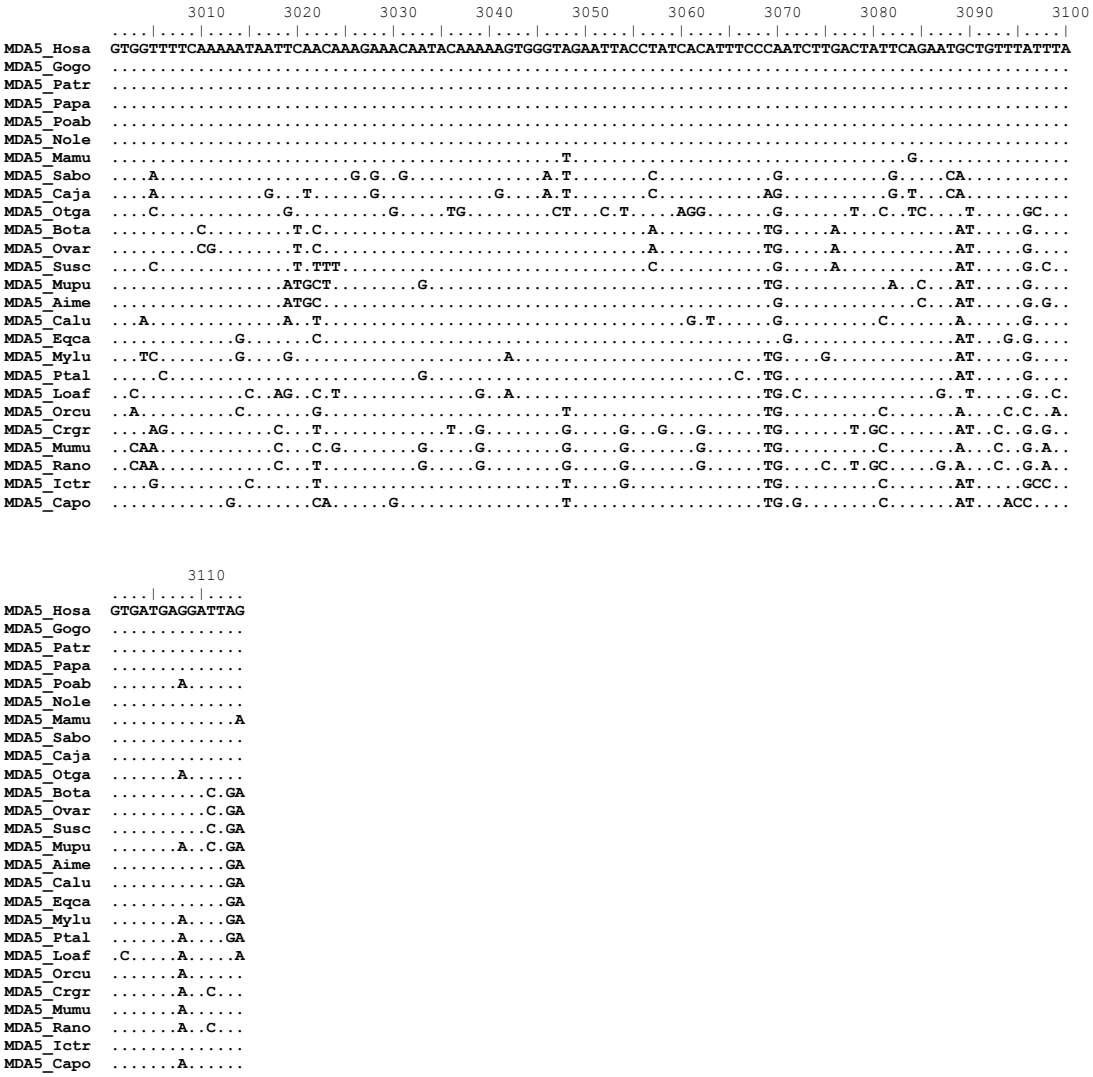

2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
MDA5_Hosa CCAGAGCTGATGAGAGCACCTACGTCCTGGTTGCTCACAGTGGTTCAGGAGTTATCGAACATGAGACAGTTAATGATTTCCGAGAGAAGATGATGTATAA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo


2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
MDA5_Hosa AGCTATACATTGTGTTCAAAATATGAAACCAGAGGAGTATGCTCATAAGATTTTGGAATTACAGATGCAAAGTATAATGGAAAAGAAAATGAAAACCAAG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

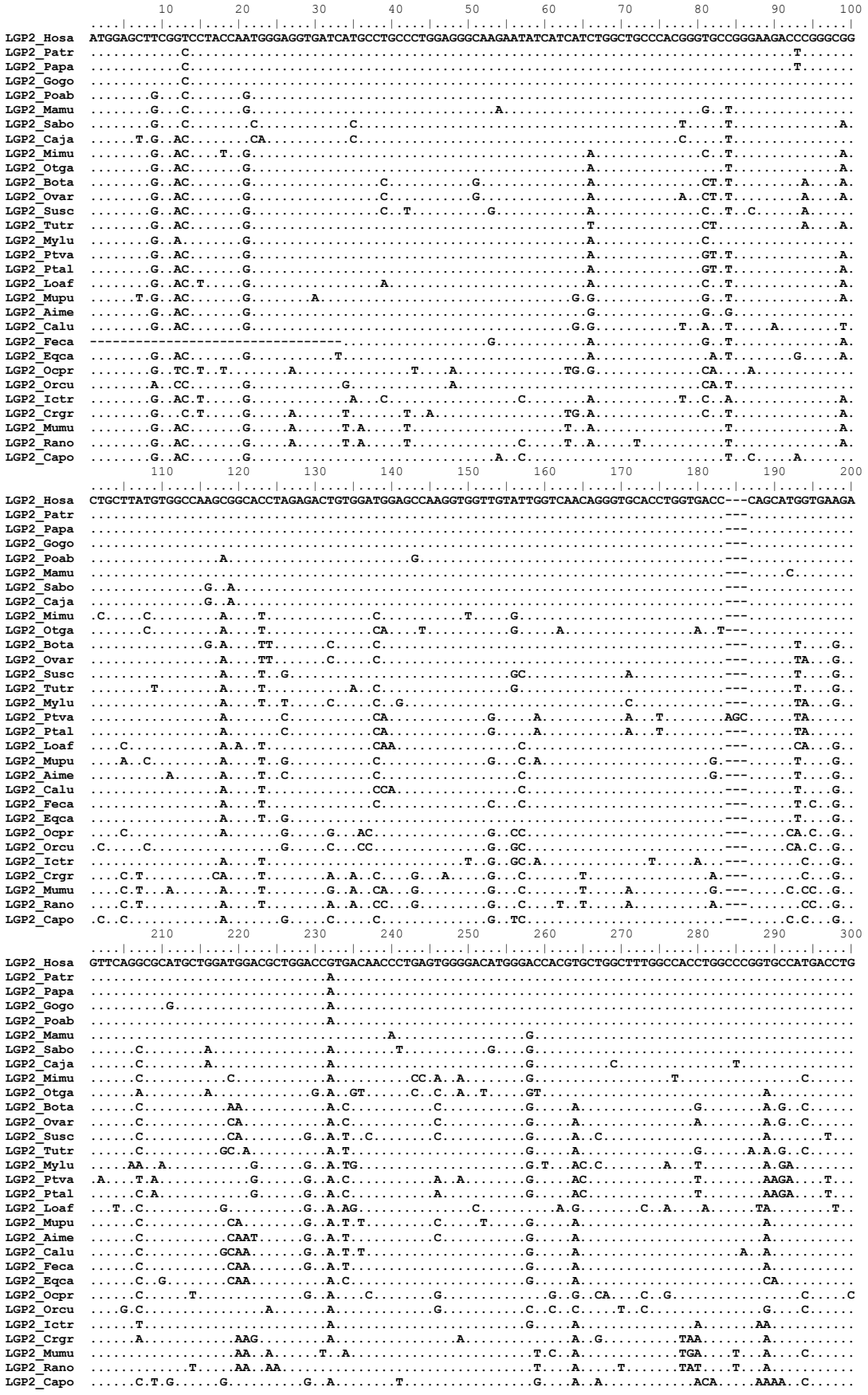

2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
MDA5_Hosa AGAAATATTGCCAAGCATTACAAGAATAACCCATCACTAATAACTTTCCTTTGCAAAAACTGCAGTGTGCTAGCCTGTTCTGGGGAAGATATCCATGTAA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

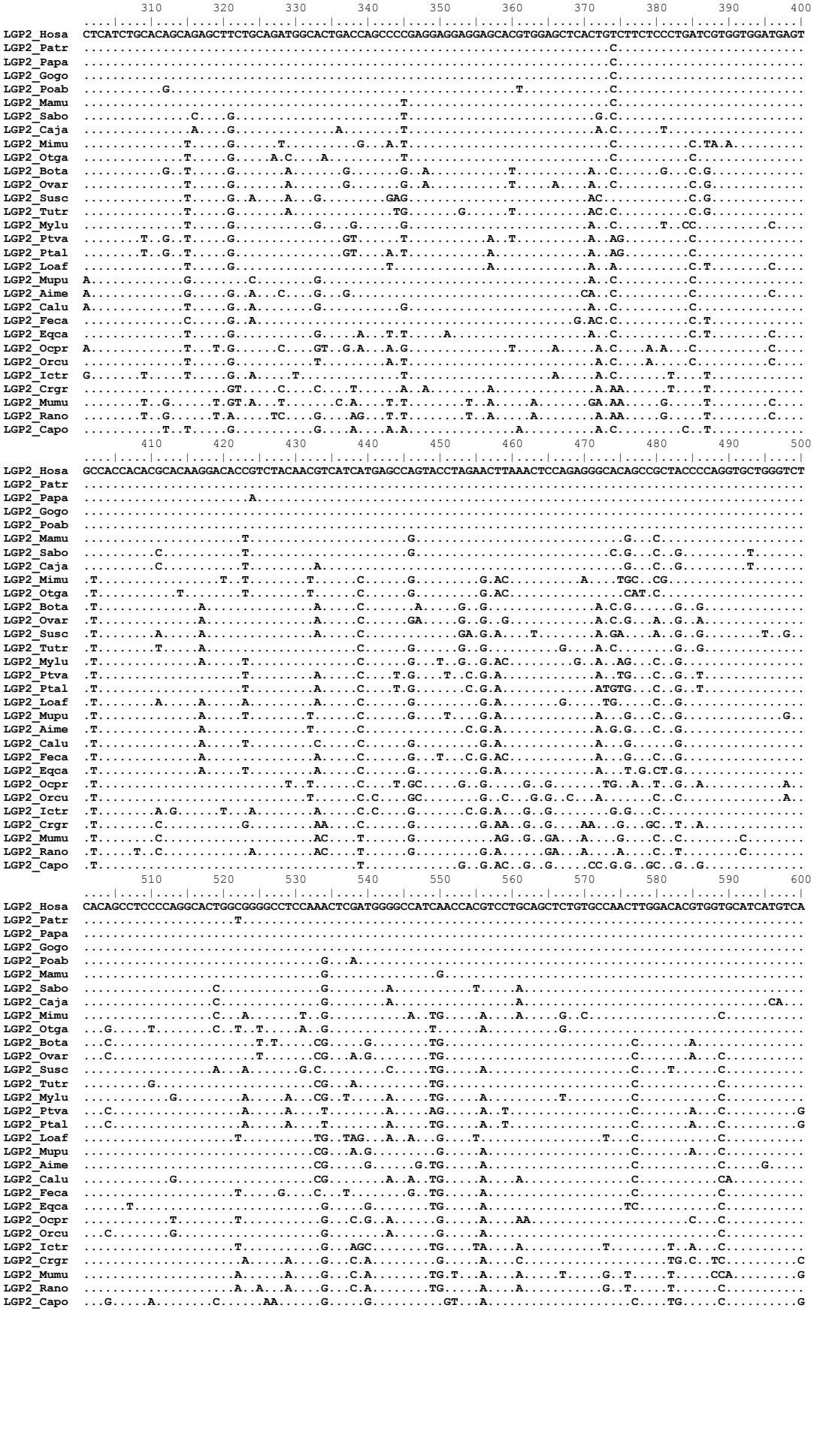

2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
MDA5_Hosa TTGAGAAAATGCATCACGTCAATATGACCCCAGAATTCAAGGAACTTTACATTGTAAGAGAAAACAAAGCACTGCAAAAGAAGTGTGCCGACTATCAAAT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

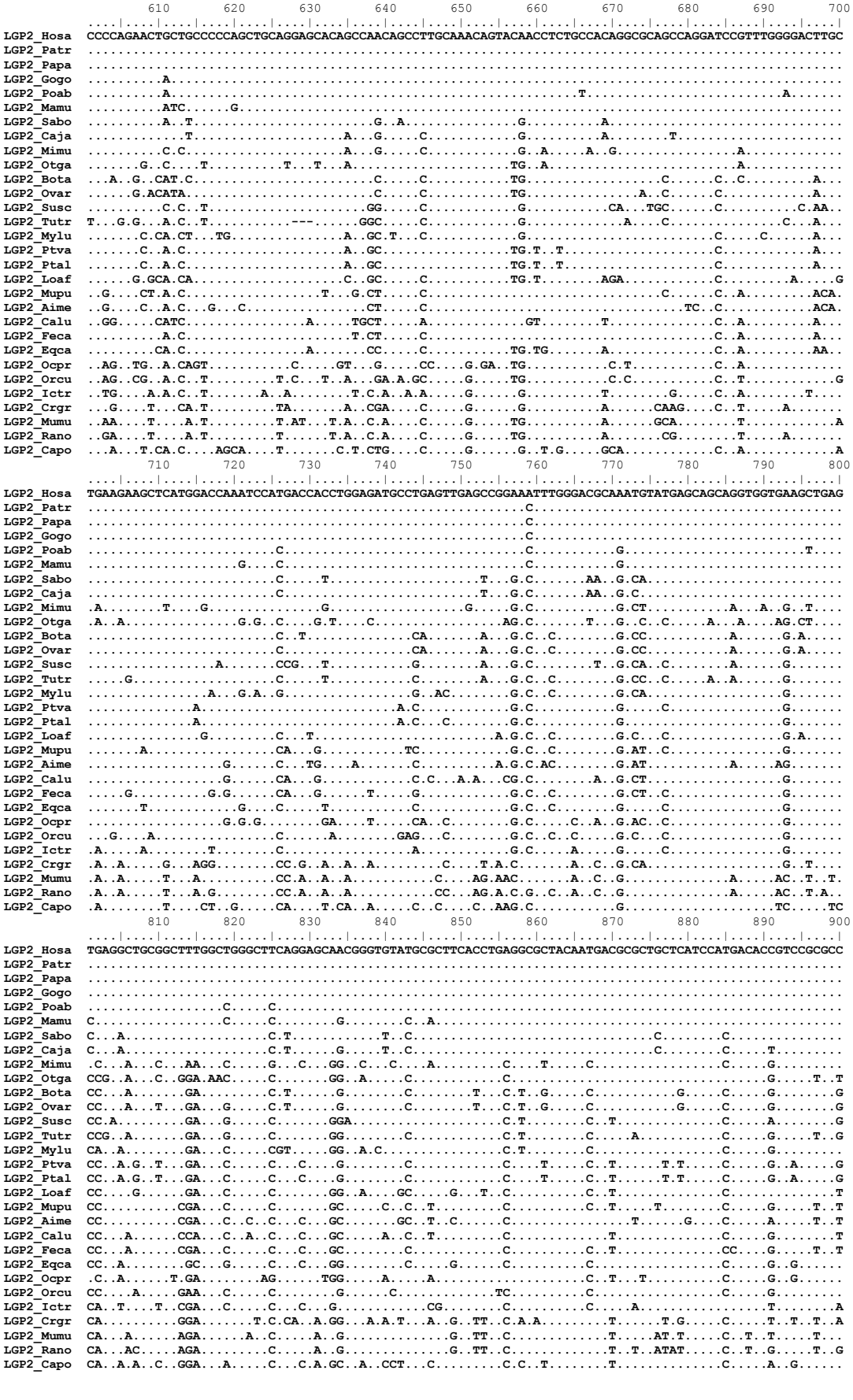

2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
MDA5_Hosa AAATGGTGAAATCATCTGCAAA---TGTGGCCAGGCTTGGGGAACAATGATGGTGCACAAAGGCTTAGATTTGCCTTGTCTCAAAATAAGGAATTTTGTA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

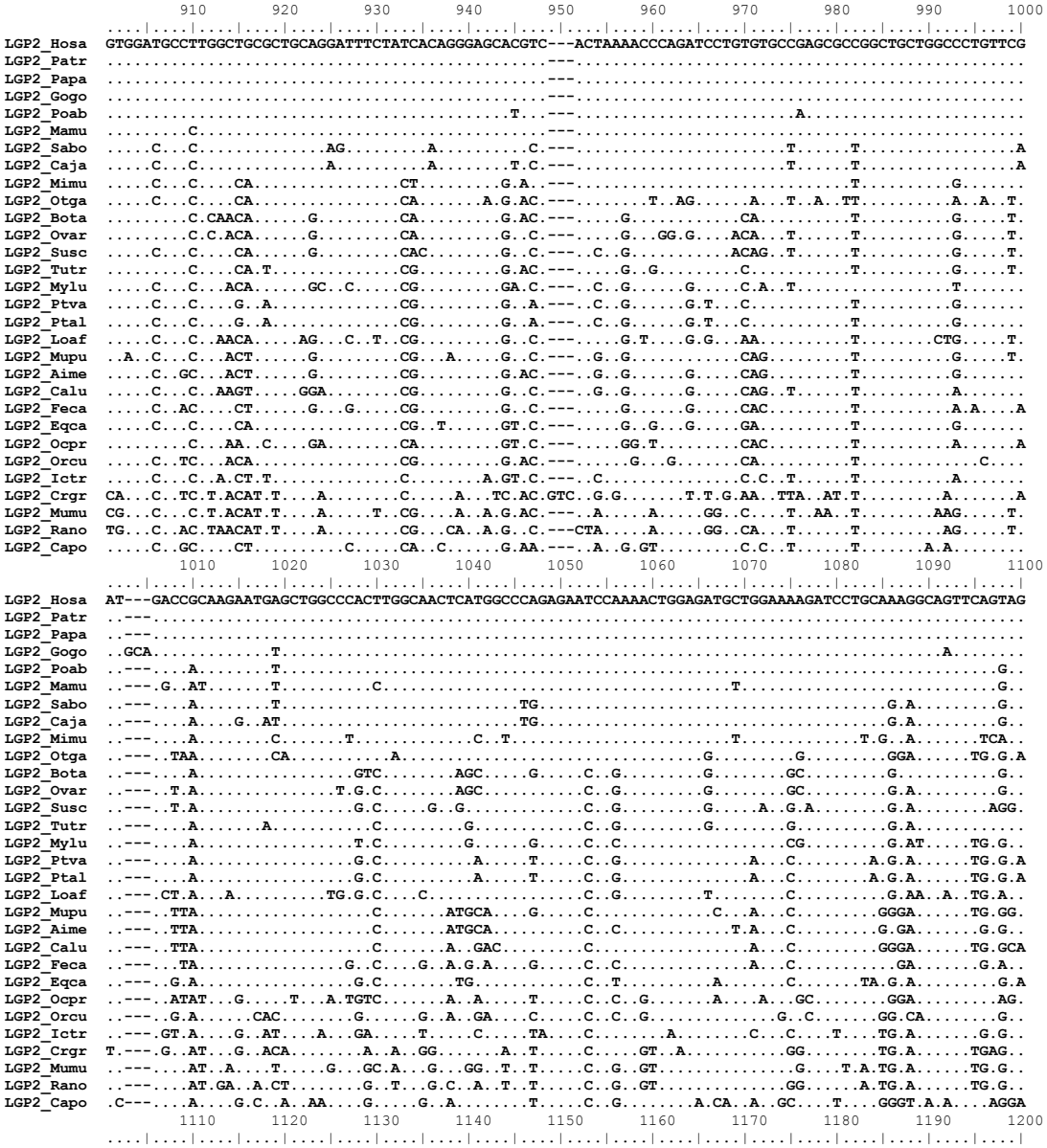

3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
MDA5_Hosa GTGGTTTTCAAAAATAATTCAACAAAGAAACAATACAAAAAGTGGGTAGAATTACCTATCACATTTCCCAATCTTGACTATTCAGAATGCTGTTTATTTA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo
3110
MDA5_Hosa GTGATGAGGATTAG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

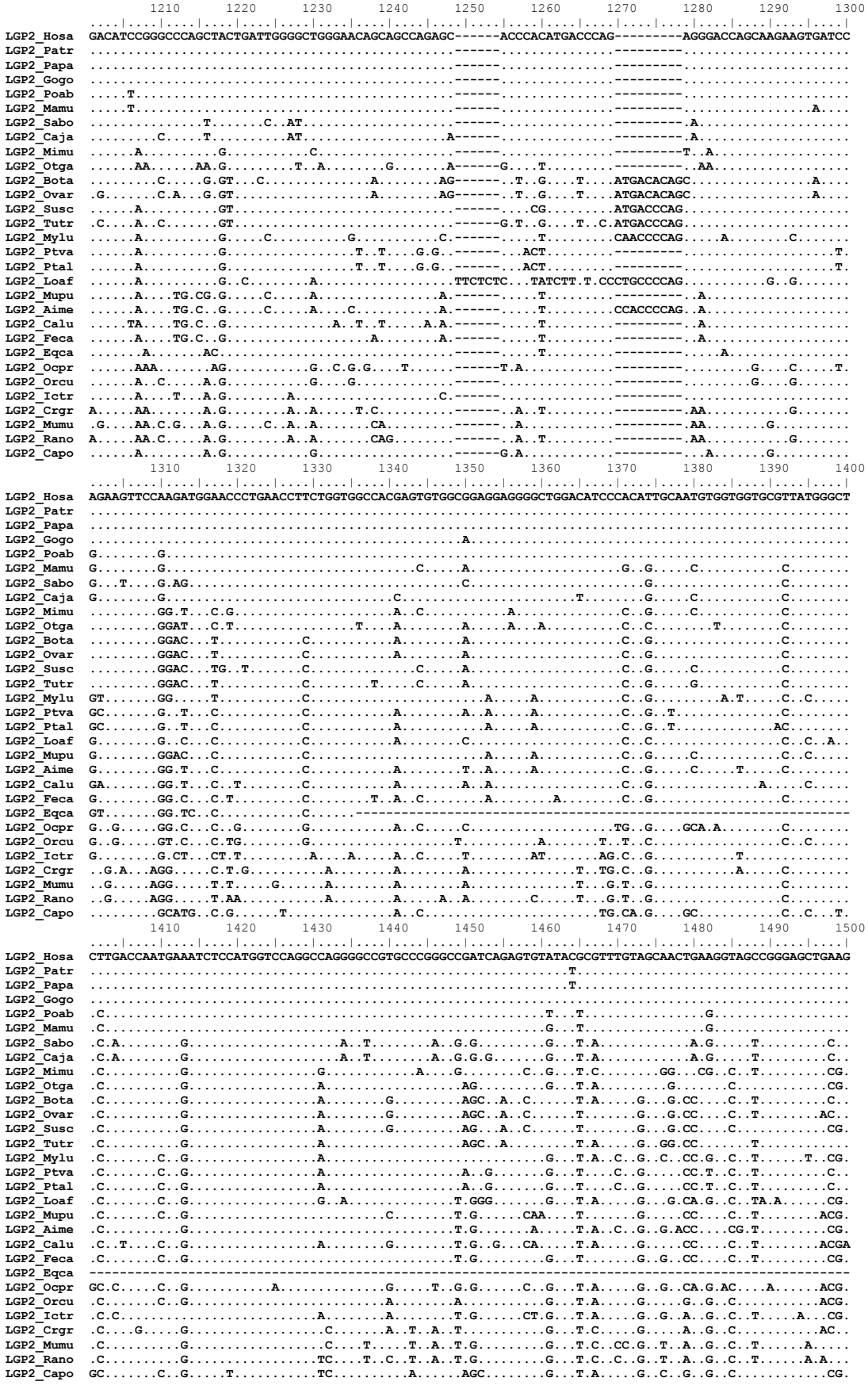
Supplementary Figure S3

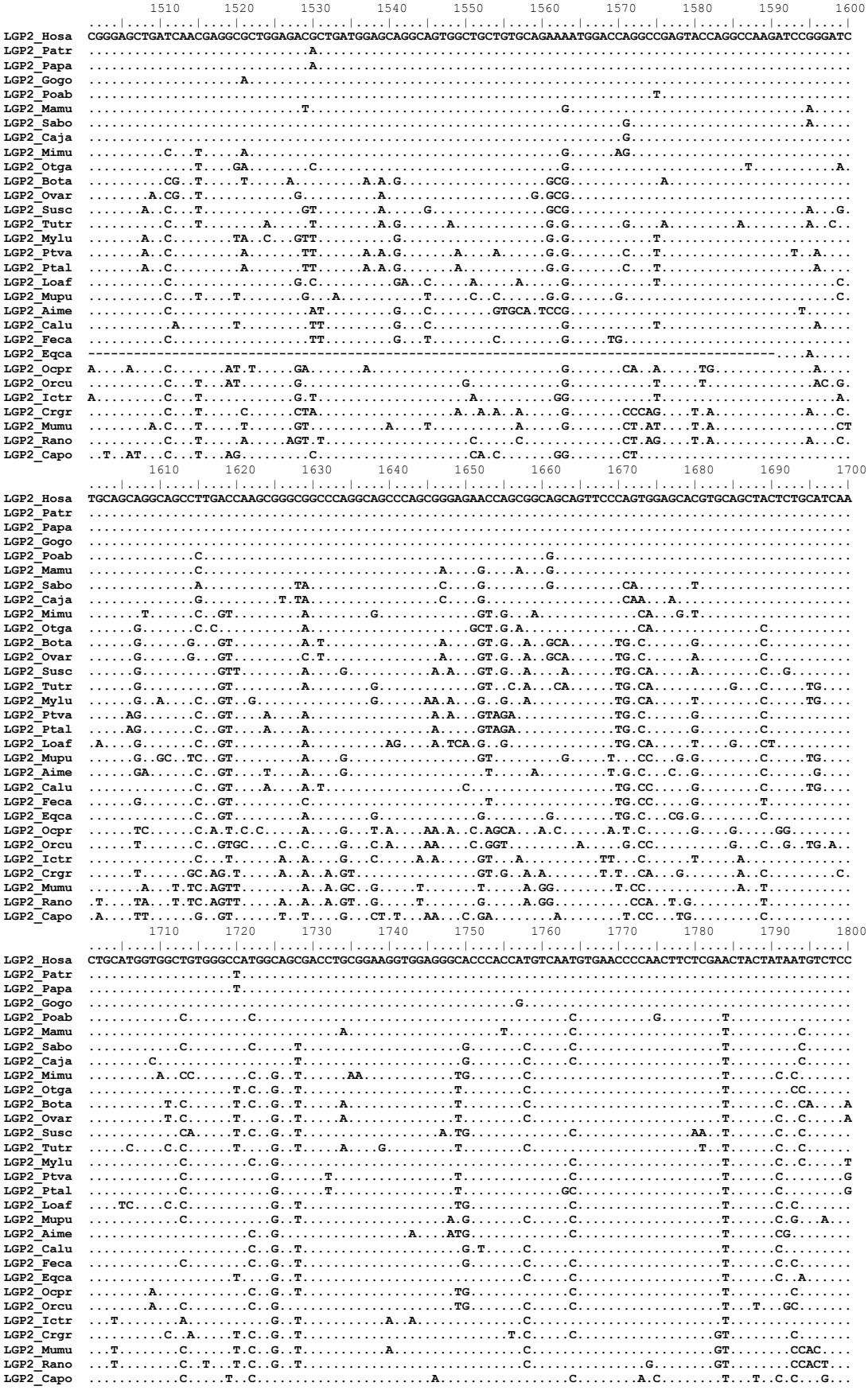


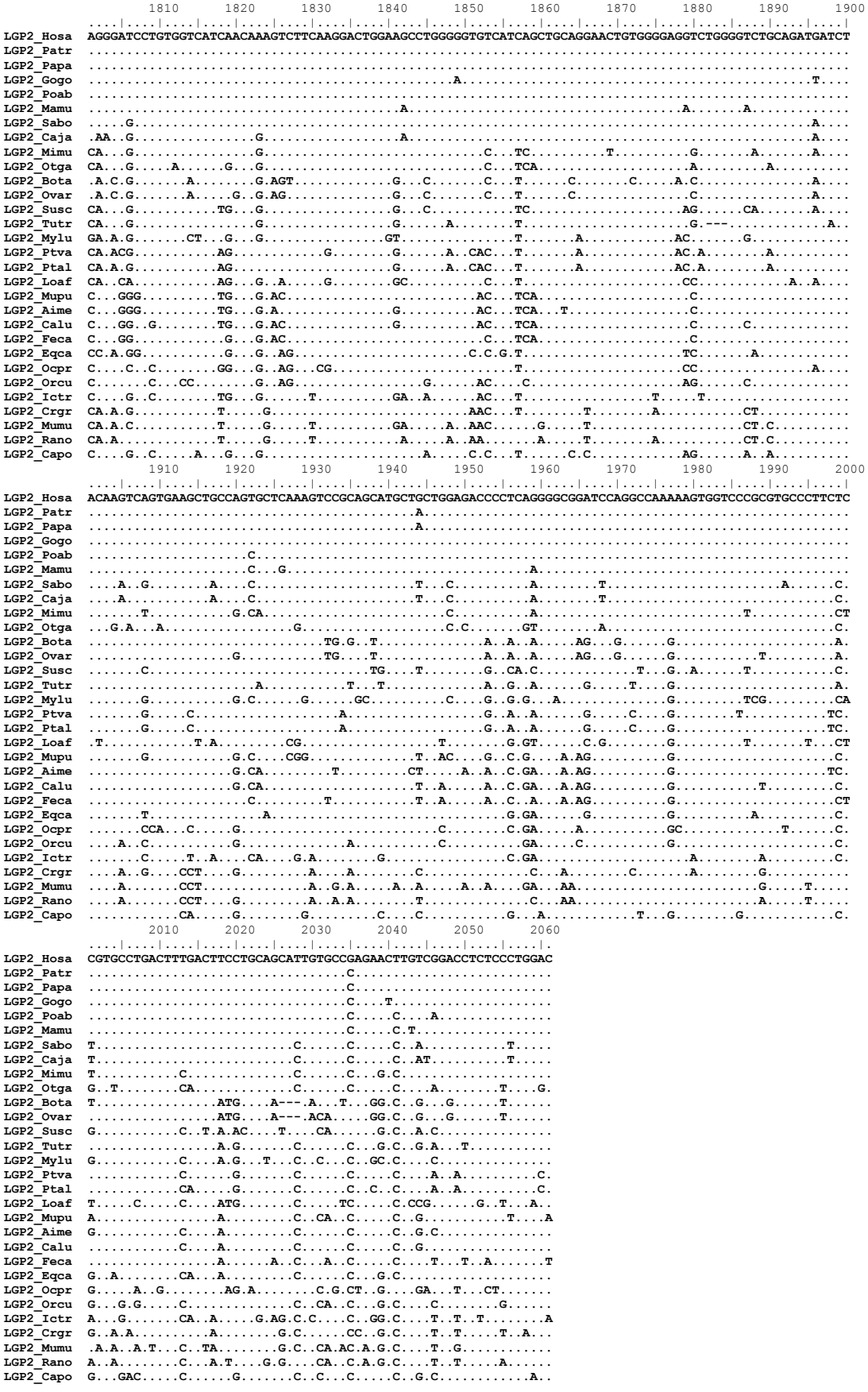

310 320 330 340 350 360 370 380 390 400
LGP2_Hosa CTCATCTGCACAGCAGAGCTTCTGCAGATGGCACTGACCAGCCCCGAGGAGGAGGAGCACGTGGAGCTCACTGTCTTCTCCCTGATCGTGGTGGATGAGT
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
410 420 430 440 450 460 470 480 490 500
LGP2_Hosa GCCACCACACGCACAAGGACACCGTCTACAACGTCATCATGAGCCAGTACCTAGAACTTAAACTCCAGAGGGCACAGCCGCTACCCCAGGTGCTGGGTCT
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
510 520 530 540 550 560 570 580 590 600
LGP2_Hosa CACAGCCTCCCCAGGCACTGGCGGGGCCTCCAAACTCGATGGGGCCATCAACCACGTCCTGCAGCTCTGTGCCAACTTGGACACGTGGTGCATCATGTCA
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo


610 620 630 640 650 660 670 680 690 700
LGP2_Hosa CCCCAGAACTGCTGCCCCCAGCTGCAGGAGCACAGCCAACAGCCTTGCAAACAGTACAACCTCTGCCACAGGCGCAGCCAGGATCCGTTTGGGGACTTGC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
710 720 730 740 750 760 770 780 790 800
LGP2_Hosa TGAAGAAGCTCATGGACCAAATCCATGACCACCTGGAGATGCCTGAGTTGAGCCGGAAATTTGGGACGCAAATGTATGAGCAGCAGGTGGTGAAGCTGAG
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
810 820 830 840 850 860 870 880 890 900
LGP2_Hosa TGAGGCTGCGGCTTTGGCTGGGCTTCAGGAGCAACGGGTGTATGCGCTTCACCTGAGGCGCTACAATGACGCGCTGCTCATCCATGACACCGTCCGCGCC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo

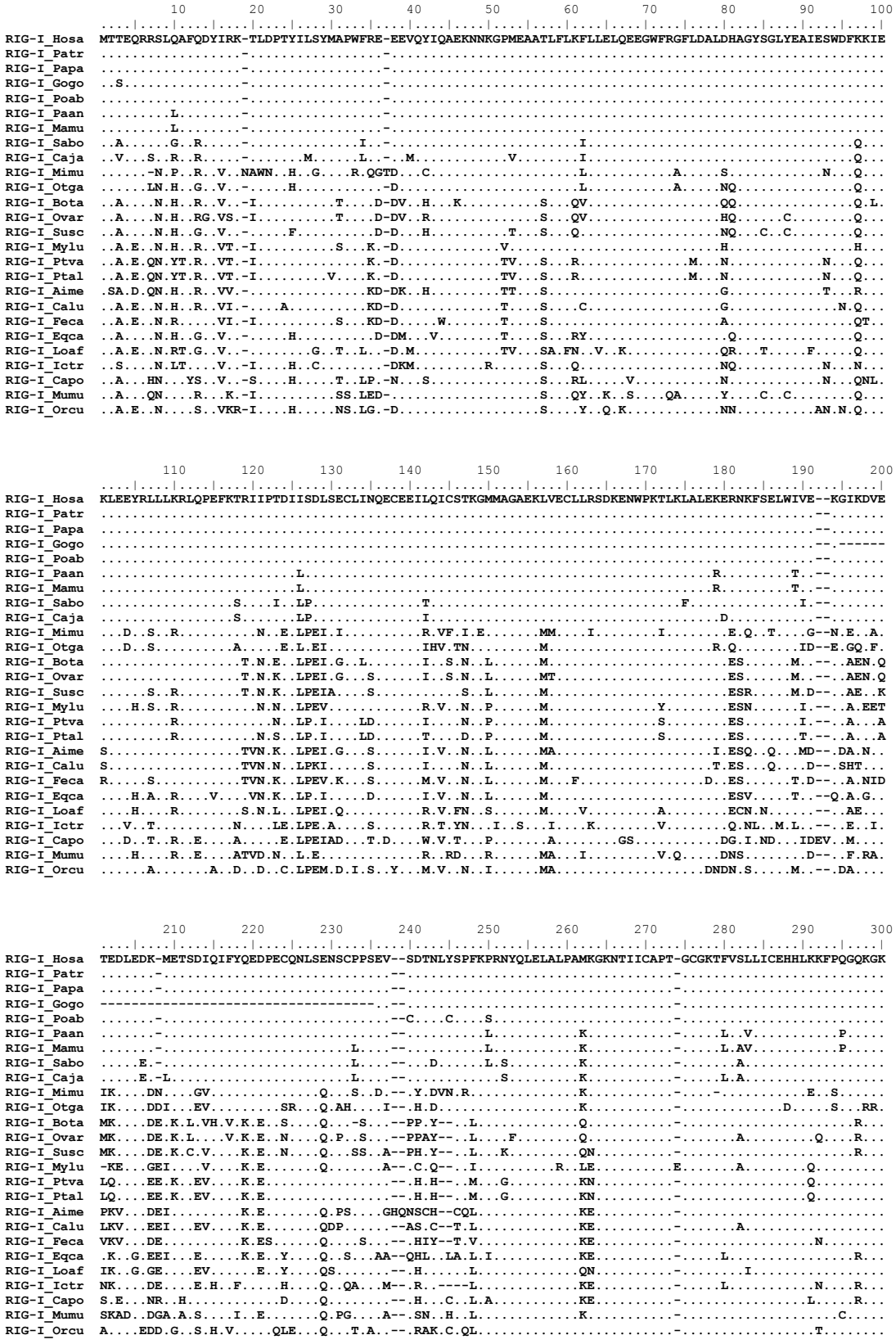

910 920 930 940 950 960 970 980 990 1000
LGP2_Hosa GTGGATGCCTTGGCTGCGCTGCAGGATTTCTATCACAGGGAGCACGTC---ACTAAAACCCAGATCCTGTGTGCCGAGCGCCGGCTGCTGGCCCTGTTCG
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
LGP2_Hosa AT---GACCGCAAGAATGAGCTGGCCCACTTGGCAACTCATGGCCCAGAGAATCCAAAACTGGAGATGCTGGAAAAGATCCTGCAAAGGCAGTTCAGTAG
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
LGP2_Hosa CTCTAACAGC---CCTCGGGGTATCATCTTCACCCGCACCCGCCAAAGCGCACACTCCCTCCTGCTCTGGCTCCAGCAGCAGCAGGGCCTGCAGACTGTG
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo

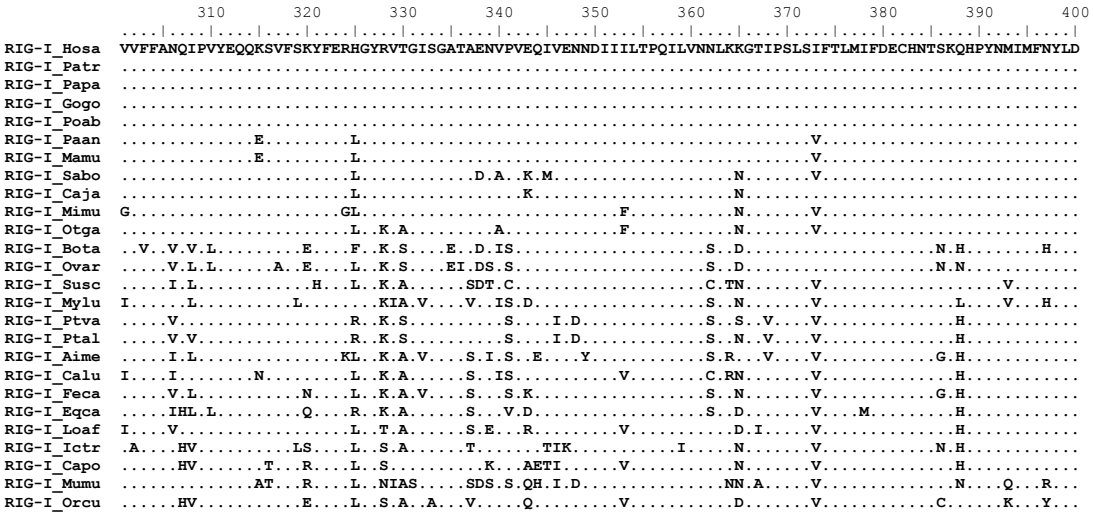

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
LGP2_Hosa GACATCCGGGCCCAGCTACTGATTGGGGCTGGGAACAGCAGCCAGAGC------ACCCACATGACCCAG---------AGGGACCAGCAAGAAGTGATCC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
LGP2_Hosa AGAAGTTCCAAGATGGAACCCTGAACCTTCTGGTGGCCACGAGTGTGGCGGAGGAGGGGCTGGACATCCCACATTGCAATGTGGTGGTGCGTTATGGGCT
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
LGP2_Hosa CTTGACCAATGAAATCTCCATGGTCCAGGCCAGGGGCCGTGCCCGGGCCGATCAGAGTGTATACGCGTTTGTAGCAACTGAAGGTAGCCGGGAGCTGAAG
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo

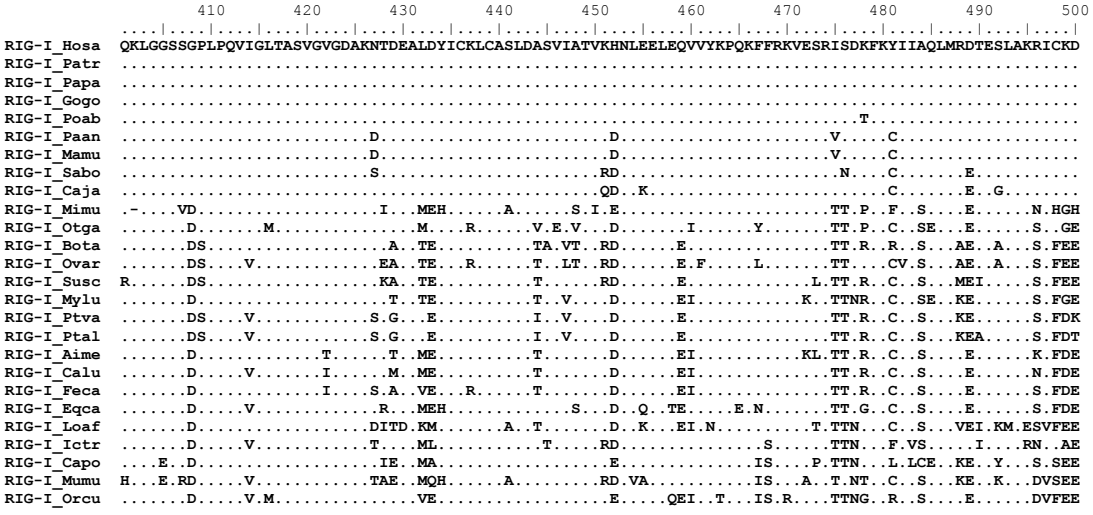

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
LGP2_Hosa CGGGAGCTGATCAACGAGGCGCTGGAGACGCTGATGGAGCAGGCAGTGGCTGCTGTGCAGAAAATGGACCAGGCCGAGTACCAGGCCAAGATCCGGGATC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
LGP2_Hosa TGCAGCAGGCAGCCTTGACCAAGCGGGCGGCCCAGGCAGCCCAGCGGGAGAACCAGCGGCAGCAGTTCCCAGTGGAGCACGTGCAGCTACTCTGCATCAA
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
LGP2_Hosa CTGCATGGTGGCTGTGGGCCATGGCAGCGACCTGCGGAAGGTGGAGGGCACCCACCATGTCAATGTGAACCCCAACTTCTCGAACTACTATAATGTCTCC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo

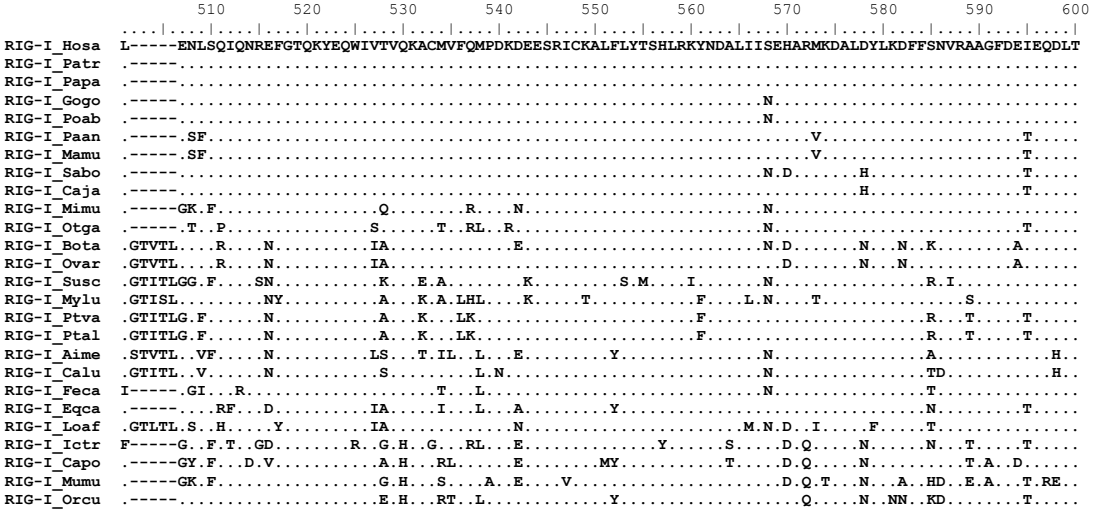

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
LGP2_Hosa AGGGATCCTGTGGTCATCAACAAAGTCTTCAAGGACTGGAAGCCTGGGGGTGTCATCAGCTGCAGGAACTGTGGGGAGGTCTGGGGTCTGCAGATGATCT
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
LGP2_Hosa ACAAGTCAGTGAAGCTGCCAGTGCTCAAAGTCCGCAGCATGCTGCTGGAGACCCCTCAGGGGCGGATCCAGGCCAAAAAGTGGTCCCGCGTGCCCTTCTC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
2010 2020 2030 2040 2050 2060
LGP2_Hosa CGTGCCTGACTTTGACTTCCTGCAGCATTGTGCCGAGAACTTGTCGGACCTCTCCCTGGAC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo

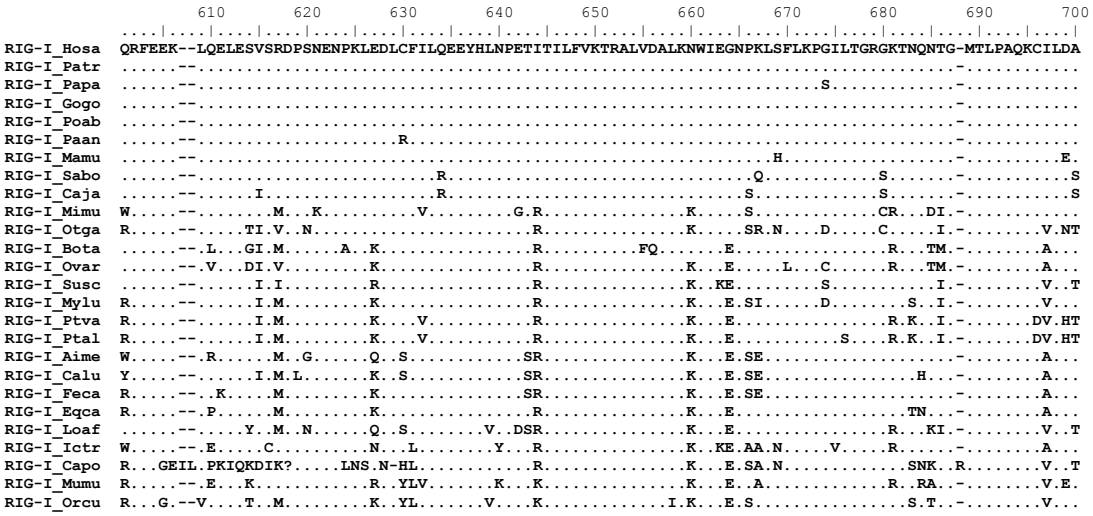
Supplementary Figure S4

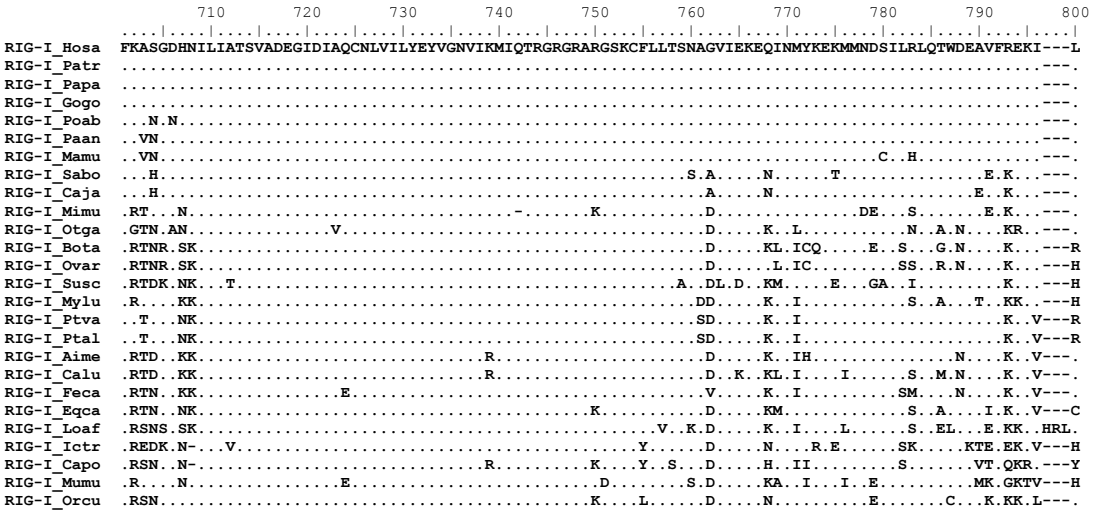


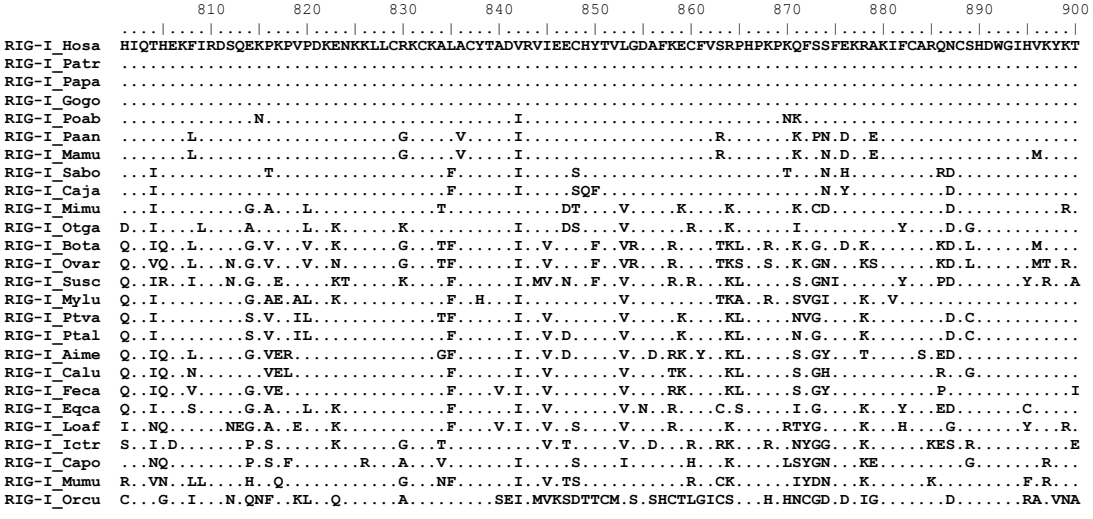

310 320 330 340 350 360 370 380 390 400
RIG-I_Hosa VVFFANQIPVYEQQKSVFSKYFERHGYRVTGISGATAENVPVEQIVENNDIIILTPQILVNNLKKGTIPSLSIFTLMIFDECHNTSKQHPYNMIMFNYLD
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu


410 420 430 440 450 460 470 480 490 500
RIG-I_Hosa QKLGGGSSGPLPQVIGLTASVGVGDAKNTDEALDYICKLCASLDASVIATVKHNLEELEQVVYKPQKFFRKVESRISDKFKYIIAQLMRDTESLAKRICKD
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu


510 520 530 540 550 560 570 580 590 600
RIG-I_Hosa L-----ENLSQIQNREFGTQKYEQWIVTVQKACMVFQMPDKDEESRICKALFLYTSHLRKYNDALIISEHARMKDALDYLKDFFSNVRAAGFDEIEQDLT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

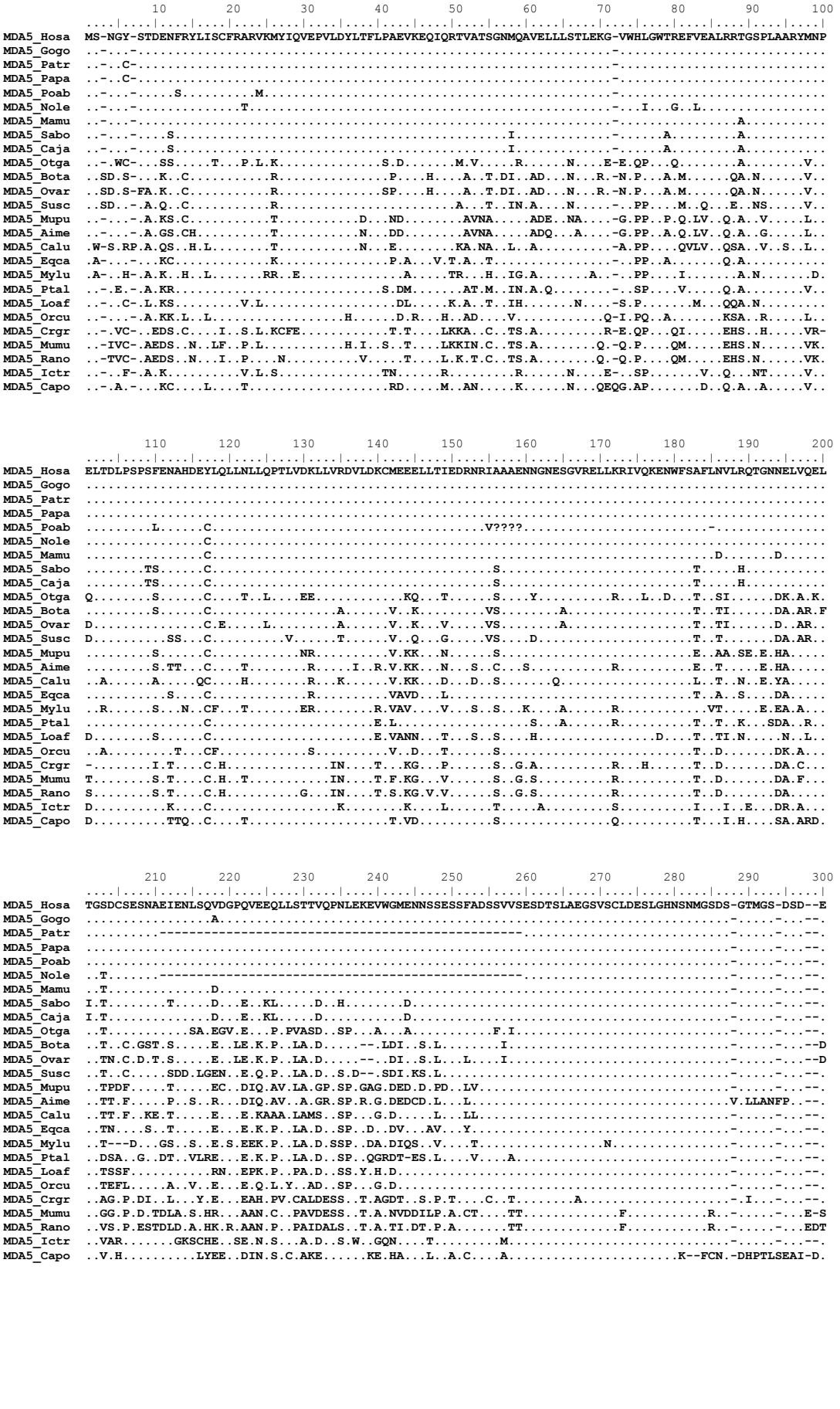

610 620 630 640 650 660 670 680 690 700
RIG-I_Hosa QRFEEK--LQELESVSRDPSNENPKLEDLCFILQEEYHLNPETITILFVKTRALVDALKNWIEGNPKLSFLKPGILTGRGKTNQNTG-MTLPAQKCILDA
RIG-I_Patr ......--.....-.....
RIG-I_Papa ......--.....S.....-.....
RIG-I_Gogo ......--.....-.....
RIG-I_Poab ......--.....-.....
RIG-I_Paan ......--.....R.....-.....
RIG-I_Mamu ......--.....H.....-.....E.
RIG-I_Sabo ......--.....R.....Q.....S.....-.....S
RIG-I_Caja ......--.....I.....R.....S.....S.....-.....S
RIG-I_Mimu W.....--.....M...K.....V.....G.R.....K.....S.....CR...DI.-.....
RIG-I_Otga R.....--.....TI.V..N.....R.....K.....SR.N....D.....C.....I.-.....V.NT
RIG-I_Bota ......--.L...GI.M.....A..K.....R.....FQ.....E.....R...TM.-.....A...
RIG-I_Ovar ......--.V...DI.V.....K.....R.....K...E.....L...C.....R...TM.-.....A...
RIG-I_Susc ......--.....I.I.....R.....R.....K..KE.....S.....I.-.....V..T
RIG-I_Mylu R.....--.....I.M.....K.....R.....K...E.SI.....D.....S..I.-.....V...
RIG-I_Ptva R.....--.....I.M.....K....V.....R.....K...E.....R.K..I.-.....DV.HT
RIG-I_Ptal R.....--.....I.M.....K....V.....R.....K...E.....S...R.K..I.-.....DV.HT
RIG-I_Aime W.....--.R.....M..G.....Q..S.....SR.....K...E.SE.....-.....A...
RIG-I_Calu Y.....--.....I.M.L.....K..S.....SR.....K...E.SE.....H...-.....A...
RIG-I_Feca R.....--..K.....M.....K.....SR.....K...E.SE.....-.....A...
RIG-I_Eqca R.....--.P.....M.....K.....R.....K...E.....TN...-.....A...
RIG-I_Loaf .....--.....Y..M..N.....Q..S.....V..DSR.....K...E.....R...KI.-.....V..T
RIG-I_Ictr W.....--.E.....C.....N...L.....Y...R.....K..KE.AA.N.....V.....R.....-.....A...
RIG-I_Capo R...GEIL.PKIQKDIK?.....LNS.N-HL.....R.....K...E.SA.N.....SNK..R.....V..T
RIG-I_Mumu R.....--.E...K.....R..YLV.....K...K.....K...E..A.....R..RA..-.....V.E.
RIG-I_Orcu R...G.--V....T..M.....K..YL.....V....K.....I.K...E.S.....S.T..-.....V...

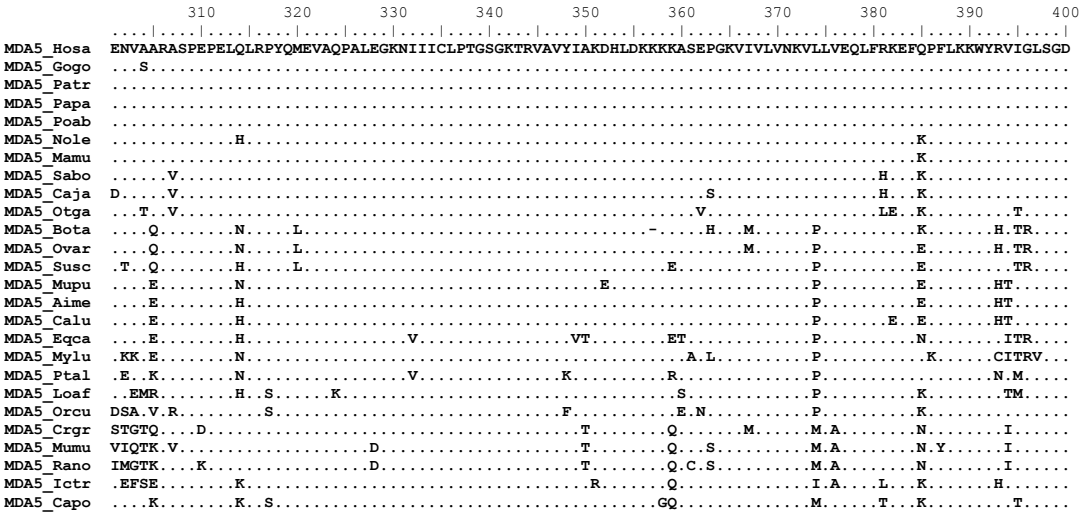

710 720 730 740 750 760 770 780 790 800
RIG-I_Hosa FKASGDHNILIATSVADEGIDIAQCNLVILYEYVGNVIKMIQTRGRGRARGSKCFLLTSNAGVIEKEQINMYKEKMMNDSILRLQTWDEAVFREKI---L
RIG-I_Patr .....---.
RIG-I_Papa .....---.
RIG-I_Gogo .....---.
RIG-I_Poab ...N.N.....---.
RIG-I_Paan ..VN.....---.
RIG-I_Mamu ..VN.....C..H.....---.
RIG-I_Sabo ...H.....S.A.....N.....T.....E.K...---.
RIG-I_Caja ...H.....A.....N.....E..K...---.
RIG-I_Mimu .RT...N.....-.....K.....D.....DE...S.....E.K...---.
RIG-I_Otga .GTN.AN.....V.....D.....K..L.....N..A.N....KR..---.
RIG-I_Bota .RTNR.SK.....D.....KL.ICQ.....E..S...G.N....K...---R
RIG-I_Ovar .RTNR.SK.....D.....L.IC.....SS..R.N....K...---H
RIG-I_Susc .RTDK.NK...T.....A..DL.D..KM.....E...GA..I.....K...---H
RIG-I_Mylu .R....KK.....DD.....K..I.....S..A...T..KK..---H
RIG-I_Ptva ..T...NK.....SD.....K..I.....K..V---R
RIG-I_Ptal ..T...NK.....SD.....K..I.....K..V---R
RIG-I_Aime .RTD..KK.....R.....D.....K..IH.....N....K..V---.
RIG-I_Calu .RTD..KK.....R.....D..K..KL.I....I.....S..M.N....K..V---.
RIG-I_Feca .RTN..KK.....E.....V.....K..I.....SM...N....K..V---.
RIG-I_Eqca .RTN..NK.....K.....D.....KM.....S..A....I.K..V---C
RIG-I_Loaf .RSNS.SK.....V..K.D.....K..I....L.....S..EL...E.KK..HRL.
RIG-I_Ictr .REDK.N-...V.....Y.....D.....N....R.E.....SK.....KTE.EK.V---H
RIG-I_Capo .RSN..N-.....R.....K....Y..S...D.....H..II.....S.....VT.QKR.---Y
RIG-I_Mumu .R....N.....E.....D.....S.D.....KA..I...I..E.....MK.GKTV---H
RIG-I_Orcu .RSN.....K....L.....D.....N.....E.....C...K.KK.L---.

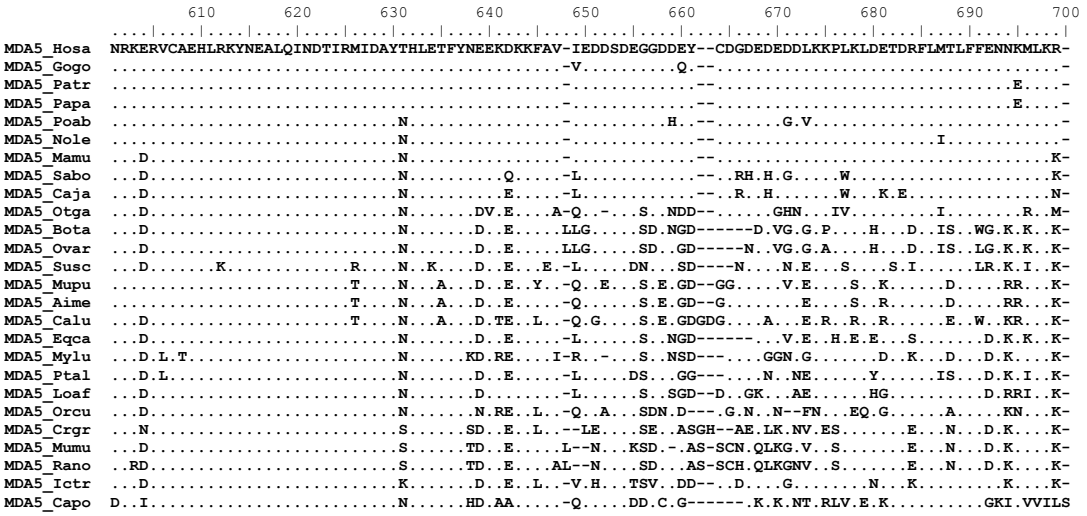

810 820 830 840 850 860 870 880 890 900
RIG-I_Hosa HIQTHEKFIRDSQEKPKPVPDKENKKLLCRKCKALACYTADVRVIEECHYTVLGDAFKECFVSRPHPKPKQFSSFEKRAKIFCARQNCSHDWGIHVKYKT
RIG-I_Patr .....
RIG-I_Papa .....
RIG-I_Gogo .....
RIG-I_Poab .....N.....I.....NK.....
RIG-I_Paan .....L.....G.....V.....I.....R.....K.PN.D..E.....
RIG-I_Mamu .....L.....G.....V.....I.....R.....K..N.D..E.....M....
RIG-I_Sabo ...I.....T.....F.....I.....S.....T...N.H.....RD.....
RIG-I_Caja ...I.....F.....I.....SQF.....N.Y.....D.....
RIG-I_Mimu ...I.....G.A...L.....T.....DT....V.....K....K.....K.CD.....D.....R.
RIG-I_Otga D..I....L....A....L..K.....K.....I....DS....V.....R...K.....I.....Y....D.G.....
RIG-I_Bota Q..IQ..L.....G.V...V..K.....G...TF.....I..V....F..VR...R....TKL..R..K.G..D.K.....KD.L.....M....
RIG-I_Ovar Q..VQ..L...N.G.V...V..N.....G...TF.....I..V....F..VR...R....TKS..S..K.GN...KS.....KD.L.....MT.R.
RIG-I_Susc Q..IR..I...N.G..E.....KT.....K....F.....I.MV.N..F..V....R.R...KL.....S.GNI.....Y...PD.....Y.R..A
RIG-I_Mylu Q..I.....G.AE.AL..K.....F..H...I.....V.....TKA..R..SVGI...K..V.....
RIG-I_Ptva Q..I.....S.V..IL.....TF.....I..V.....V.....K....KL.....NVG....K.....D.C.....
RIG-I_Ptal Q..I.....S.V..IL.....F.....I..V.D.....V.....K....KL.....N.G....K.....D.C.....
RIG-I_Aime Q..IQ..L.....G.VER.....GF.....I..V.D.....V..D.RK.Y..KL.....S.GY...T.....S.ED.....
RIG-I_Calu Q..IQ..N.....VEL.....F.....I..V.....V....TK....KL.....S.GH.....R..G.....
RIG-I_Feca Q..IQ..V.....G.VE.....F....V.I..V.....V....RK....KL.....S.GY.....P.....I
RIG-I_Eqca Q..I...S.....G.A...L..K.....F.....I..V.....V.N..R....C.S.....I.G....K...Y...ED.....C.....
RIG-I_Loaf I..NQ.....NEG.A..E...K.....F....V.I..V..S....V....R.....K.....RTYG....K...H....G.....Y...R.
RIG-I_Ictr S..I.D.....P.S.....K.....G...T.....V.T.....V..D...R..RK...R..NYGG...K.....KES.R.....E
RIG-I_Capo ...NQ.....P.S.F.....R...A...V.....I.....S....I.....H...K.....LSYGN...KE.....G.....R...
RIG-I_Mumu R..VN..LL....H..Q.....G...NF.....I..V.TS.....R..CK.....IYDN...K.....K.....F.R...
RIG-I_Orcu C...G..I...N.QNF..KL..Q.....A.....SEI.MVKSDTTCM.S.SHCTLGICS...H.HNCGD.D.IG.....D.....RA.VNA

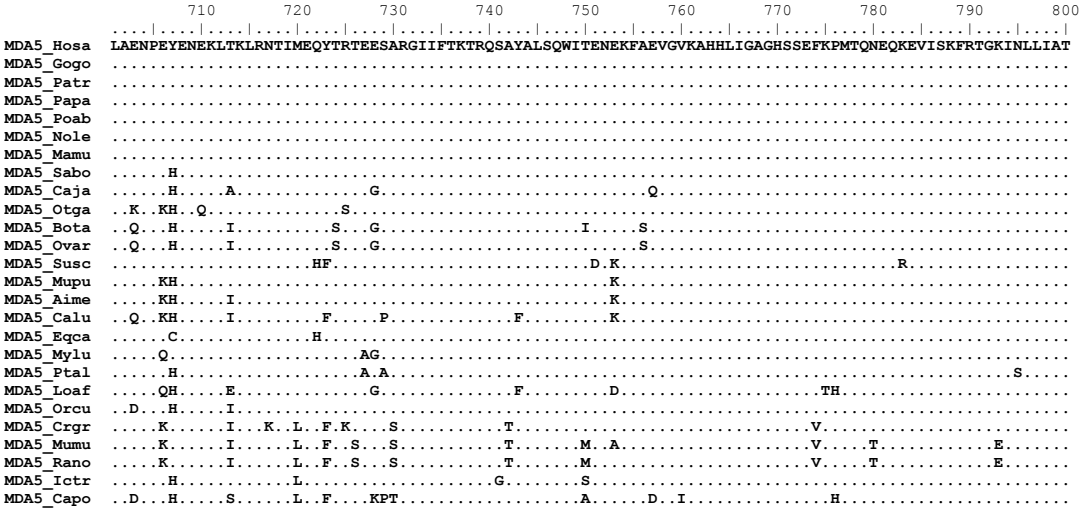
```
                910       920       930       940       950       960       970
           ....|....|....|....|....|....|....|....|....|....|....|....|....|..
RIG-I_Hosa FEIPVIKIESFVVEDIATGVQTLYSKWKDFHFEKIPFDPAEMSK*------------------------------
RIG-I_Patr ..........................................*------------------------------
RIG-I_Papa ..........................................*------------------------------
RIG-I_Gogo ..........................................*------------------------------
RIG-I_Poab ..........................................*------------------------------
RIG-I_Paan ........................................A.*------------------------------
RIG-I_Mamu ........................................A.*------------------------------
RIG-I_Sabo .......M..........................A....TISEPQPSASGNE*--------------------
RIG-I_Caja ..................................A.....*------------------------------
RIG-I_Mimu ..............V.....R.............A....R*------------------------------
RIG-I_Otga ................A.R......E.....T..HAV..Y.*------------------------------
RIG-I_Bota ..............V...A....A.....N.......A....PWAQDLNLQGVDGLE*------------------
RIG-I_Ovar .........G......V...A..........N.......A....AGAQDLNLQAMNGLE*------------------
RIG-I_Susc ..-.F......AL.EY......VHAS....N...LS..A...AGGAQDMGLQGWATLSEGEIGLGLNHGSPVPLLR*
RIG-I_Mylu ....I..............E.R..A..R......L.........*------------------------------
RIG-I_Ptva ....F..................A..R......L...A...P.*------------------------------
RIG-I_Ptal ....F..............A....A..R....G.L...A...P.*------------------------------
RIG-I_Aime ..................A.K..A..R..PLQ.....A.....*------------------------------
RIG-I_Calu ..................A.K..A.....P......GSP.IPE*------------------------------
RIG-I_Feca ..................A.K..A.............AK..P.*------------------------------
RIG-I_Eqca ................R.....A..R..N.......A....N*------------------------------
RIG-I_Loaf ..................A.S..A.............A.....*------------------------------
RIG-I_Ictr ....................R............Q..A..YPNDFRA*----------------------------
RIG-I_Capo ..............V..L.KQ.PR..A.P.K..E..AT....*------------------------------
RIG-I_Mumu ..............VS...NRH.........R.Q.......V*------------------------------
RIG-I_Orcu ....L...Q......TV..K.MCP......N......N.E.L.E*------------------------------
```

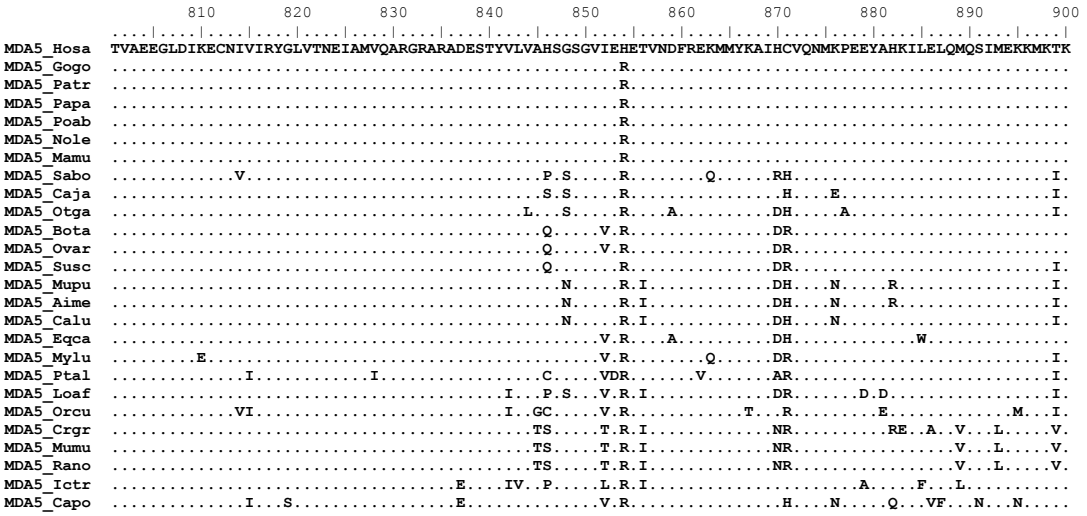
Supplementary Figure S5

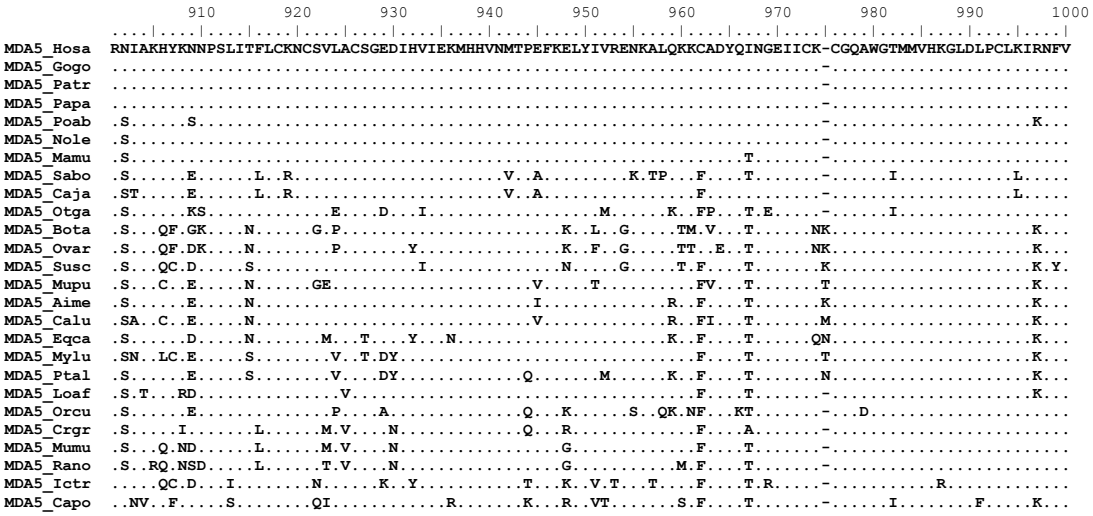


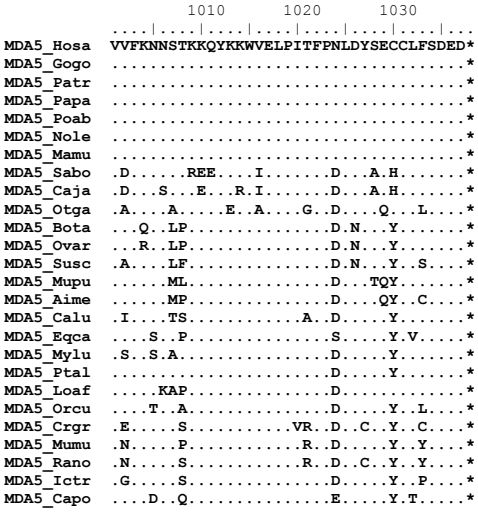

310 320 330 340 350 360 370 380 390 400
MDA5_Hosa ENVAARASPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIVLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGD
MDA5_Gogo ...S.....
MDA5_Patr .....
MDA5_Papa .....
MDA5_Poab .....
MDA5_Nole .....H.....K.....
MDA5_Mamu .....K.....
MDA5_Sabo .....V.....H...K.....
MDA5_Caja D.....V.....S.....H...K.....
MDA5_Otga ...T..V.....V.....LE..K.....T.....
MDA5_Bota ....Q.....N.....L.....-.....H...M.....P.....K.....H.TR....
MDA5_Ovar ....Q.....N.....L.....M.....P.....E.....H.TR....
MDA5_Susc .T..Q.....H.....L.....E.....P.....E.....TR....
MDA5_Mupu ....E.....N.....E.....P.....E.....HT.....
MDA5_Aime ....E.....H.....P.....E.....HT.....
MDA5_Calu ....E.....H.....P.....E..E.....HT.....
MDA5_Eqca ....E.....H.....V.....VT.....ET.....P.....N.....ITR....
MDA5_Mylu .KK.E.....N.....A.L.....P.....K.....CITRV...
MDA5_Ptal .E..K.....N.....V.....K.....R.....P.....N.M.....
MDA5_Loaf ..EMR.....H..S.....K.....S.....P.....K.....TM.....
MDA5_Orcu DSA.V.R.....S.....F.....E.N.....P.....K.....
MDA5_Crgr STGTQ....D.....T.....Q.....M.....M.A.....N.....I.....
MDA5_Mumu VIQTK.V.....D.....T.....Q...S.....M.A.....N.Y.....I.....
MDA5_Rano IMGTK....K.....D.....T.....Q.C.S.....M.A.....N.....I.....
MDA5_Ictr .EFSE.....K.....R.....Q.....I.A...L...K.....H.....
MDA5_Capo ....K.....K..S.....GQ.....M.....T...K.....T.....
410 420 430 440 450 460 470 480 490 500
MDA5_Hosa TQLKISFPEVVKSCDIIISTAQILENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQILGLTASPGV
MDA5_Gogo .....R.....
MDA5_Patr .....R.....
MDA5_Papa .....R.....
MDA5_Poab .....RH.....
MDA5_Nole .....R.....
MDA5_Mamu .....R.....M.....
MDA5_Sabo .....K.....E.....R..K.....N.....S.....
MDA5_Caja .....E.....R..K.....IK.....QS.....
MDA5_Otga .....Q.H.V.....T..E.....F.....R.VK.....R..V.....AV.....A.....
MDA5_Bota .....T.....H.V.....S.E...D.IE.....RF.K.....K.....V.....
MDA5_Ovar .....T.....H.V.....S.E...D.IE.....R..K.....K.....V.....
MDA5_Susc .....T.....V.....S.E...P.....R..K.....K....E.....
MDA5_Mupu .....TY.V.....S.S.K...D...F.....V.....R..K.....K....C.....
MDA5_Aime .....TY.V.....S.K...D...F.....V.....R..K.....K....C.....
MDA5_Calu .....I..TY.V.....S.K...D...F...T.....R..K.....K....Y.....
MDA5_Eqca .....H.V.....S.K.....V.....R..K.....
MDA5_Mylu ..M.....KF...Y.....S.K...D.....R..K.....K.R.....M.....
MDA5_Ptal .....K....H.V.....S.K...D.....V.....R..K...R..K.....
MDA5_Loaf .....V.....S.K.....V.....R..E.....Q.....
MDA5_Orcu .....H.V.....F.....L.....R..K.....A.....
MDA5_Crgr .....Y.V.....S...D.....R..K.....K...Q....VH.....
MDA5_Mumu .....Y.V.....S.D.D.....R..K...R..D...Q...A.....
MDA5_Rano .....Y.V.....S...D.....R..K.....HK...Q...T.H.....
MDA5_Ictr .....Y.V.....S.S.....S.....R..K.....K..E.....
MDA5_Capo .....H.V.....S.S.....D.....R..K.....KA.Q....S.V.....
510 520 530 540 550 560 570 580 590 600
MDA5_Hosa GGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFAIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKEG
MDA5_Gogo .....
MDA5_Patr .....
MDA5_Papa .....
MDA5_Poab .....M.....D.....
MDA5_Nole .....D.....
MDA5_Mamu .....D.....
MDA5_Sabo ...R.....S.....E.....K.....D.....P.....V.....
MDA5_Caja ...R.....S.....E.....N.....T.....P.....V.....
MDA5_Otga .....N.....DH.K.....I....D.....I....K.....V.....S....V.....E..R..
MDA5_Bota ...K.....I.....V.....Q.INL..E..K.....V...D.KK....D.....K...F..IN.....L.....
MDA5_Ovar ...K.....I.....V.....Q.INL..E..K.....V...D.KK....D.....K..IF..I.....L.....
MDA5_Susc ...K.....I.....K.IN...D..K.....V...D....L..DR.....N...F..I.....L.....RD.
MDA5_Mupu ...KR..E.KD....I.....C.....IH...D..K.....D.....D.....NK..SF.....P.....TT.....R..
MDA5_Aime ...K...E.K.....I.....C.....IV...D..K.....D.....D.....NK..SF.....TT.....R..
MDA5_Calu R..KR..E.QQ....I.....C..I.....IN...G.MK.....D.....D.....NK..SF.....T.....R..
MDA5_Eqca .....I.....SV.....I....D..K.....D.....D..V.....F...N..P.....K.....
MDA5_Mylu .....I.....IC...E..K.....V...D.....DR.....K...F...N..P.....T.....R..
MDA5_Ptal ...N.....I.....I....D..K.....D.....D.....K...F...N..P.....T.....R..
MDA5_Loaf ...R.....D.....I.....S.....VGE.REH.K.....D.....K...F...N.....T.....R..
MDA5_Orcu ...N.....I.....D.MK.....D.....H.....
MDA5_Crgr ...K...E..K...NI.....G...H..K.....V...D...N.....S.....L.....S.
MDA5_Mumu .A.K..SE..K...NI.....G...H..K.....V...D...N.....AS.....K.....H.....D.
MDA5_Rano .A.K..SE..K...NI.....S...H..K.....V...D...N.....AS.....K..L.....H.....D.
MDA5_Ictr .....I.....E...D..K.....V...D.....VK...S..NH.....T.....K.....R..
MDA5_Capo .....I.....S.....DYYN...D..K.....V...D.....D.....SS..A...V.S.....N....T...SSL.SR..


610 620 630 640 650 660 670 680 690 700
MDA5_Hosa NRKERVCAEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAV-IEDDSDEGGDDEY--CDGDEDEDDLKKPLKLDETDRFLMTLFFENNKMLKR-
MDA5_Gogo .....-V.....Q.--.....-
MDA5_Patr .....-.....--.....E....-
MDA5_Papa .....-.....--.....E....-
MDA5_Poab .....N.....-.....H..--.....G.V.....-
MDA5_Nole .....N.....-.....--.....I.....-
MDA5_Mamu ...D.....N.....-.....--.....K-
MDA5_Sabo ...D.....N.....Q.....-L.....---.RH.H.G.....W.....K-
MDA5_Caja ...D.....N.....E.....-L.....---.R..H.....W...K.E.....N-
MDA5_Otga ...D.....N.....DV.E....A-Q..-...S..NDD--.....GHN...IV.....I.....R..M-
MDA5_Bota ...D.....N.....D..E.....LLG.....SD.NGD------D.VG.G.P....H...D..IS..WG.K.K..K-
MDA5_Ovar ...D.....N.....D..E.....LLG.....SD..GD-----N..VG.G.A....H...D..IS..LG.K.K..K-
MDA5_Susc ...D.....K.....R....N..K....D..E...E.-L.....DN...SD----N....N.E...S....S.I.....LR.K.I..K-
MDA5_Mupu .....T....N...A...D..E..Y..-Q..E...S.E.GD--GG.....V.E....S..K.....D.....RR...K-
MDA5_Aime .....T....N...A...D..E.....-Q.....S.E.GD--G.....E....S..R.....D.....RR...K-
MDA5_Calu ...D.....T....N...A...D.TE..L..-Q.G....S.E.GDGDG....A...E.R..R..R.....E..W..KR...K-
MDA5_Eqca ...D.....N.....D..E.....-L.....S..NGD------...V.E..H.E.E...S.....D.K.K..K-
MDA5_Mylu ...D.L.T.....N.....KD.RE....I-R..-...S..NSD---....GGN.G.....D..K...D...D.K....K-
MDA5_Ptal ...D.L.....N.....D..E.....-L.....DS...GG---....N..NE.....Y.....IS...D.K.I..K-
MDA5_Loaf ...D.....N.....D.....-L.....S..SGD--D..GK...AE.....HG.....D.RRI..K-
MDA5_Orcu ...D.....N.....N.RE..L..-Q..A...SDN.D---.G.N..N--FN...EQ.G.....A.....KN...K-
MDA5_Crgr ...N.....S.....SD..E..L..--LE....SE..ASGH--AE.LK.NV.ES.....E...N...D.K....K-
MDA5_Mumu ...D.....S.....TD..E.....L--N...KSD.-.AS-SCN.QLKG.V..S.....E...N...D.K....K-
MDA5_Rano ..RD.....S.....TD..E....AL--N....SD...AS-SCH.QLKGNV..S.....E...N...D.K....K-
MDA5_Ictr ...D.....K.....D..E..L..-V.H...TSV..DD---..D....G.....N...K.....K....K-
MDA5_Capo D..I.....N.....HD.AA.....-Q.....DD.C.G------.K.K.NT.RLV.E.K.....GKI.VVILS

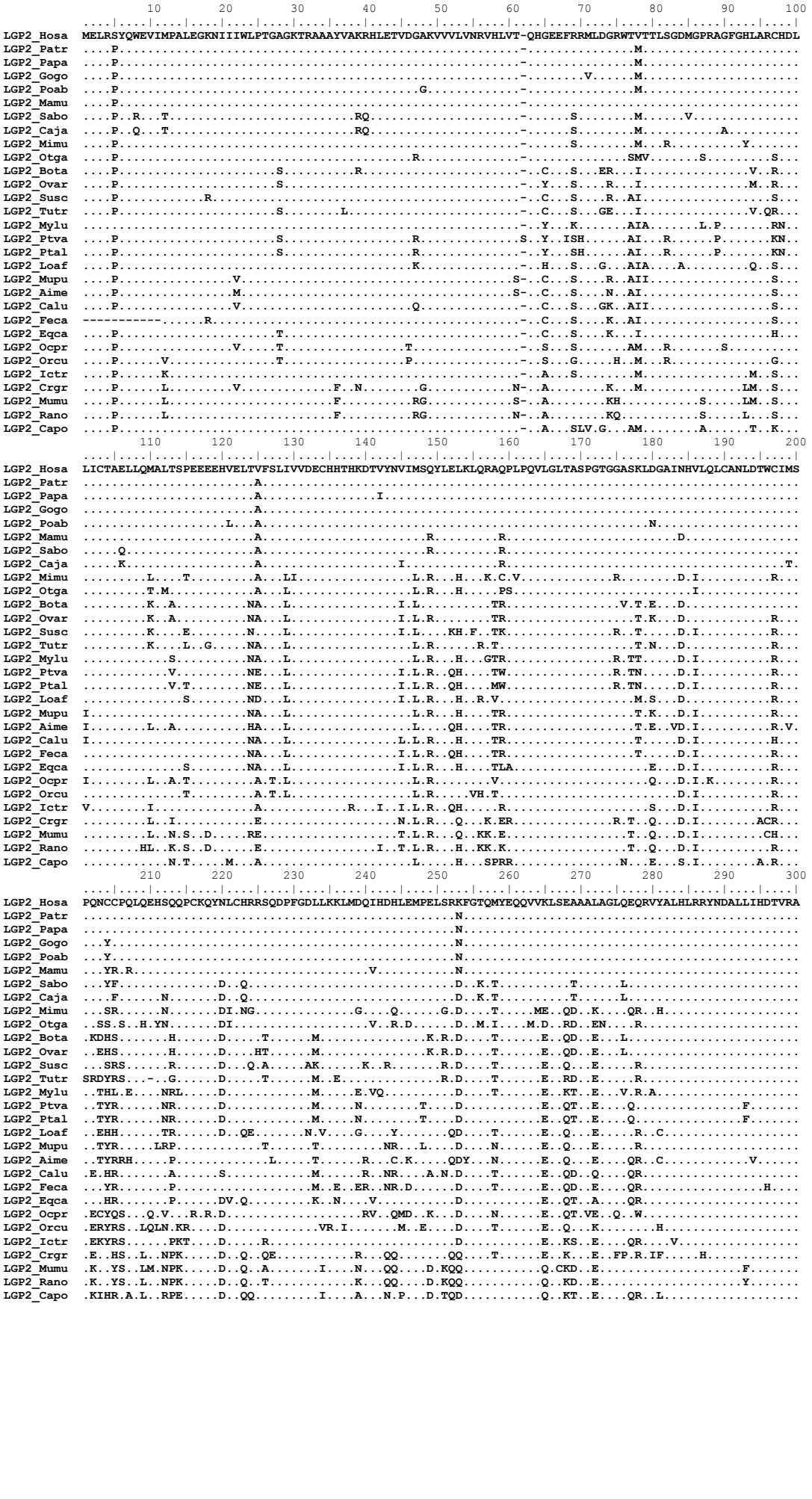

710 720 730 740 750 760 770 780 790 800
MDA5_Hosa LAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIAT
MDA5_Gogo .....
MDA5_Patr .....
MDA5_Papa .....
MDA5_Poab .....
MDA5_Nole .....
MDA5_Mamu .....
MDA5_Sabo .....H.....
MDA5_Caja .....H.....A.....G.....Q.....
MDA5_Otga ..K..KH..Q.....S.....
MDA5_Bota ..Q...H.....I.....S...G.....I.....S.....
MDA5_Ovar ..Q...H.....I.....S...G.....S.....
MDA5_Susc .....HF.....D.K.....R.....
MDA5_Mupu .....KH.....K.....
MDA5_Aime .....KH.....I.....K.....
MDA5_Calu ..Q..KH.....I.....F.....P.....F.....K.....
MDA5_Eqca .....C.....H.....
MDA5_Mylu .....Q.....AG.....
MDA5_Ptal .....H.....A.A.....S.....
MDA5_Loaf .....QH.....E.....G.....F.....D.....TH.....
MDA5_Orcu ..D...H.....I.....
MDA5_Crgr .....K.....I...K..L..F.K....S.....T.....V.....
MDA5_Mumu .....K.....I.....L..F..S...S.....T.....M..A.....V.....T.....E.....
MDA5_Rano .....K.....I.....L..F..S...S.....T.....M.....V.....T.....E.....
MDA5_Ictr .....H.....L.....G.....S.....
MDA5_Capo ..D...H.....S.....L..F....KPT.....A.....D..I.....H.....

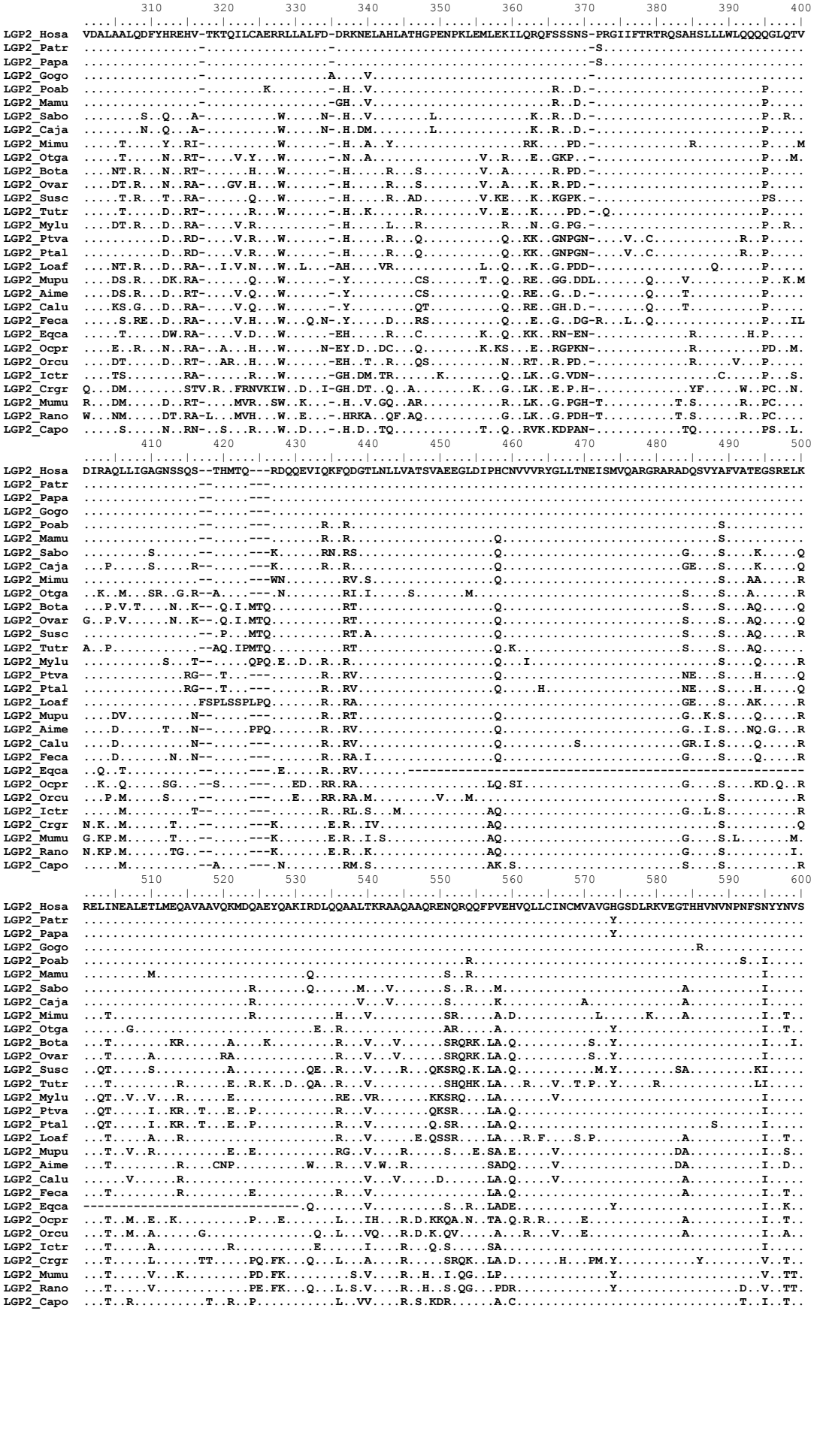

810 820 830 840 850 860 870 880 890 900
MDA5_Hosa TVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTK
MDA5_Gogo .....R.....
MDA5_Patr .....R.....
MDA5_Papa .....R.....
MDA5_Poab .....R.....
MDA5_Nole .....R.....
MDA5_Mamu .....R.....
MDA5_Sabo .....V.....P.S.....R.....Q.....RH.....I.
MDA5_Caja .....S.S.....R.....H....E.....I.
MDA5_Otga .....L...S.....R....A.....DH.....A.....I.
MDA5_Bota .....Q.....V.R.....DR.....
MDA5_Ovar .....Q.....V.R.....DR.....
MDA5_Susc .....Q.....R.....DR.....I.
MDA5_Mupu .....N.....R.I.....DH....N.....R.....I.
MDA5_Aime .....N.....R.I.....DH....N.....R.....I.
MDA5_Calu .....N.....R.I.....DH....N.....I.
MDA5_Eqca .....V.R....A.....DH.....W.....
MDA5_Mylu .....V.R.....Q.....DR.....I.
MDA5_Ptal .....C.....VDR.....V.....AR.....I.
MDA5_Loaf .....I...P.S...V.R.I.....DR.....D.D.....I.
MDA5_Orcu .....I..GC.....V.R.....T...R.....E.....M...I.
MDA5_Crgr .....TS.....T.R.I.....NR.....RE..A..V...L.....V.
MDA5_Mumu .....TS.....T.R.I.....NR.....V...L.....V.
MDA5_Rano .....TS.....T.R.I.....NR.....V...L.....V.
MDA5_Ictr .....E....IV..P.....L.R.I.....A.....F...L.....
MDA5_Capo .....I...S.....E.....V.R.....H....N.....Q...VF...N...N.....

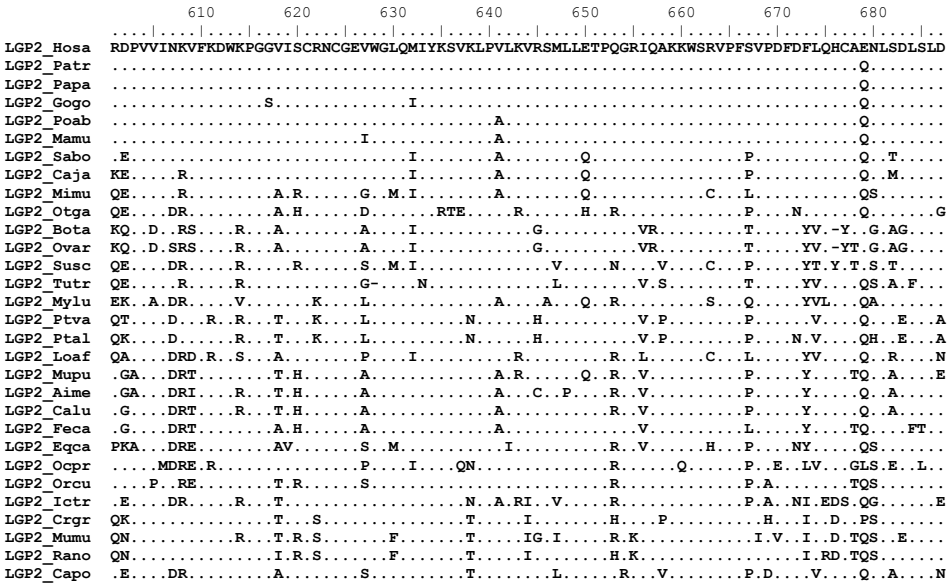

910 920 930 940 950 960 970 980 990 1000
MDA5_Hosa RNIAKHYKNNPSLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEIICK-CGQAWGTMMVHKGLDLPCLKIRNFV
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo


1010 1020 1030
MDA5_Hosa VVFKNNSTKKQYKKWVELPITFPNLDYSECCLFSDED*
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

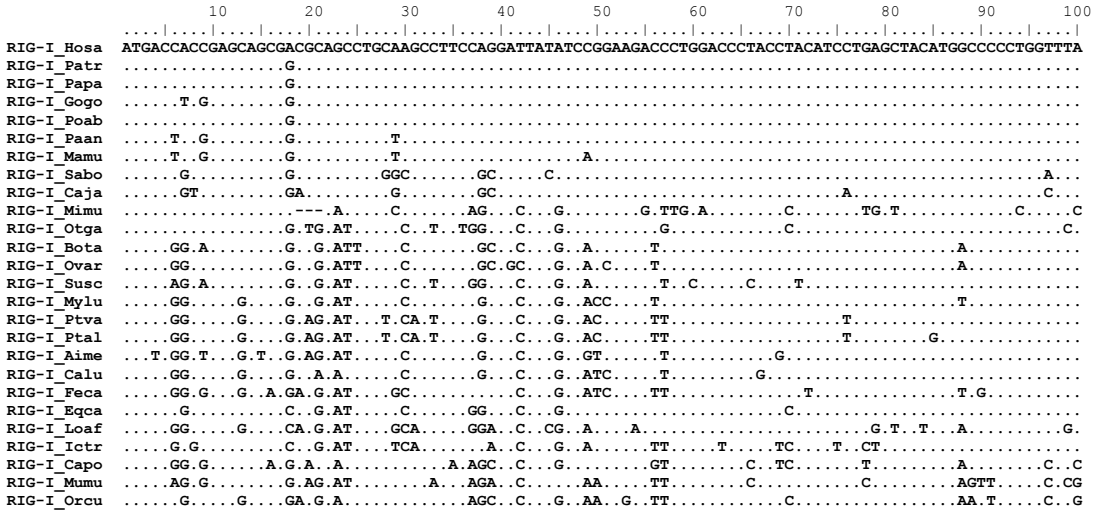
Supplementary Figure S6

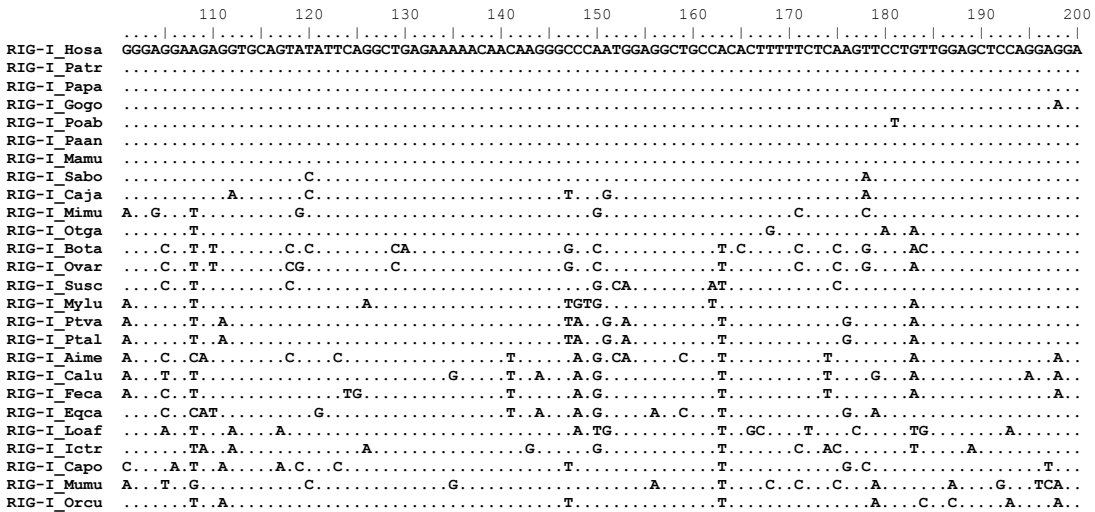


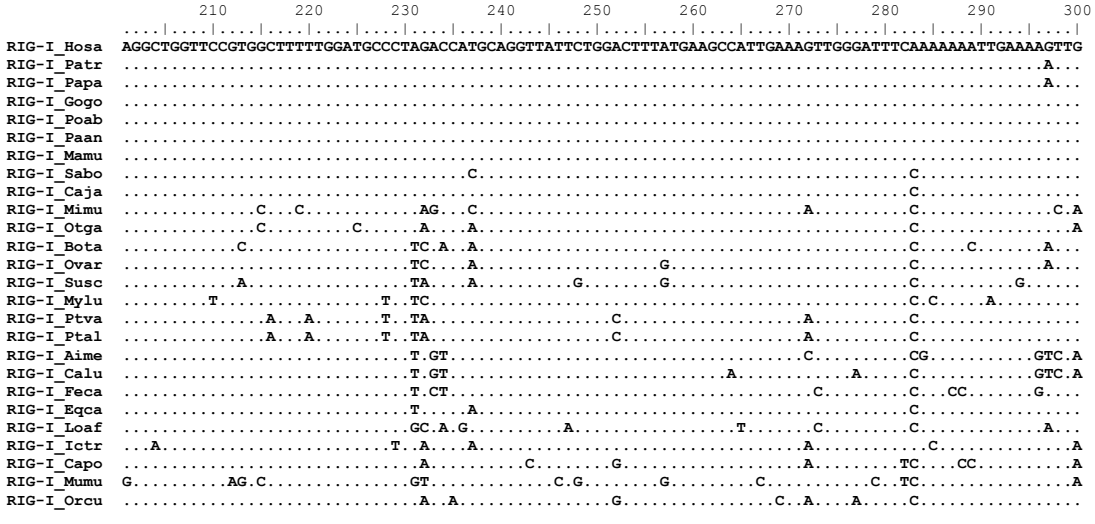

310 320 330 340 350 360 370 380 390 400
LGP2_Hosa VDALAALQDFYHREHV-TKTQILCAERRLLALFD-DRKNELAHLATHGPENPKLEMLEKILQRQFSSSNS-PRGIIFTRTRQSAHSLLLWLQQQQGLQTV
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
410 420 430 440 450 460 470 480 490 500
LGP2_Hosa DIRAQLLIGAGNSSQS--THMTQ---RDQQEVIQKFQDGTLNLLVATSVAEEGLDIPHCNVVVRYGLLTNEISMVQARGRARADQSVYAFVATEGSRELK
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
510 520 530 540 550 560 570 580 590 600
LGP2_Hosa RELINEALETLMEQAVAAVQKMDQAEYQAKIRDLQQAALTKRAAQAAQRENQRQQFPVEHVQLLCINCMVAVGHGSDLRKVEGTHHVNVNPNFSNYYNVS
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo

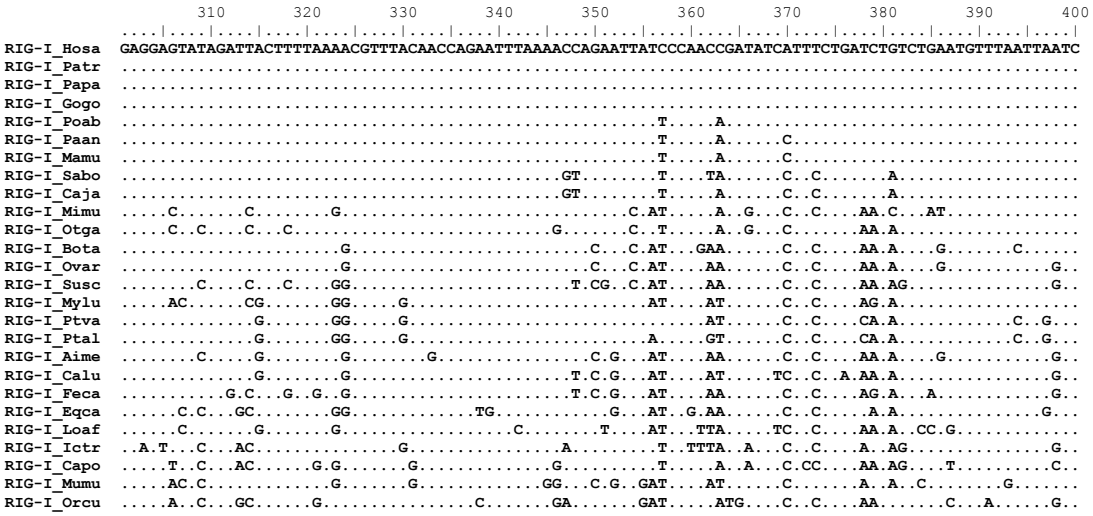

610 620 630 640 650 660 670 680
LGP2_Hosa RDPVVINKVFKDWKPGGVISCRNCGEVWGLQMIYKSVKLPVLKVRSMLLETPQGRIQAKKWSRVPFSVPDFDFLQHCAENLSDLSLD
LGP2_Patr .....Q.....
LGP2_Papa .....Q.....
LGP2_Gogo .....S.....I.....Q.....
LGP2_Poab .....A.....Q.....
LGP2_Mamu .....I.....A.....Q.....
LGP2_Sabo .E.....I.....A.....Q.....P.....Q..T.....
LGP2_Caja KE.....R.....I.....A.....Q.....P.....Q..M.....
LGP2_Mimu QE.....R.....A.R.....G..M.I.....A.....Q.....C...L.....QS.....
LGP2_Otga QE....DR.....A.H.....D.....RTE.....R.....H..R.....P....N.....Q.....G
LGP2_Bota KQ..D..RS....R...A.....A....I.....G.....VR.....T.....YV.-Y..G.AG....
LGP2_Ovar KQ..D.SRS....R...A.....A....I.....G.....VR.....T.....YV.-YT.G.AG....
LGP2_Susc QE....DR.....R.....R.....S..M.I.....V.....N....V....C...P.....YT.Y.T.S.T.....
LGP2_Tutr QE.....R.....R.....G-....N.....L.....V.S.....T.....YV....QS.A.F...
LGP2_Mylu EK..A.DR.....V.....K....L.....A....A...Q..R.....S...Q.....YVL...QA.....
LGP2_Ptva QT....D...R..R...T...K....L.....N.....H.....V.P.....P.....V....Q...E...A
LGP2_Ptal QK....D.....R...T...K....L.....N.....H.....V.P.....P....N.V....QH..E...A
LGP2_Loaf QA....DRD.R..S...A.....P....I.....R.....R..L.....C...L.....YV....Q..R....N
LGP2_Mupu .GA...DRT.....T.H.....A.....A.R.....Q..R..V.....P.....Y....TQ..A....E
LGP2_Aime .GA...DRI....R...T.H.....A.....A...C..P....R..V.....P.....Y.....Q..A.....
LGP2_Calu .G....DRT....R...T.H.....A.....A.....R..V.....P.....Y.....Q..A.....
LGP2_Feca .G....DRT.....A.H.....A.....A.....V.....L.....Y....TQ....FT..
LGP2_Eqca PKA...DRE.....AV.....S..M.....I.....R..V.....H...P....NY.....QS.....
LGP2_Ocpr .....MDRE.R.....P....I....QN.....R.....Q.....P..E..LV...GLS.E..L..
LGP2_Orcu ....P..RE.....T.R.....S.....R.....P.A.....TQS.....
LGP2_Ictr .E....DR.....R...T.....N..A.RI..V.....R.....P.A..NI.EDS.QG.....E
LGP2_Crgr QK.....T...S.....T.....I.....H....P.....H...I..D..PS.....
LGP2_Mumu QN.....R...T.R.S.....F.....T.....IG.I.....R.K.....I.V..I..D.TQS..E....
LGP2_Rano QN.....I.R.S.....F.....T.....I.....H.K.....I.RD.TQS.....
LGP2_Capo .E....DR.....A.....S.....T.....L.....R...V.....P.D....V....Q..A....N

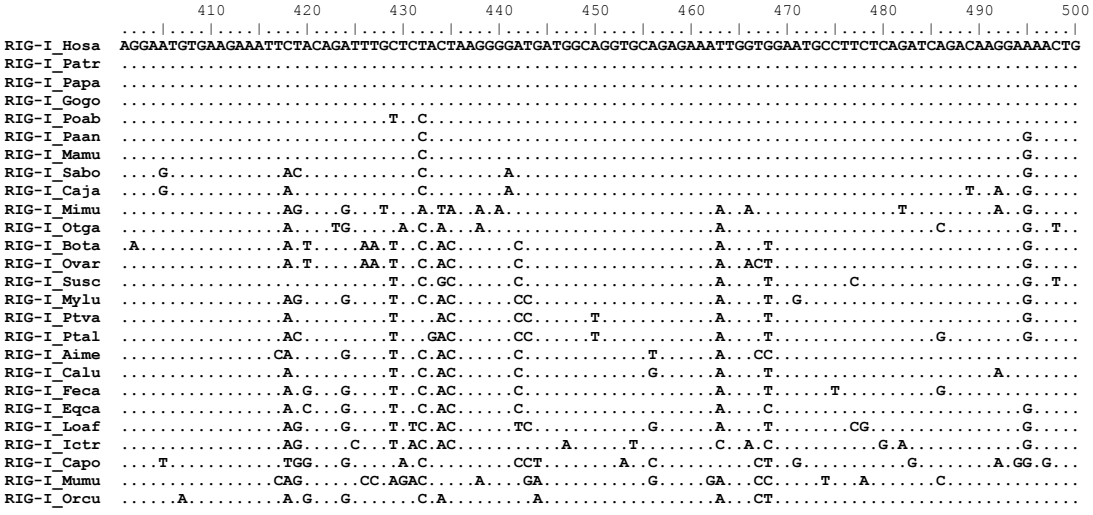
Supplementary Figure S7

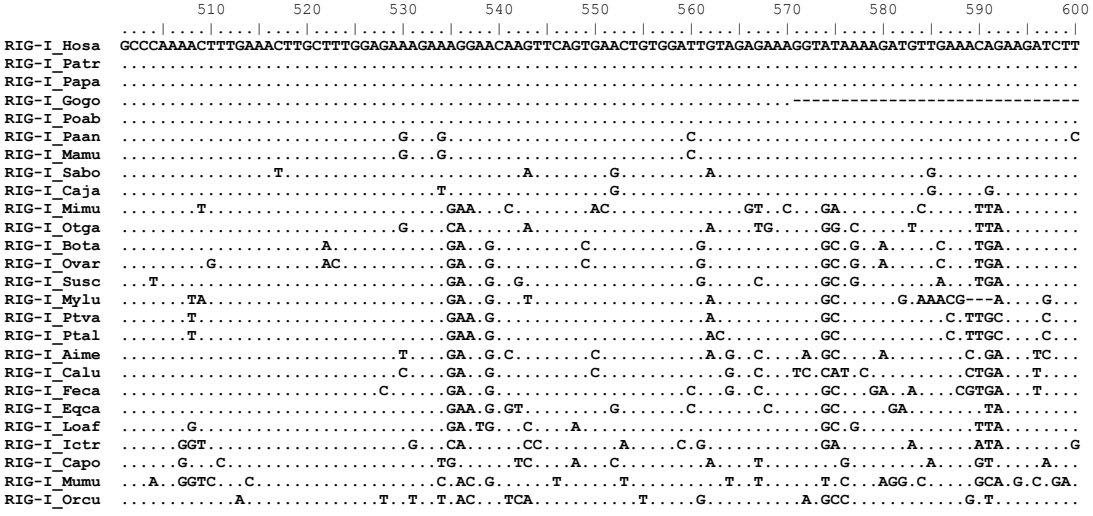







310 320 330 340 350 360 370 380 390 400
RIG-I_Hosa GAGGAGTATAGATTACTTTTAAAACGTTTACAACCAGAATTTAAAACCAGAATTATCCCAACCGATATCATTTCTGATCTGTCTGAATGTTTAATTAATC
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

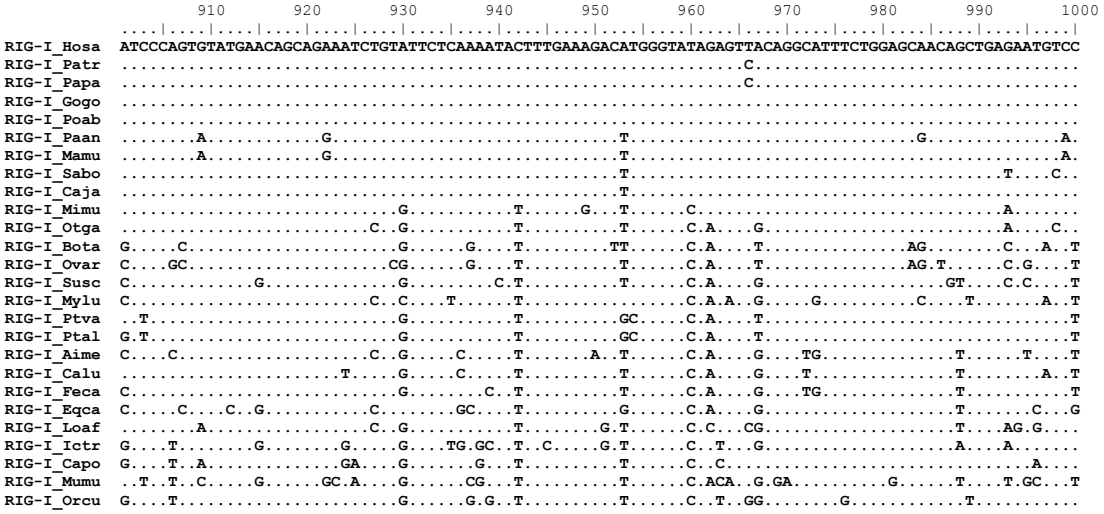

410 420 430 440 450 460 470 480 490 500
RIG-I_Hosa AGGAATGTGAAGAAATTCTACAGATTTGCTCTACTAAGGGGATGATGGCAGGTGCAGAGAAATTGGTGGAATGCCTTCTCAGATCAGACAAGGAAAACTG
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

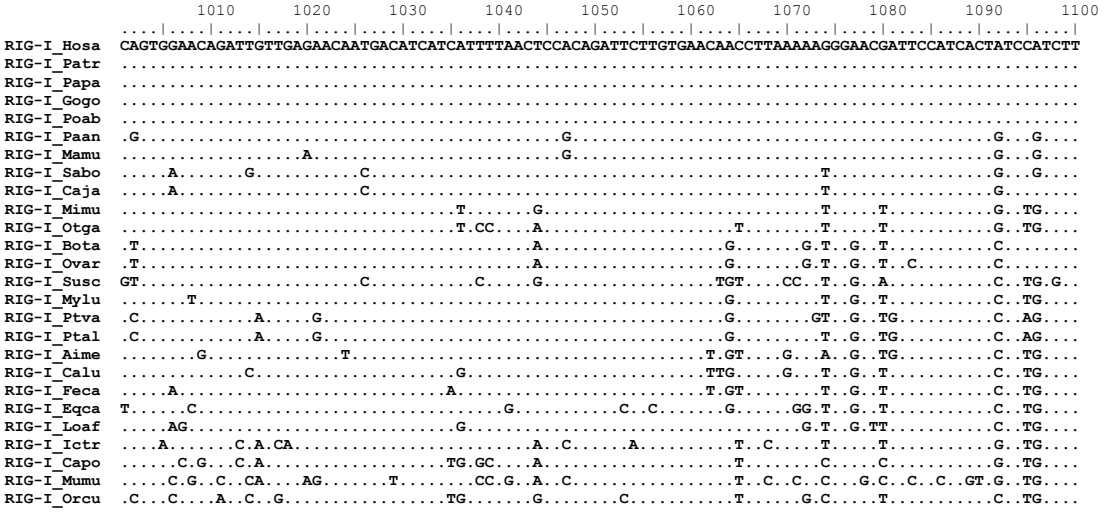

510 520 530 540 550 560 570 580 590 600
RIG-I_Hosa GCCCAAAACTTTGAAACTTGCTTTGGAGAAAGAAAGGAACAAGTTCAGTGAACTGTGGATTGTAGAGAAAGGTATAAAAGATGTTGAAACAGAAGATCTT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

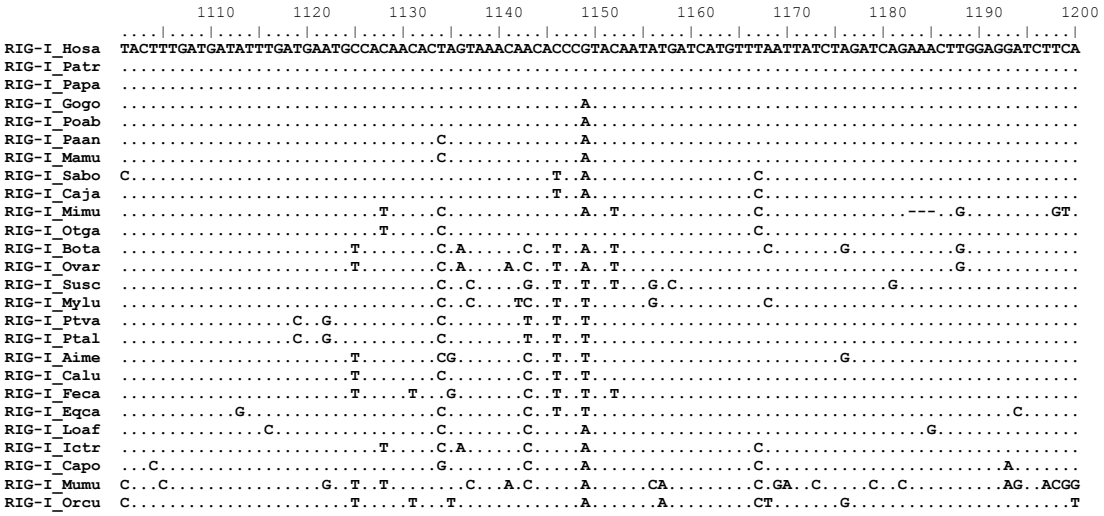
```
                610       620       630       640       650       660       670       680       690       700
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
RIG-I_Hosa GAGGATAAG---ATGGAAACTTCTGACATACAGATTTTCTACCAAGAAGATCCAGAATGCCAGAATCTTAGTGAGAATTCATGTCCACCTTCAGAAGTGT
RIG-I_Patr .........---..........................................................................................
RIG-I_Papa .........---..........................................................................................
RIG-I_Gogo ----------------------------------------------------------------------------------------.......
RIG-I_Poab .........---..........................................................................................
RIG-I_Paan .........---..........................................................................................
RIG-I_Mamu .........---.....................................................................T..............
RIG-I_Sabo .....G...---.....................................................C.....................TG.............
RIG-I_Caja .....G...---T.....................................................C..G.................TG.............
RIG-I_Mimu ......G.TAAC..........C.G.G.................G.........................C.......G...T..........T....
RIG-I_Otga ......G.CGAC..T........A..AG....A...........G........C..G.C..G......C...C.....G..CA.............A.A.
RIG-I_Bota ......G.TGAA...A......TG..TG....C...G.....A.......A......A.............CC.......C...---T...........C
RIG-I_Ovar ......G.TGAA...A......TG............G.....A.G.....A.....AA............CC.....C.C.....TT...........C
RIG-I_Susc .....CG.CGAA...A......G...TG..............A.G.....A......AA...........C.......G...T.CT.........CAC
RIG-I_Mylu ......GG.GAA..A........C...G..............A.G.....A.....................C.........................CA.
RIG-I_Ptva ......G..GAA...A.......C..AG..............A.G.....A.....................................................A.
RIG-I_Ptal ......G..GAA...A.......C..AG..............A.G.....A.....................................................A.
RIG-I_Aime ......G.TGAA..A.............................TA.G.....A.....................C....CC...C...G..........GAA
RIG-I_Calu ......G.AGAA..A........C..AG..............TA.G.....A.....................C..G.CC.....................G
RIG-I_Feca ......G.TGAA.....................................A.G.....AT.T...................C..............T..........C.
RIG-I_Eqca .G....G..GAA..A...........G...............A.G.....A.......A.........G...C........A....G.......CT.CTC
RIG-I_Loaf .G....GGTGAA..............AG....A..A........G.....A.......A.............C...G.....................
RIG-I_Ictr ......G.TGAA..............A.....C.......T...G............CA.............C.......CCAGG....C......A...
RIG-I_Capo ........TAGA......CA.......................G............GA...A.........C....C..C.....................
RIG-I_Mumu ......G.TGGAGC...GG.G..CAG...C.........ATT..G.....G.....G..T.........C...C.....C.CG.G..T..........C..
RIG-I_Orcu .....GG.CGAC....G......CAG......T...G.......G.........C.G.TGG.......C...C...........A.....G.......A.
```

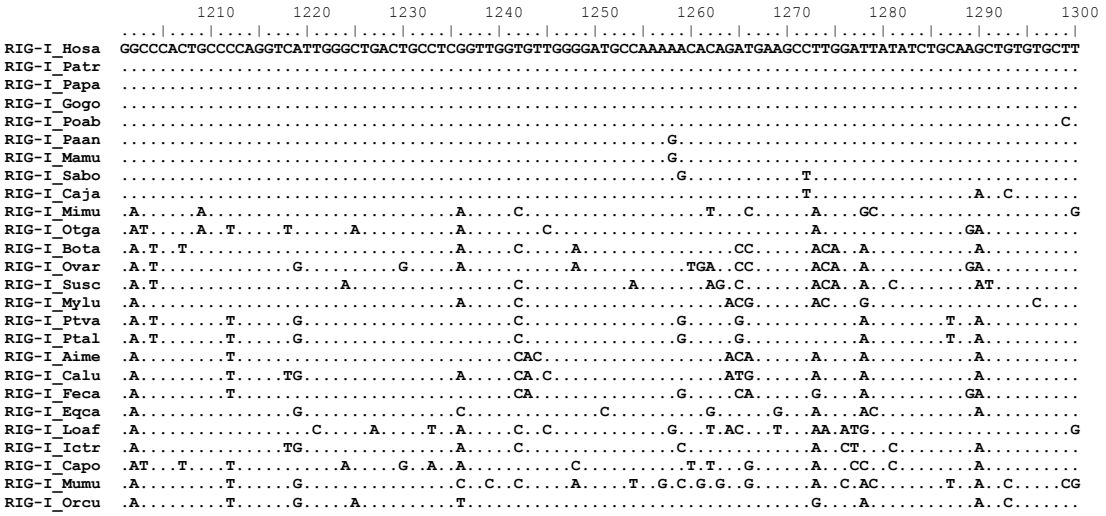
```
                710       720       730       740       750       760       770       780       790       800
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
RIG-I_Hosa CTGATACAAACTTGTACAGCCCATTTAAACCAAGAAATTACCAATTAGAGCTTGCTTTGCCTGCTATGAAAGGAAAAAACACAATAATATGTGCTCCTAC
RIG-I_Patr ..................................................................................................
RIG-I_Papa ..................................................................................................
RIG-I_Gogo ..................................................................................................
RIG-I_Poab G..............G.............T.....................................................................
RIG-I_Paan ............................T.....................................A.....................................
RIG-I_Mamu ............................T.....................................A.....................................
RIG-I_Sabo ........G.....................T.....G.............................C.A.....................................C..
RIG-I_Caja ..................................G.............................C.A.....................................
RIG-I_Mimu ..T....TG.TG..A......G.....................GC.................C....A.....................................
RIG-I_Otga ..C.C..TG.....................................GC.....................A.....................................
RIG-I_Bota ..CC...TT.T------........G.................C.......G.....A......CA...G.....G..T.....................
RIG-I_Ovar ..CC.G.TT.T------........G.............T....C.............A......CA...G.....G..T.....................
RIG-I_Susc ..C....TT..------.....C.G..G........G......C.G...........A......CA...T.....................................
RIG-I_Mylu ..TG...TC.G------......A....G..............C...........CGT......T.AG.....................................
RIG-I_Ptva ..C....TC..------......A.G..G......GG.......C.....................A...C.....................................
RIG-I_Ptal ..C....TC..------......A.G..G......GG.......C.....................A...C.....................................
RIG-I_Aime A.TC.TGCC.T------T...A.C.G..G..............C.T.....C..G..........A.G.....................................
RIG-I_Calu ..TC...TTG.------.C....C.G..G..............C....A..............A.G..........T.....................
RIG-I_Feca ..C...TTT..------.C....G.G..G..............C.....................A.G.....................................
RIG-I_Eqca AGC..CT...TC...TGGC......G..GAT.............C.G.....C...........C.A.G.....................................
RIG-I_Loaf ..C....T.............C..G..G.....................C.............CA...T........T.....C.....................
RIG-I_Ictr ..CGA..T...------------..G.................C.....................A.G..........T.....................
RIG-I_Capo ..C....T..T....G.......C.G...G.............C.............A.......A.G.....................................
RIG-I_Mumu ..TC..AT..T..AC..........G.................C.G........CC.......C.A......G.....T..............C.....
RIG-I_Orcu ..CGCG.T..A..A.G.....A.C.G.................GC...........C.............G.....................................
```

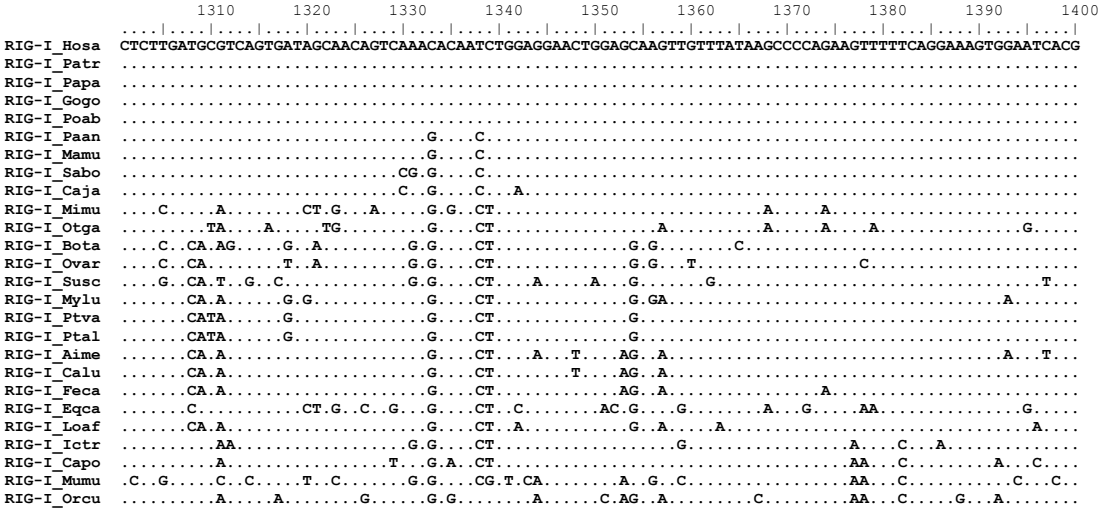
```
                810       820       830       840       850       860       870       880       890       900
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
RIG-I_Hosa AGGTTGTGGAAAAACCTTTGTTTCACTGCTTATATGTGAACATCATCTTAAAAAATTCCCACAAGGACAAAAGGGGAAAGTTGTCTTTTTTGCGAATCAG
RIG-I_Patr ..................................................................................................T.....
RIG-I_Papa ..................................................................................................T.....
RIG-I_Gogo ..................................................................................................T.....
RIG-I_Poab ..................................................................................................T.....
RIG-I_Paan ................C.G.....G.C.....................................C.............................A.....
RIG-I_Mamu ................C.G...G..G.C.....................................C.............................A.....
RIG-I_Sabo ....................G....C.....................................T.............................T.....
RIG-I_Caja ..................A...G....C.....................................G.............................C..T.....
RIG-I_Mimu T............---...........T..C.................G.......TT....G........A.....G.G...........T.....A
RIG-I_Otga T.....................T.............G.................TT...........GAA.A..G.....T........T.....A
RIG-I_Bota T.....................C..T........C.............G.....T.........G......A.........G.......TGT...A
RIG-I_Ovar T.................G.C..T........C.............GC....T..G.......GC.....................TGT...A
RIG-I_Susc T.....................T..T.....................G.....T.........G.........GA.....T.....T.T...A
RIG-I_Mylu T........G...........G....T..C.................C.G.....T..G.....................GA....G........T.....A
RIG-I_Ptva T.....................T.....................C.G.....T.............A.....G.....G........TGT...A
RIG-I_Ptal T.....................T.....................C.......T.............A.....G.....G........TGT...A
RIG-I_Aime T.....................C..T.....................T.............A.....G.....T........T.T....
RIG-I_Calu T.................G.C..T.....................T..C.................GA.............T.T...A
RIG-I_Feca T.....................T...........G.............T..T..G.................G.............TGT...A
RIG-I_Eqca T.............G.....G..T.....................G.....T.........G.........G..G...........T.T...T
RIG-I_Loaf T.....................CA.T.....................T.............A.....GA.............AGT...A
RIG-I_Ictr T........G........A........T.....T.................T..T.........G...A..........CT........T.....T
RIG-I_Capo T........G.............T.....................CT...G..T.........GG..A.....................T..C..T
RIG-I_Mumu T.....................G..G..T.............C.................TGT.........A........G.....C..C..T..C..A
RIG-I_Orcu T.....................T.....................C...T.............A.....G..G..G........T.....C
```

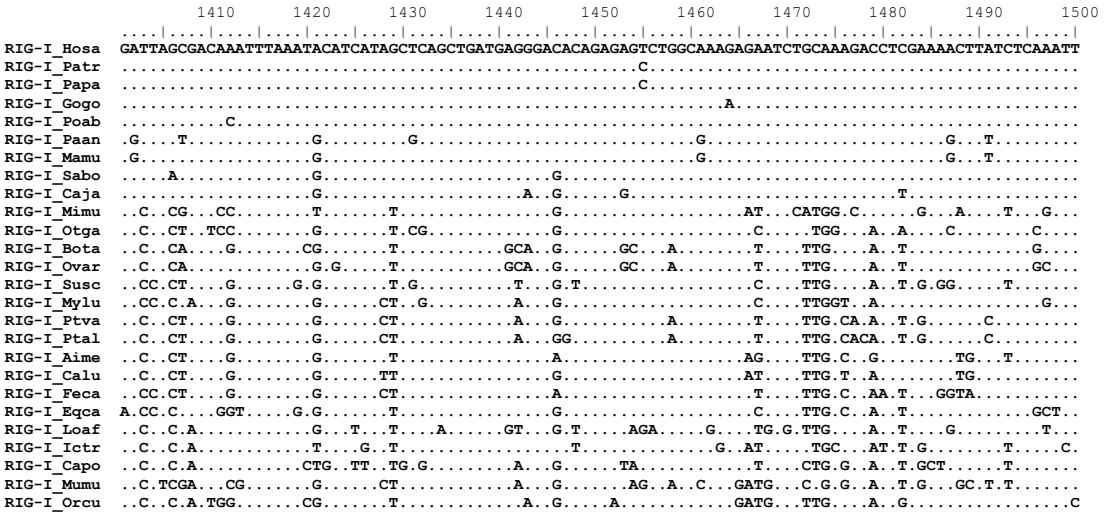

910 920 930 940 950 960 970 980 990 1000
RIG-I_Hosa ATCCCAGTGTATGAACAGCAGAAATCTGTATTCTCAAAATACTTTGAAAGACATGGGTATAGAGTTACAGGCATTTCTGGAGCAACAGCTGAGAATGTCC
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

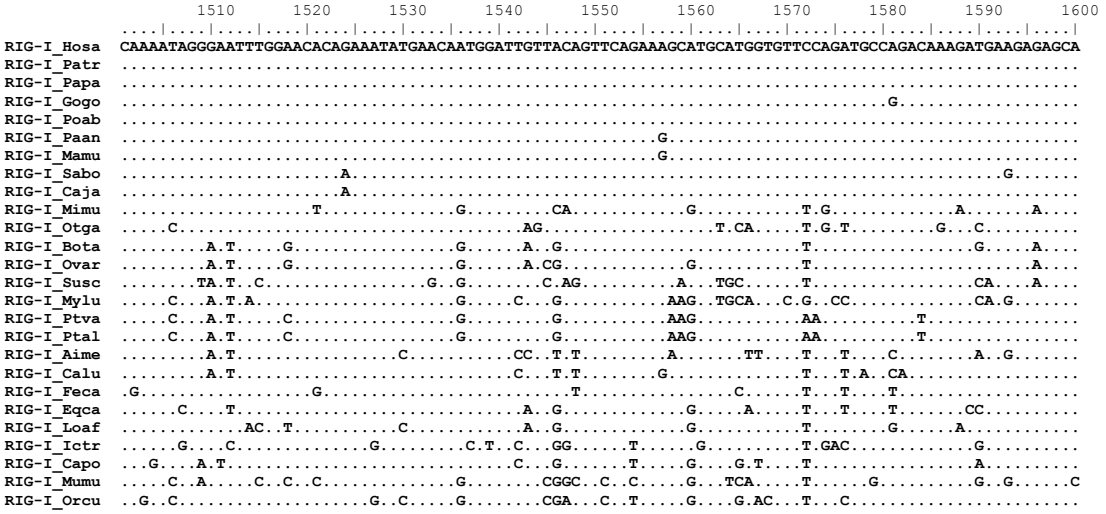

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
RIG-I_Hosa CAGTGGAACAGATTGTTGAGAACAATGACATCATCATTTTAACTCCACAGATTCTTGTGAACAACCTTAAAAAGGGAACGATTCCATCACTATCCATCTT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

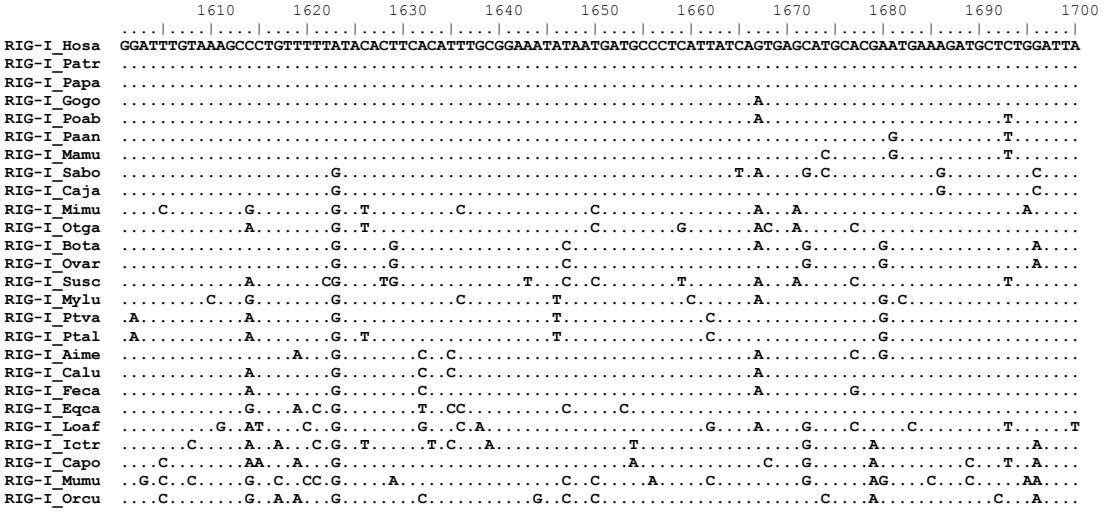

1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
RIG-I_Hosa TACTTTGATGATATTTGATGAATGCCACAACACTAGTAAACAACACCCGTACAATATGATCATGTTTAATTATCTAGATCAGAAACTTGGAGGATCTTCA
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

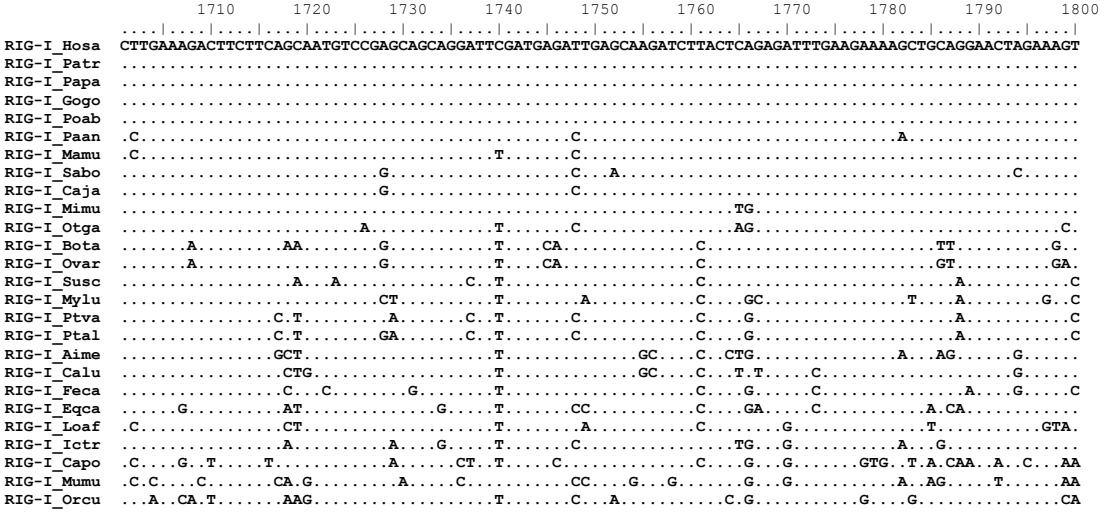

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
RIG-I_Hosa GGCCCACTGCCCCAGGTCATTGGGCTGACTGCCTCGGTTGGTGTTGGGGATGCCAAAAACACAGATGAAGCCTTGGATTATATCTGCAAGCTGTGTGCTT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

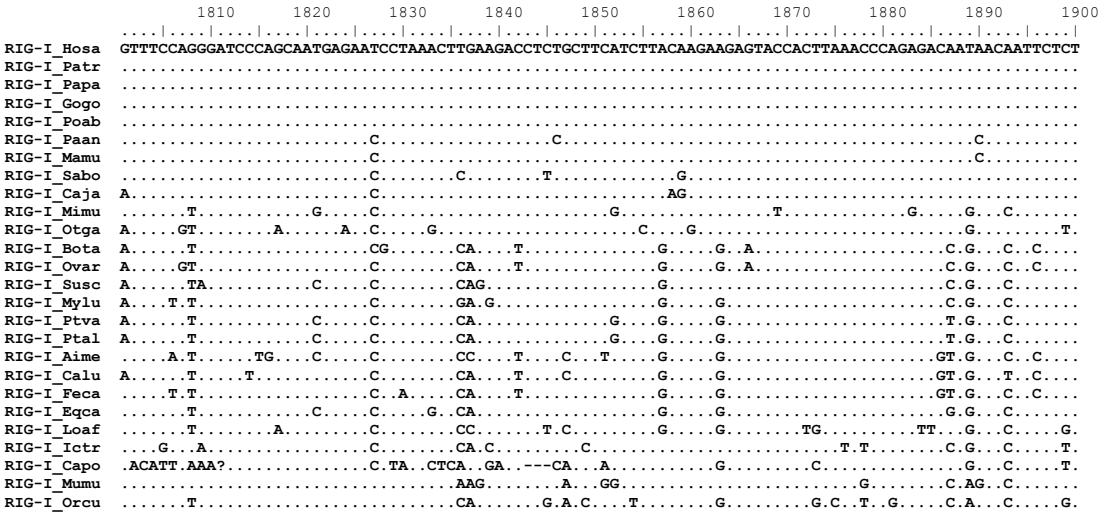

1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
RIG-I_Hosa CTCTTGATGCGTCAGTGATAGCAACAGTCAAACACAATCTGGAGGAACTGGAGCAAGTTGTTTATAAGCCCCAGAAGTTTTTCAGGAAAGTGGAATCACG
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

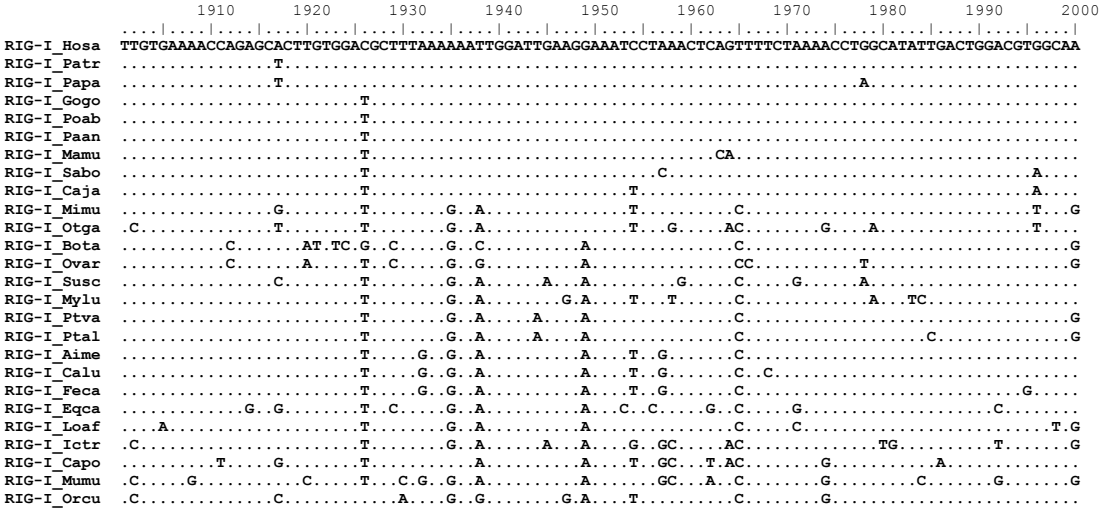

1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
RIG-I_Hosa GATTAGCGACAAATTTAAATACATCATAGCTCAGCTGATGAGGGACACAGAGAGTCTGGCAAAGAGAATCTGCAAAGACCTCGAAAACTTATCTCAAATT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

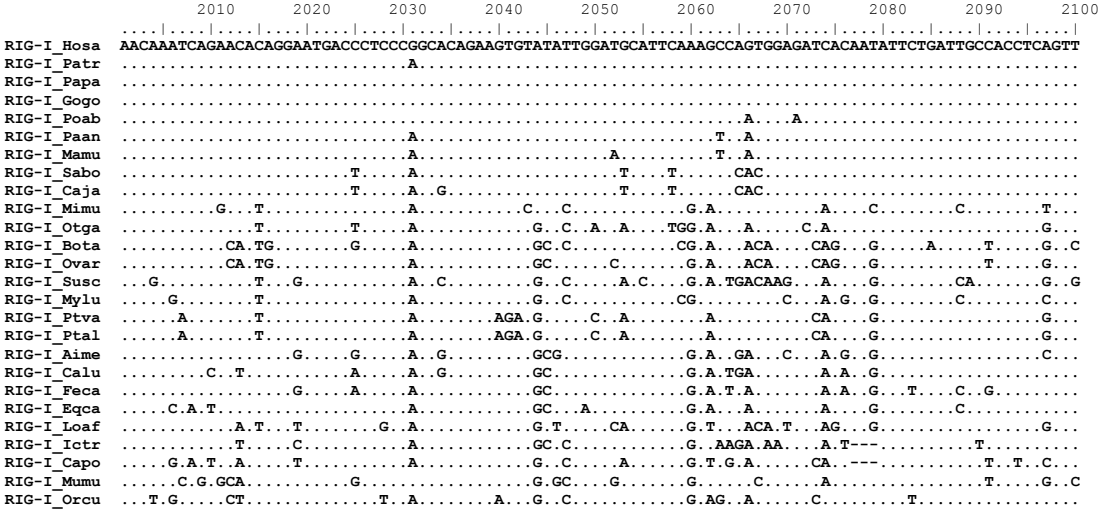

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
RIG-I_Hosa CAAAATAGGGAATTTGGAACACAGAAATATGAACAATGGATTGTTACAGTTCAGAAAGCATGCATGGTGTTCCAGATGCCAGACAAAGATGAAGAGAGCA
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

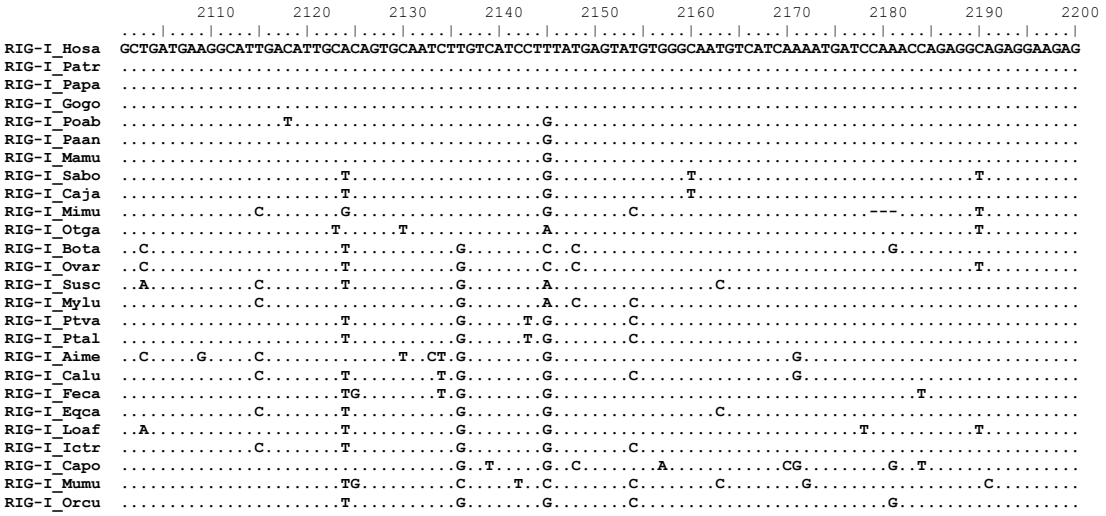

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
RIG-I_Hosa GGATTTGTAAAGCCCTGTTTTTATACACTTCACATTTGCGGAAATATAATGATGCCCTCATTATCAGTGAGCATGCACGAATGAAAGATGCTCTGGATTA
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

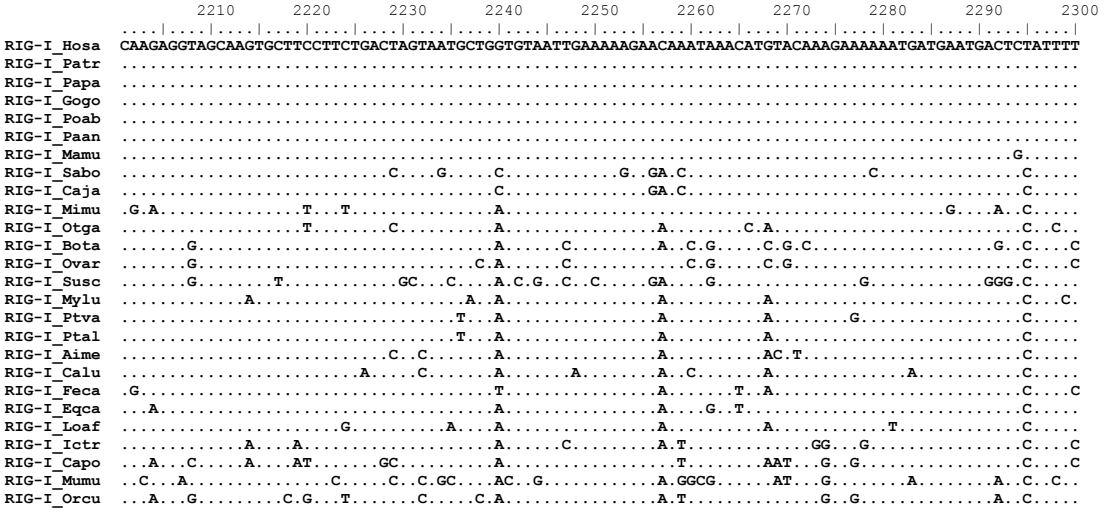

1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
RIG-I_Hosa CTTGAAAGACTTCTTCAGCAATGTCCGAGCAGCAGGATTCGATGAGATTGAGCAAGATCTTACTCAGAGATTTGAAGAAAAGCTGCAGGAACTAGAAAGT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

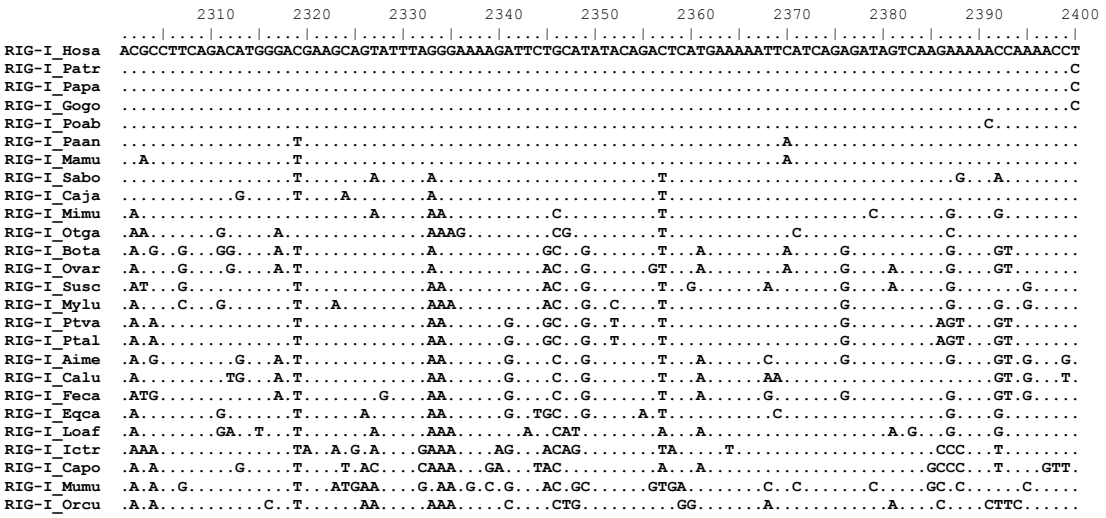

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
RIG-I_Hosa GTTTCCAGGGATCCCAGCAATGAGAATCCTAAACTTGAAGACCTCTGCTTCATCTTACAAGAAGAGTACCACTTAAACCCAGAGACAATAACAATTCTCT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

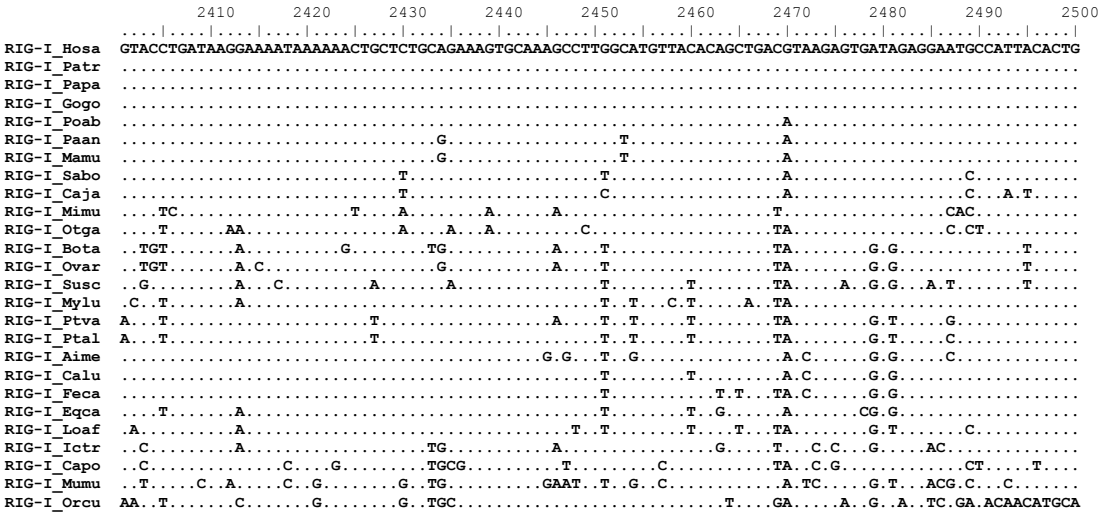

1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
RIG-I_Hosa TTGTGAAAACCAGAGCACTTGTGGACGCTTTAAAAAATTGGATTGAAGGAAATCCTAAACTCAGTTTTCTAAAACCTGGCATATTGACTGGACGTGGCAA
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

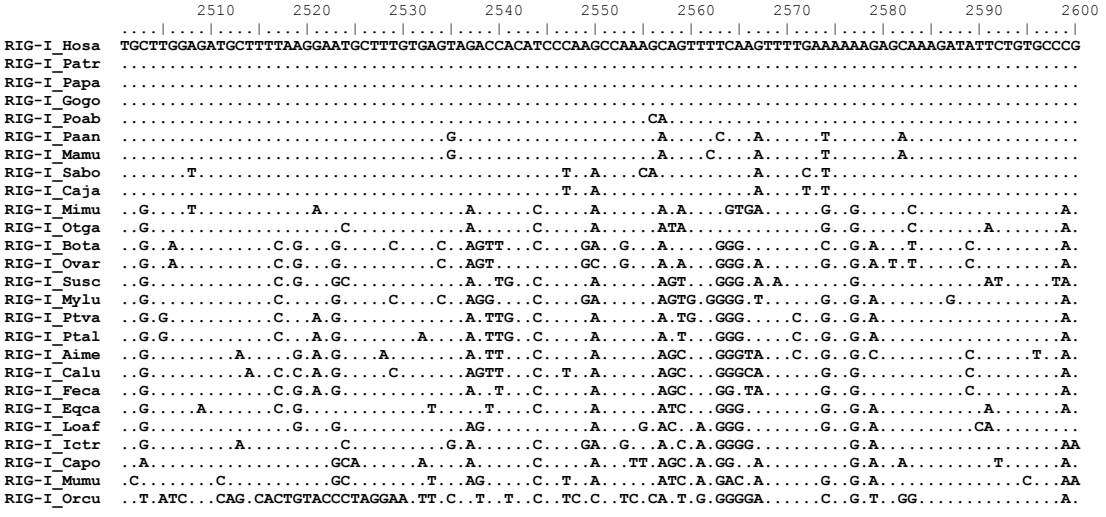

2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
RIG-I_Hosa AACAAATCAGAACACAGGAATGACCCTCCCGGCACAGAAGTGTATATTGGATGCATTCAAAGCCAGTGGAGATCACAATATTCTGATTGCCACCTCAGTT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

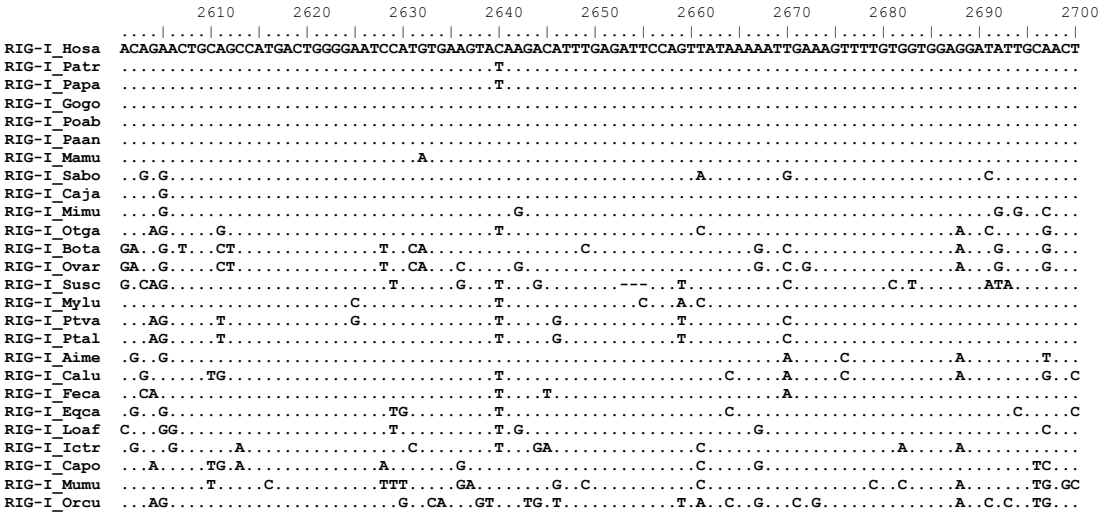

2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
RIG-I_Hosa GCTGATGAAGGCATTGACATTGCACAGTGCAATCTTGTCATCCTTTATGAGTATGTGGGCAATGTCATCAAAATGATCCAAACCAGAGGCAGAGGAAGAG
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

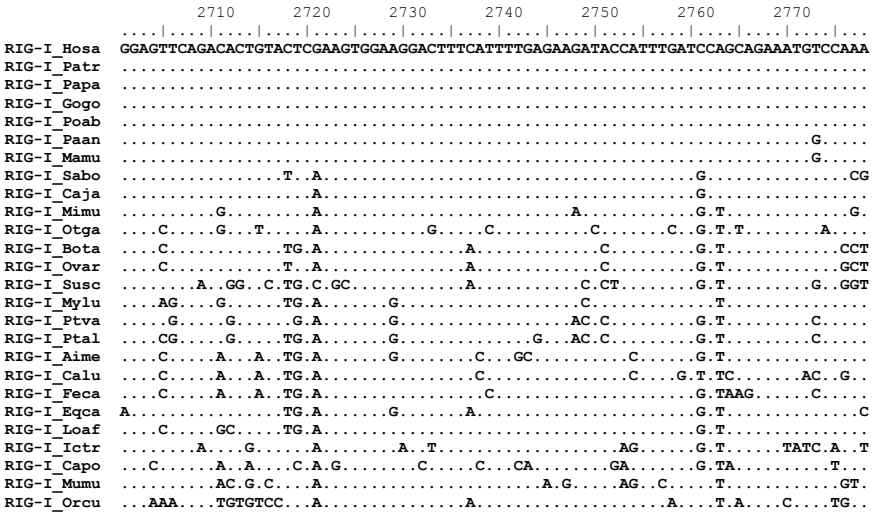

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
RIG-I_Hosa CAAGAGGTAGCAAGTGCTTCCTTCTGACTAGTAATGCTGGTGTAATTGAAAAAGAACAAATAAACATGTACAAAGAAAAAATGATGAATGACTCTATTTT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu


2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
RIG-I_Hosa ACGCCTTCAGACATGGGACGAAGCAGTATTTAGGGAAAAGATTCTGCATATACAGACTCATGAAAAATTCATCAGAGATAGTCAAGAAAAACCAAAACCT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

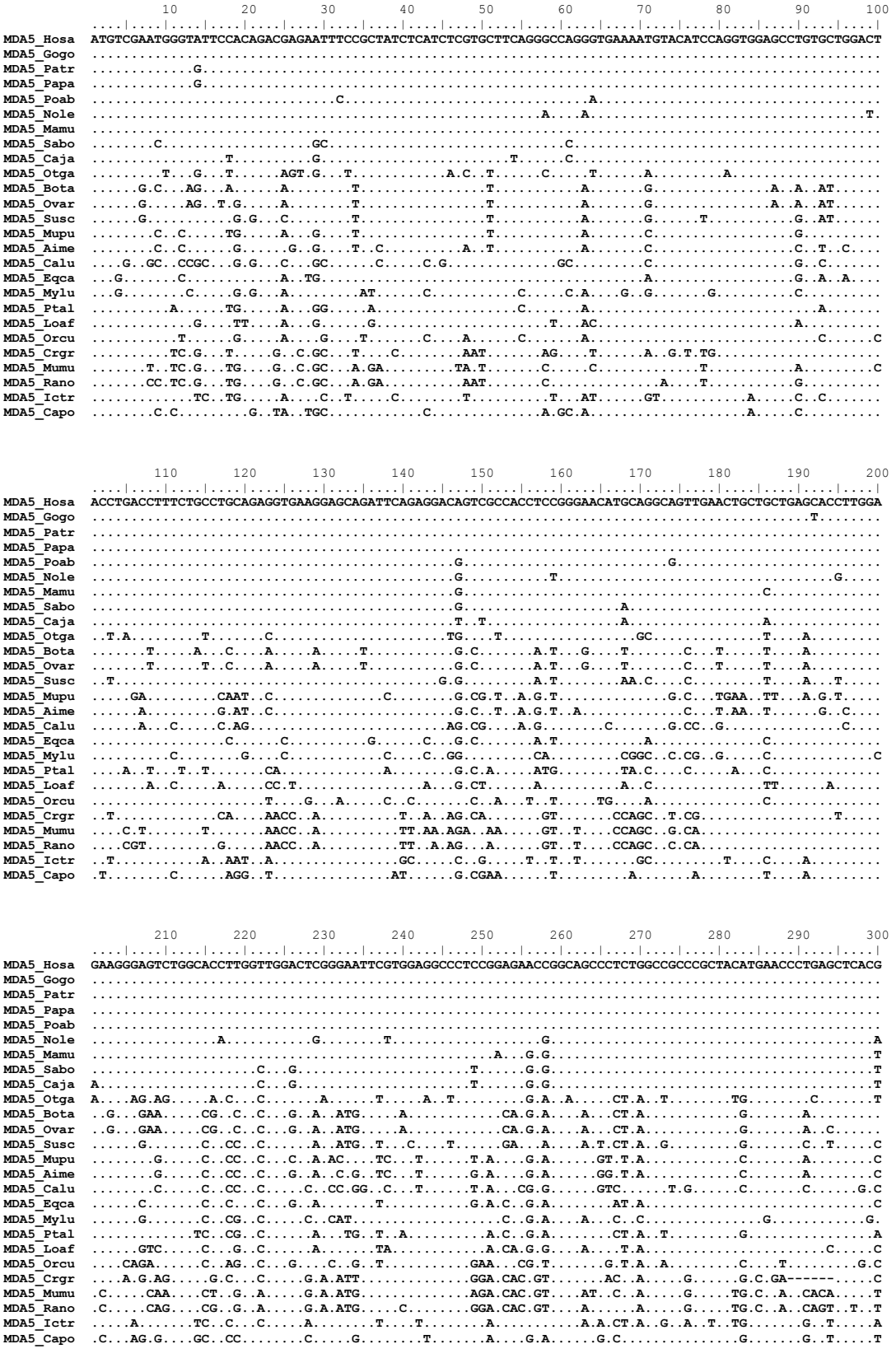

2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
RIG-I_Hosa GTACCTGATAAGGAAAATAAAAAACTGCTCTGCAGAAAGTGCAAAGCCTTGGCATGTTACACAGCTGACGTAAGAGTGATAGAGGAATGCCATTACACTG
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

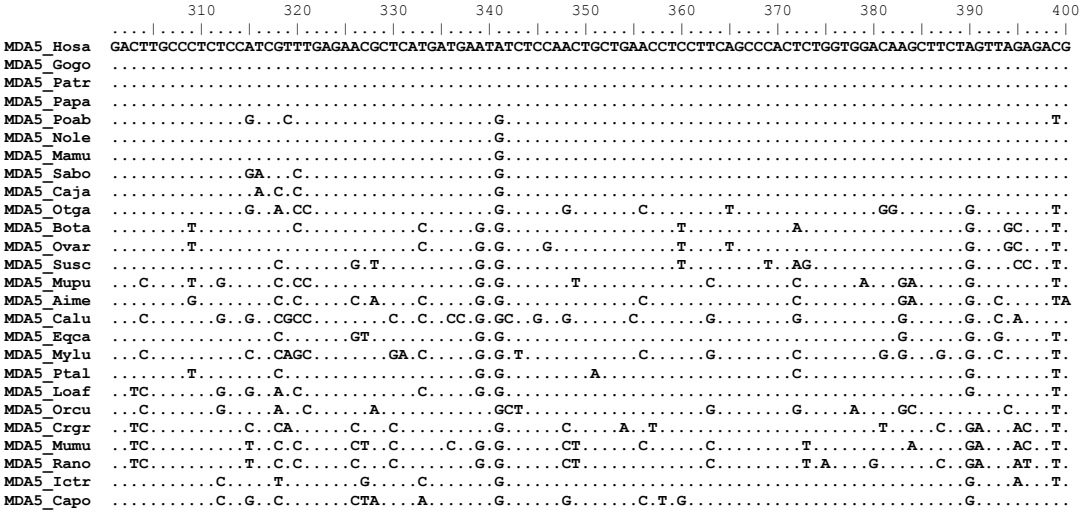

2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
RIG-I_Hosa TGCTTGGAGATGCTTTTAAGGAATGCTTTGTGAGTAGACCACATCCCAAGCCAAAGCAGTTTTCAAGTTTTGAAAAAAGAGCAAAGATATTCTGTGCCCG
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

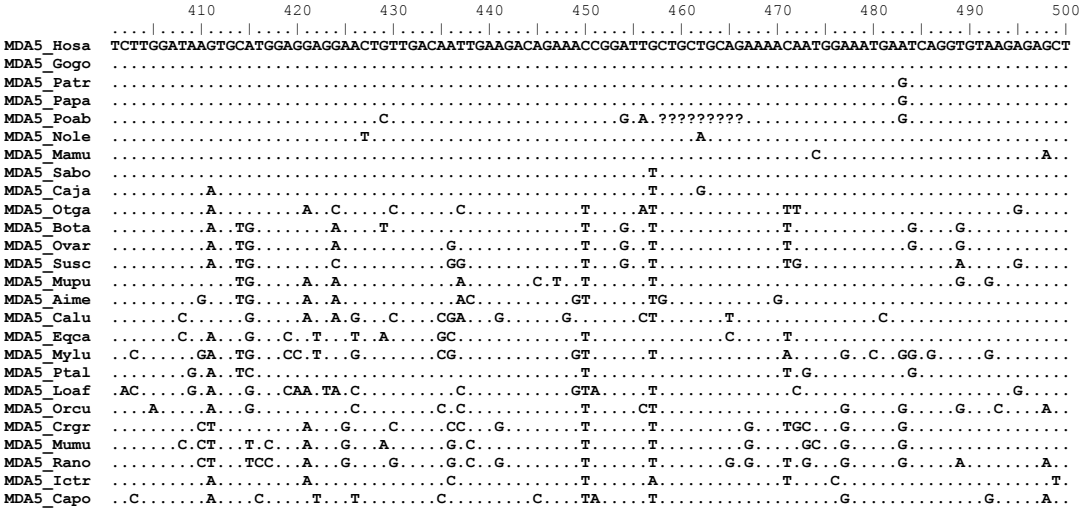

2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
RIG-I_Hosa ACAGAACTGCAGCCATGACTGGGGAATCCATGTGAAGTACAAGACATTTGAGATTCCAGTTATAAAAATTGAAAGTTTTGTGGTGGAGGATATTGCAACT
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

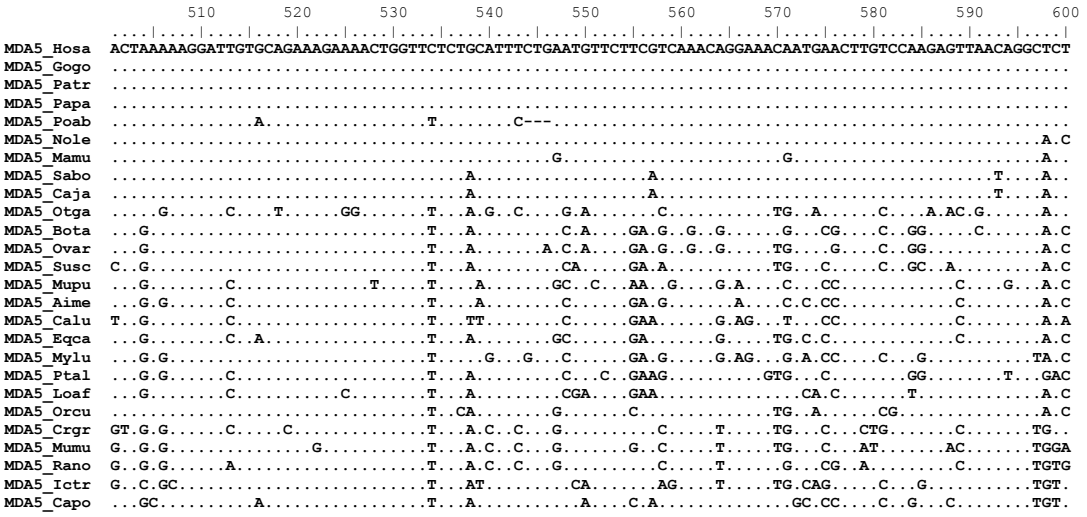

2710 2720 2730 2740 2750 2760 2770
RIG-I_Hosa GGAGTTCAGACACTGTACTCGAAGTGGAAGGACTTTCATTTTGAGAAGATACCATTTGATCCAGCAGAAATGTCCAAA
RIG-I_Patr
RIG-I_Papa
RIG-I_Gogo
RIG-I_Poab
RIG-I_Paan
RIG-I_Mamu
RIG-I_Sabo
RIG-I_Caja
RIG-I_Mimu
RIG-I_Otga
RIG-I_Bota
RIG-I_Ovar
RIG-I_Susc
RIG-I_Mylu
RIG-I_Ptva
RIG-I_Ptal
RIG-I_Aime
RIG-I_Calu
RIG-I_Feca
RIG-I_Eqca
RIG-I_Loaf
RIG-I_Ictr
RIG-I_Capo
RIG-I_Mumu
RIG-I_Orcu

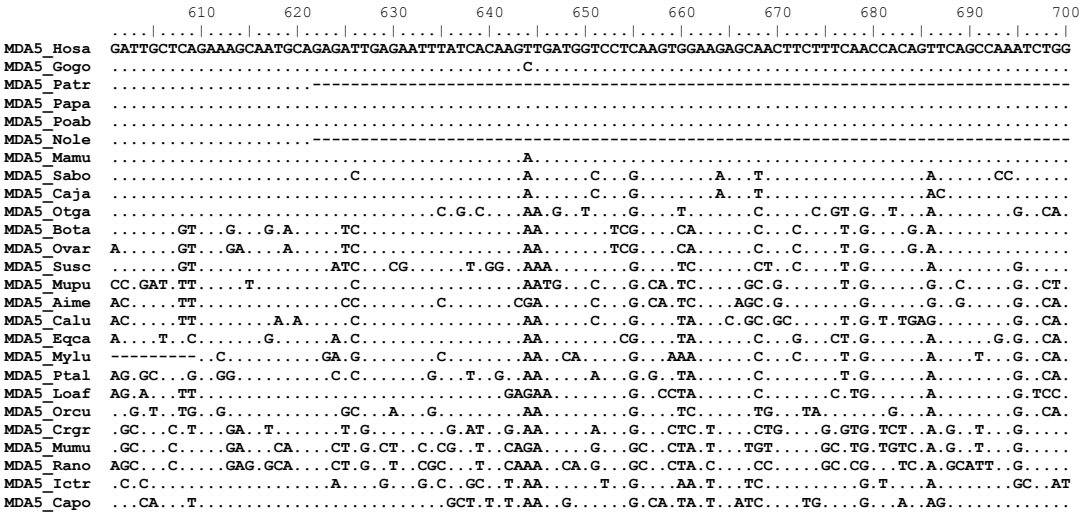
Supplementary Figure S8

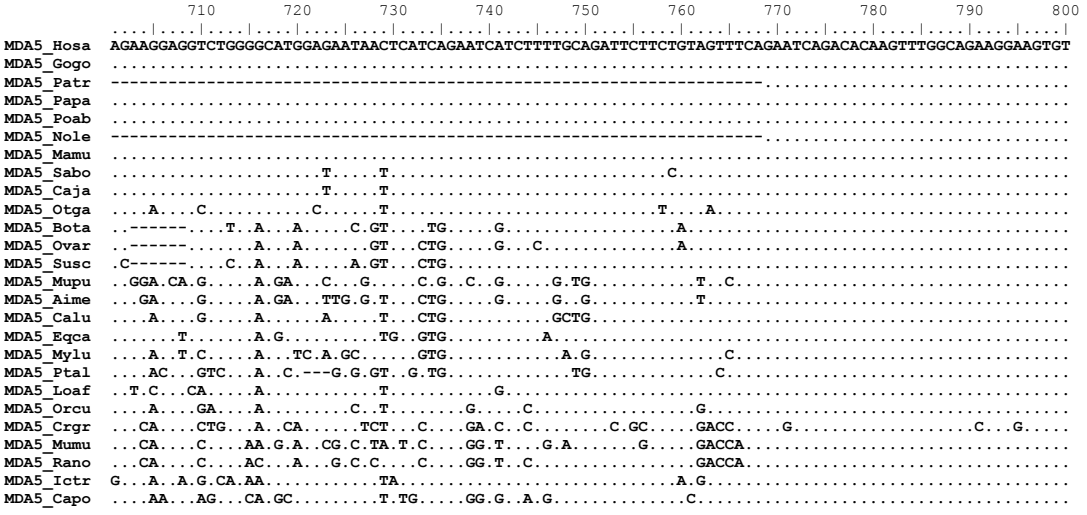


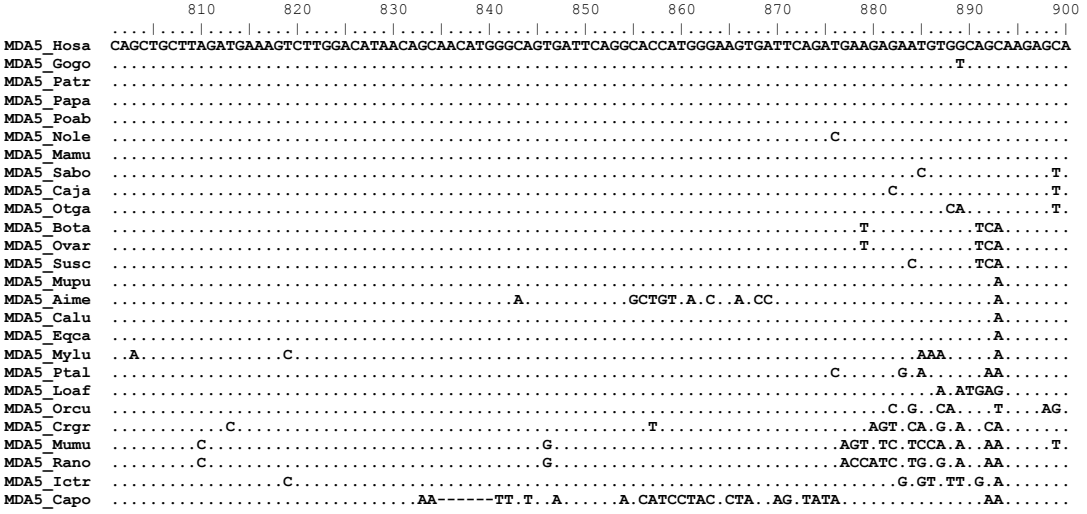

310 320 330 340 350 360 370 380 390 400
MDA5_Hosa GACTTGCCCTCTCCATCGTTTGAGAACGCTCATGATGAATATCTCCAACTGCTGAACCTCCTTCAGCCCACTCTGGTGGACAAGCTTCTAGTTAGAGACG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

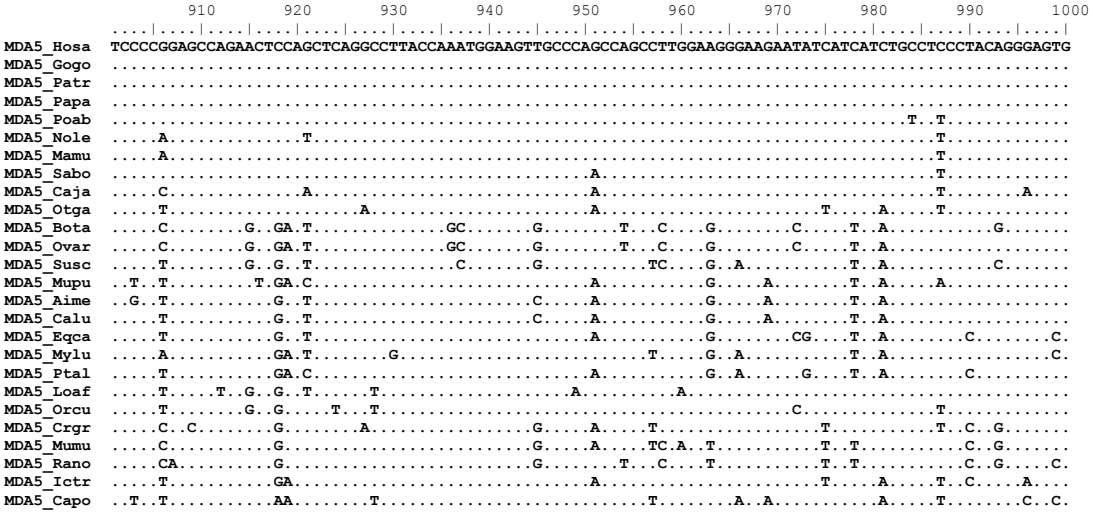

410 420 430 440 450 460 470 480 490 500
MDA5_Hosa TCTTGGATAAGTGCATGGAGGAGGAACTGTTGACAATTGAAGACAGAAACCGGATTGCTGCTGCAGAAAACAATGGAAATGAATCAGGTGTAAGAGAGCT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

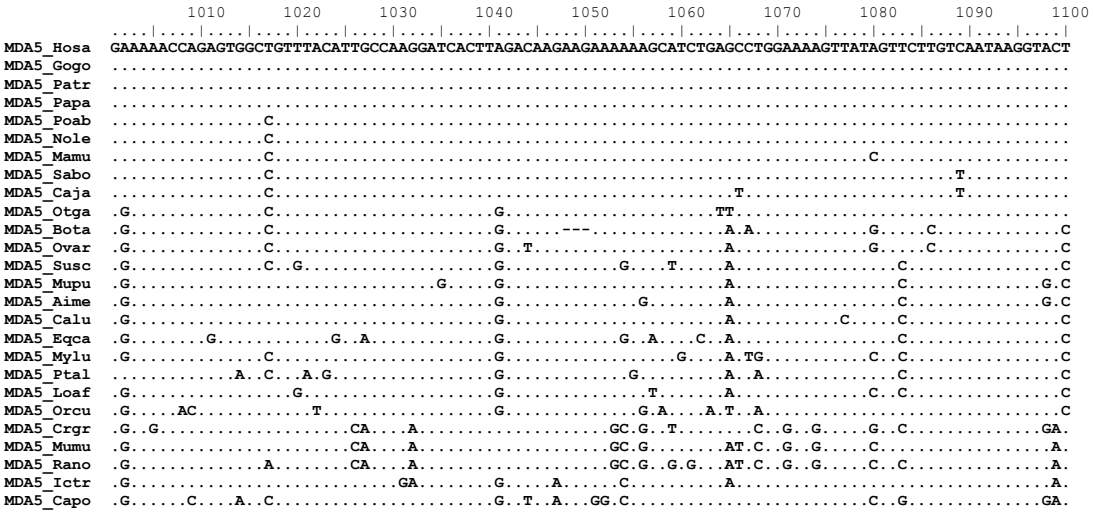

510 520 530 540 550 560 570 580 590 600
MDA5_Hosa ACTAAAAAGGATTGTGCAGAAAGAAAACTGGTTCTCTGCATTTCTGAATGTTCTTCGTCAAACAGGAAACAATGAACTTGTCCAAGAGTTAACAGGCTCT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

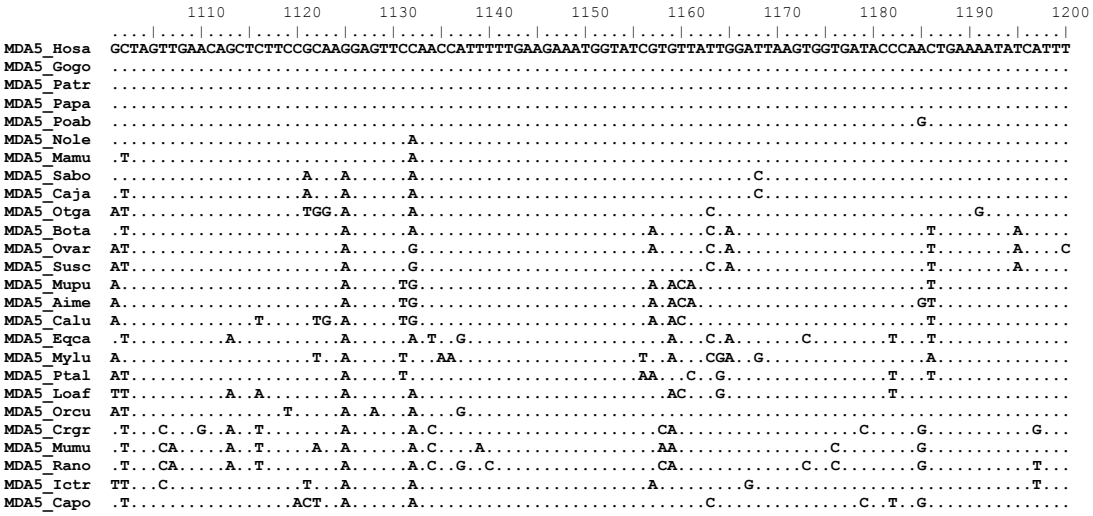

610 620 630 640 650 660 670 680 690 700
MDA5_Hosa GATTGCTCAGAAAGCAATGCAGAGATTGAGAATTTATCACAAGTTGATGGTCCTCAAGTGGAAGAGCAACTTCTTTCAACCACAGTTCAGCCAAATCTGG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

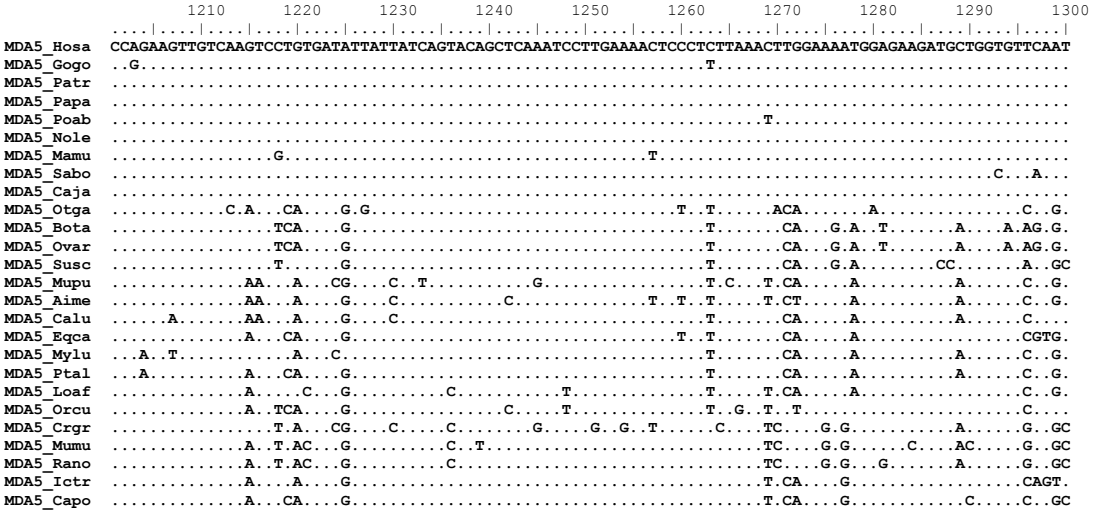

710 720 730 740 750 760 770 780 790 800
MDA5_Hosa AGAAGGAGGTCTGGGGCATGGAGAATAACTCATCAGAATCATCTTTTGCAGATTCTTCTGTAGTTTCAGAATCAGACACAAGTTTGGCAGAAGGAAGTGT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

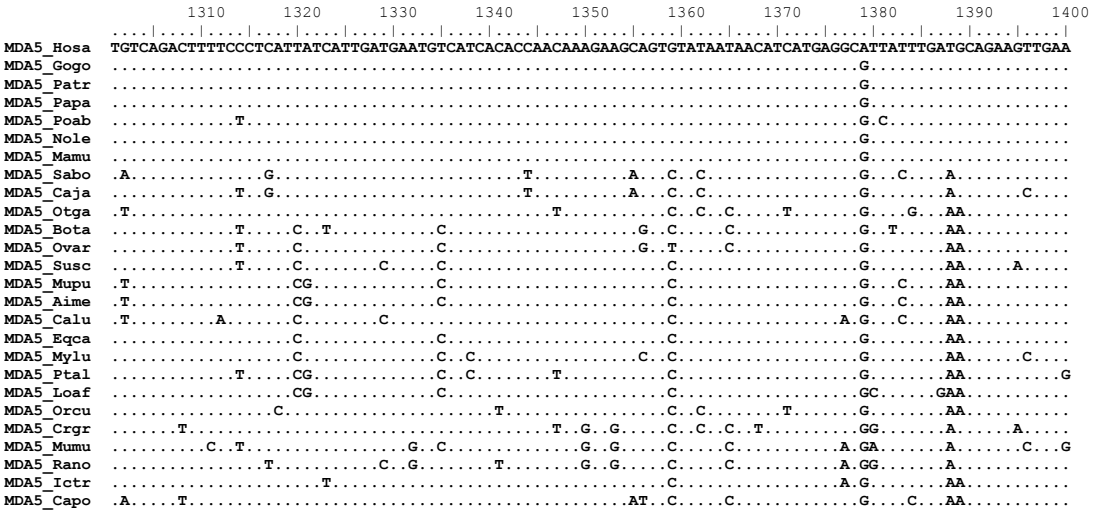

810 820 830 840 850 860 870 880 890 900
MDA5_Hosa CAGCTGCTTAGATGAAAGTCTTGGACATAACAGCAACATGGGCAGTGATTCAGGCACCATGGGAAGTGATTCAGATGAAGAGAATGTGGCAGCAAGAGCA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

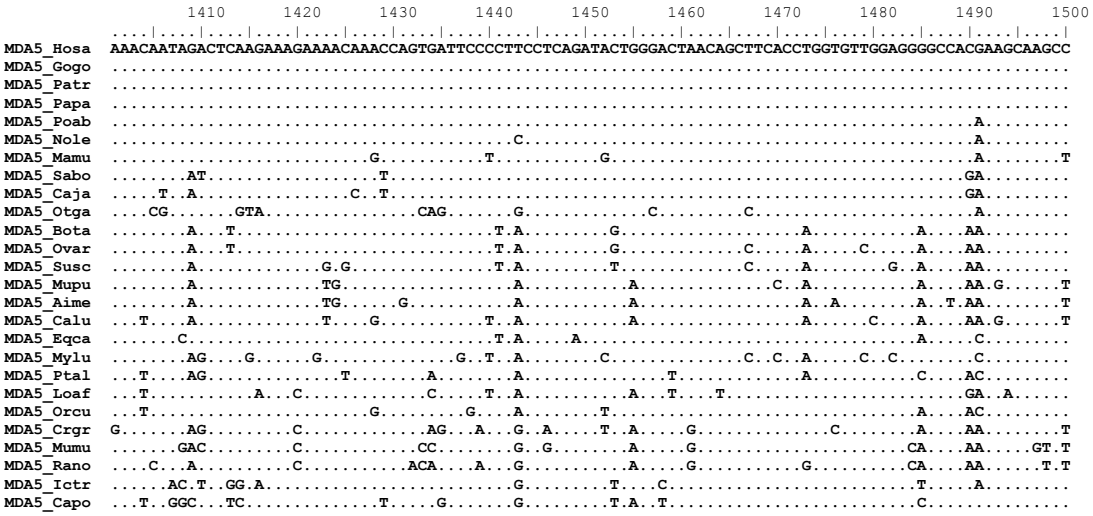

910 920 930 940 950 960 970 980 990 1000
MDA5_Hosa TCCCCGGAGCCAGAACTCCAGCTCAGGCCTTACCAAATGGAAGTTGCCCAGCCAGCCTTGGAAGGGAAGAATATCATCATCTGCCTCCCTACAGGGAGTG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

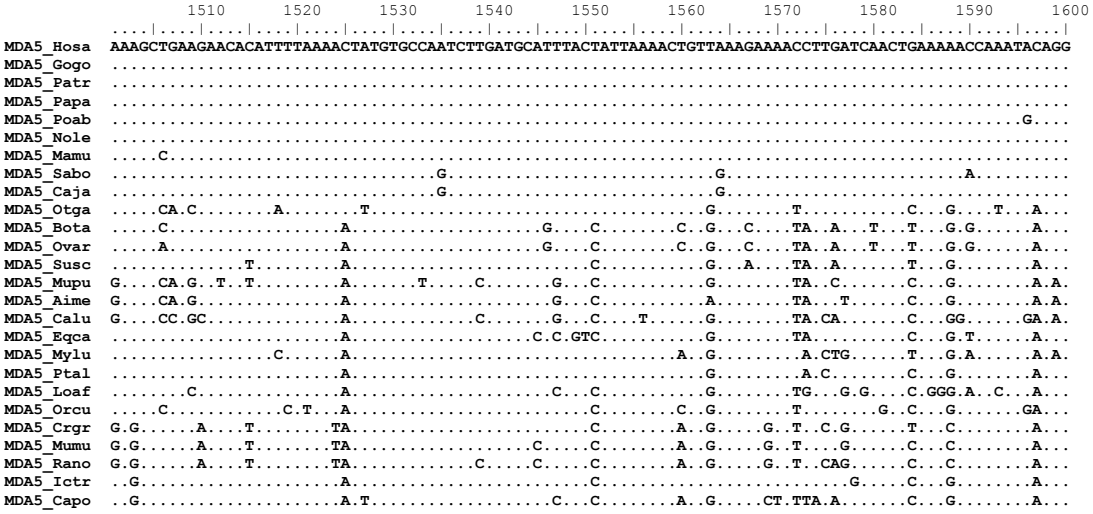

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
MDA5_Hosa GAAAAACCAGAGTGGCTGTTTACATTGCCAAGGATCACTTAGACAAGAAGAAAAAAGCATCTGAGCCTGGAAAAGTTATAGTTCTTGTCAATAAGGTACT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

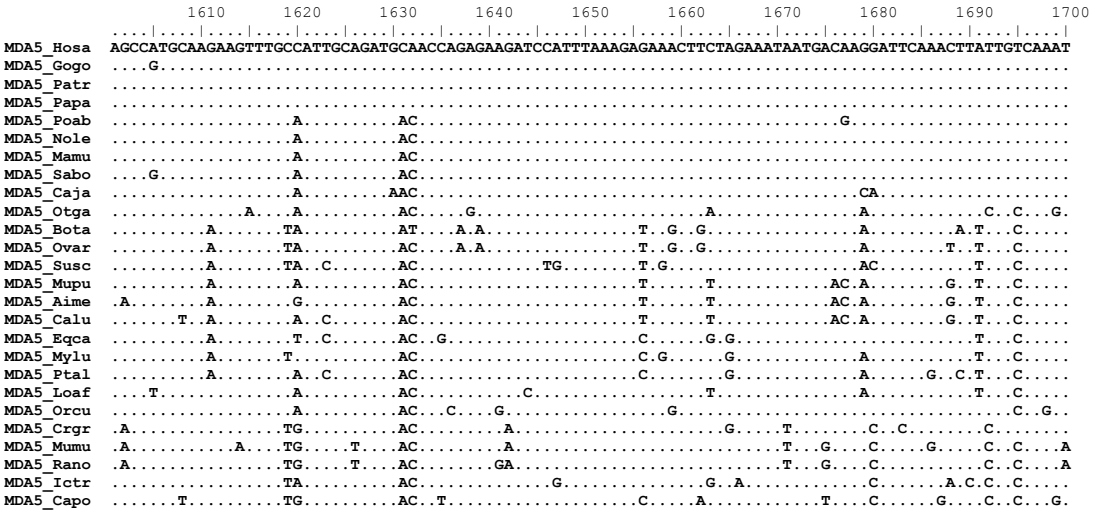

1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
MDA5_Hosa GCTAGTTGAACAGCTCTTCCGCAAGGAGTTCCAACCATTTTTGAAGAAATGGTATCGTGTTATTGGATTAAGTGGTGATACCCAACTGAAAATATCATTT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

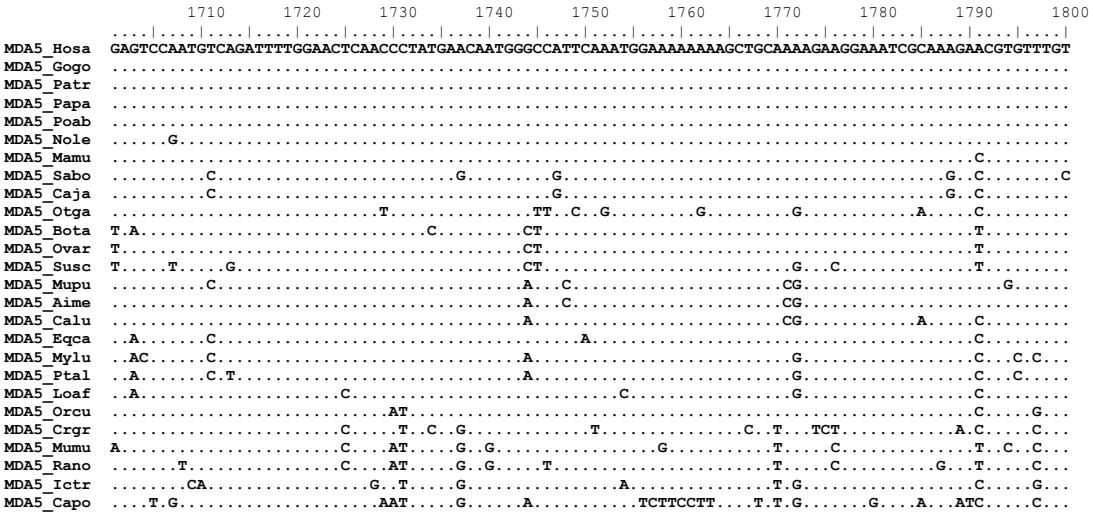

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
MDA5_Hosa CCAGAAGTTGTCAAGTCCTGTGATATTATTATCAGTACAGCTCAAATCCTTGAAAACTCCCTCTTAAACTTGGAAAATGGAGAAGATGCTGGTGTTCAAT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

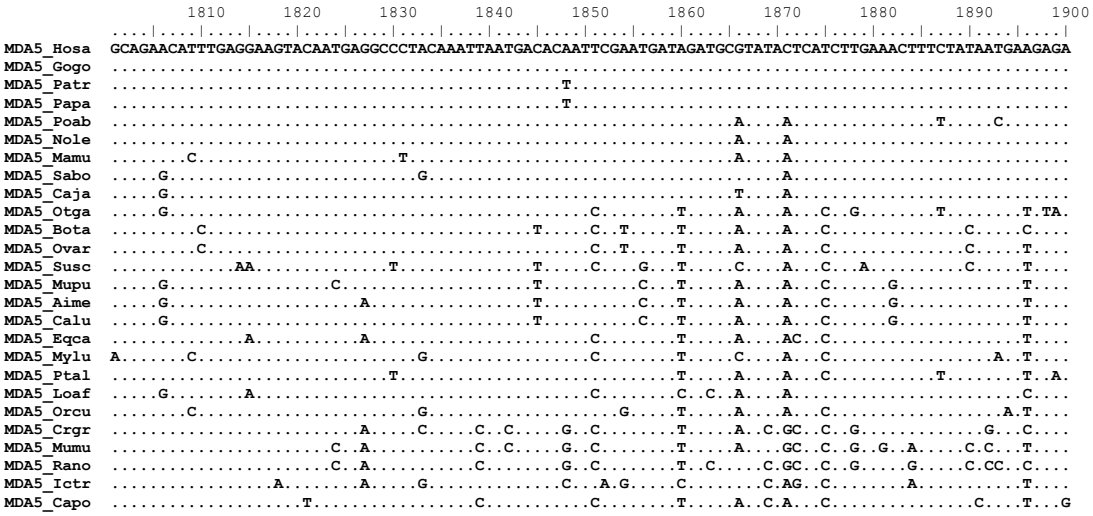

1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
MDA5_Hosa TGTCAGACTTTTCCCTCATTATCATTGATGAATGTCATCACACCAACAAAGAAGCAGTGTATAATAACATCATGAGGCATTATTTGATGCAGAAGTTGAA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

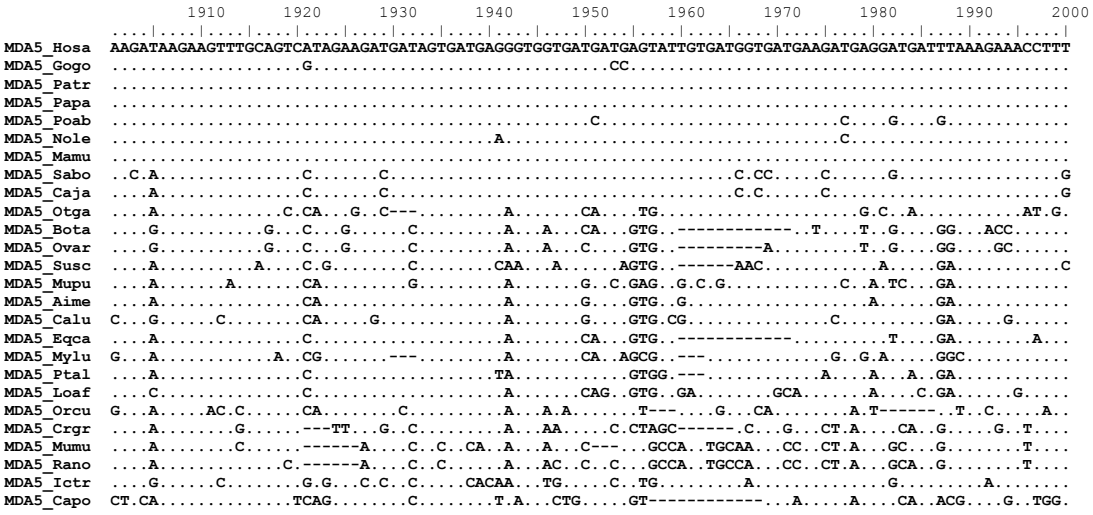

1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
MDA5_Hosa AAACAATAGACTCAAGAAAGAAAACAAACCAGTGATTCCCCTTCCTCAGATACTGGGACTAACAGCTTCACCTGGTGTTGGAGGGGCCACGAAGCAAGCC
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

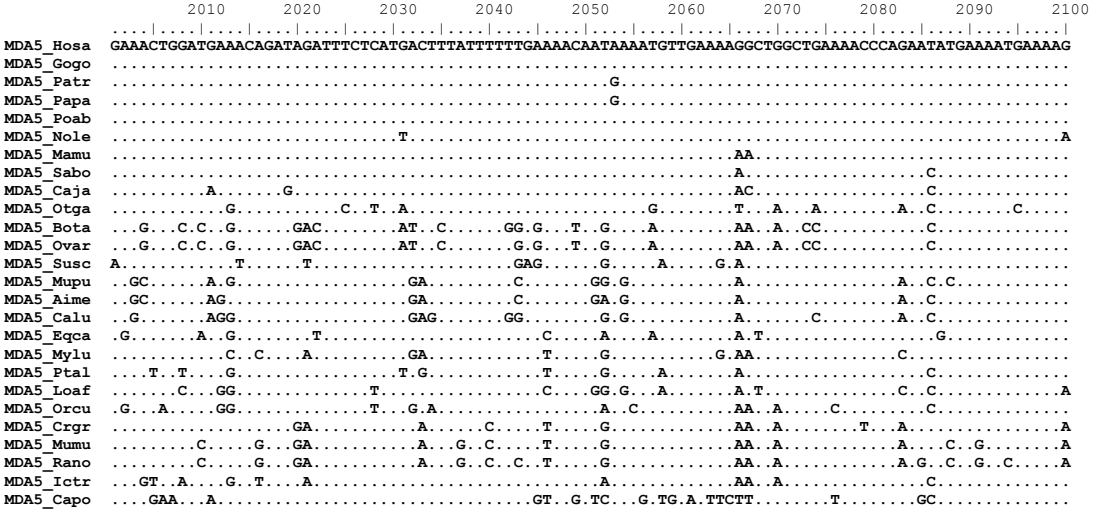

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
MDA5_Hosa AAAGCTGAAGAACACATTTTAAAACTATGTGCCAATCTTGATGCATTTACTATTAAAACTGTTAAAGAAAACCTTGATCAACTGAAAAACCAAATACAGG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

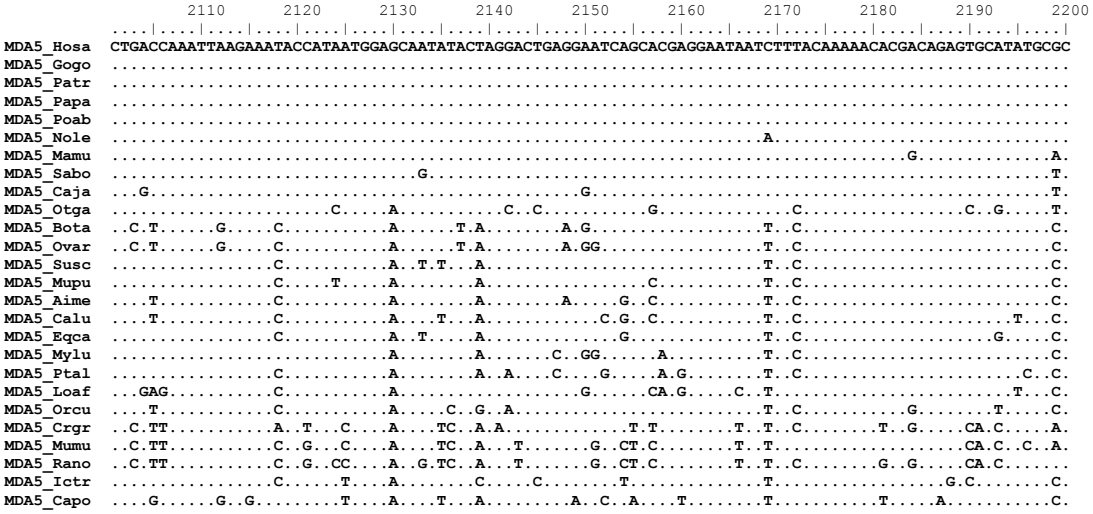

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
MDA5_Hosa AGCCATGCAAGAAGTTTGCCATTGCAGATGCAACCAGAGAAGATCCATTTAAAGAGAAACTTCTAGAAATAATGACAAGGATTCAAACTTATTGTCAAAT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

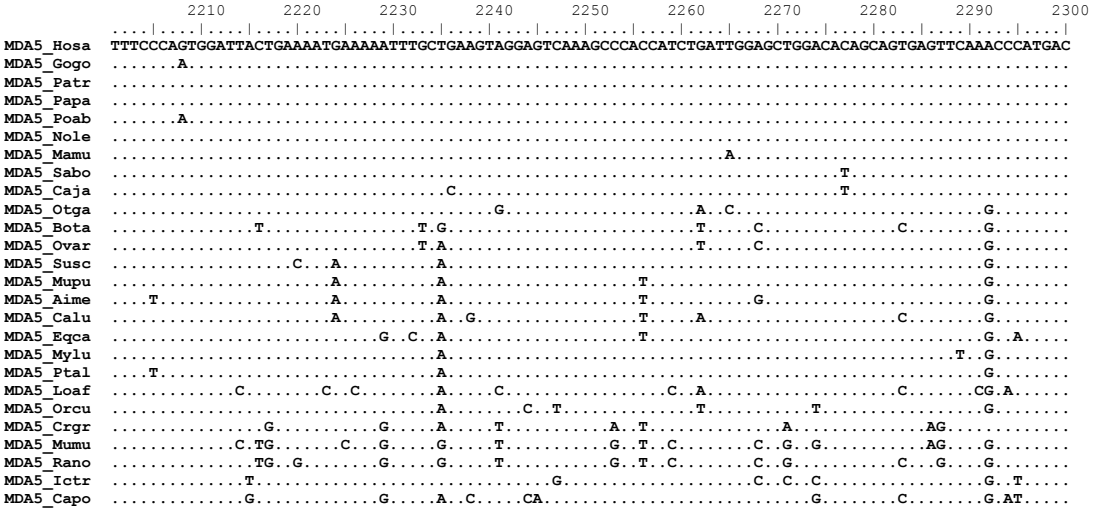

1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
MDA5_Hosa GAGTCCAATGTCAGATTTTGGAACTCAACCCTATGAACAATGGGCCATTCAAATGGAAAAAAAAGCTGCAAAAGAAGGAAATCGCAAAGAACGTGTTTGT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

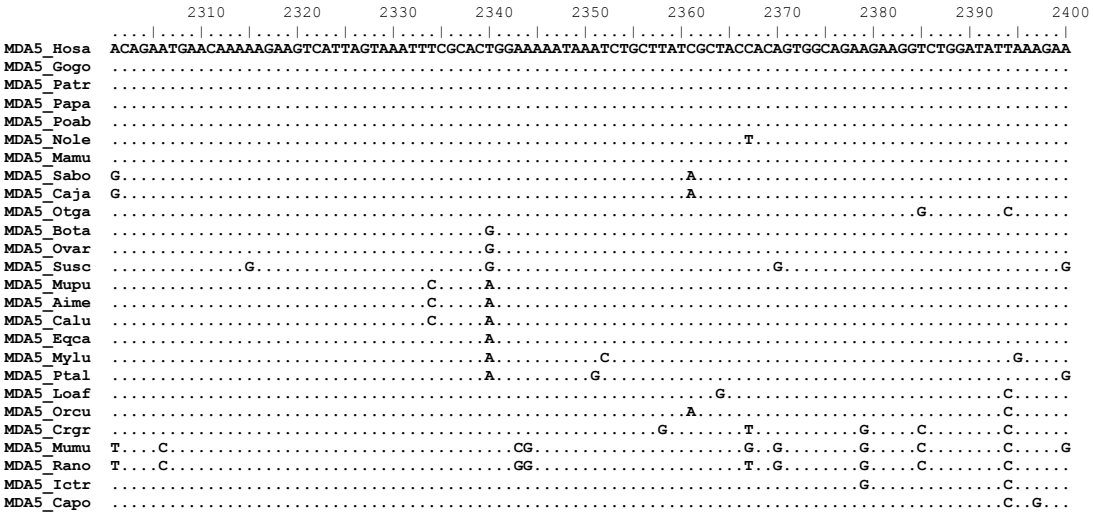

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
MDA5_Hosa GCAGAACATTTGAGGAAGTACAATGAGGCCCTACAAATTAATGACACAATTCGAATGATAGATGCGTATACTCATCTTGAAACTTTCTATAATGAAGAGA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

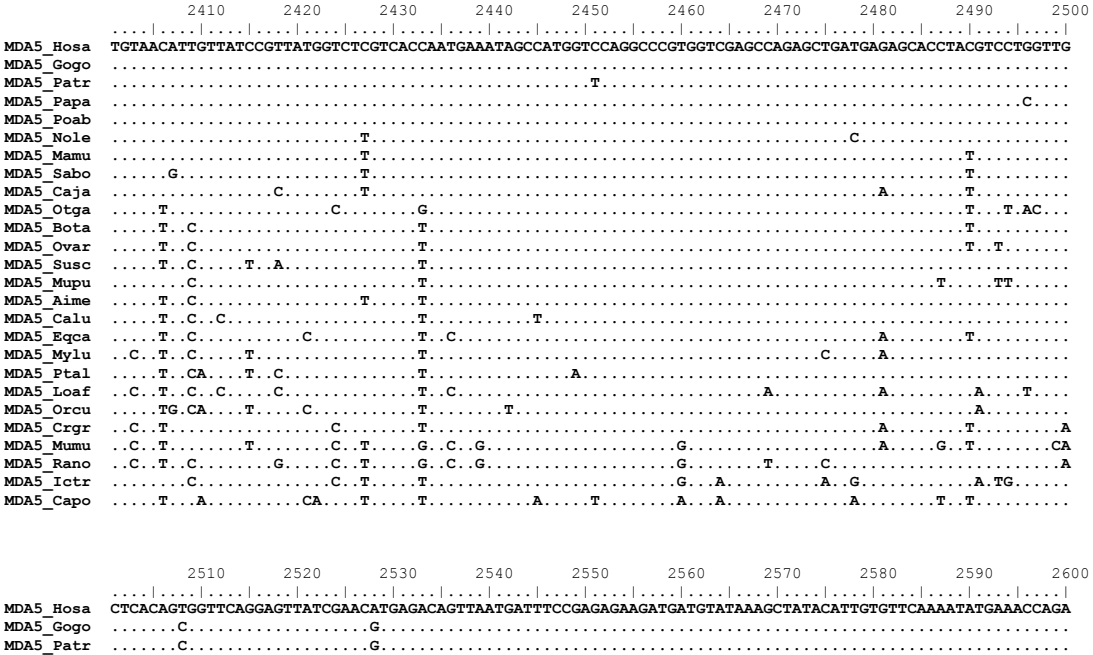

1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
MDA5_Hosa AAGATAAGAAGTTTGCAGTCATAGAAGATGATAGTGATGAGGGTGGTGATGATGAGTATTGTGATGGTGATGAAGATGAGGATGATTTAAAGAAACCTTT
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

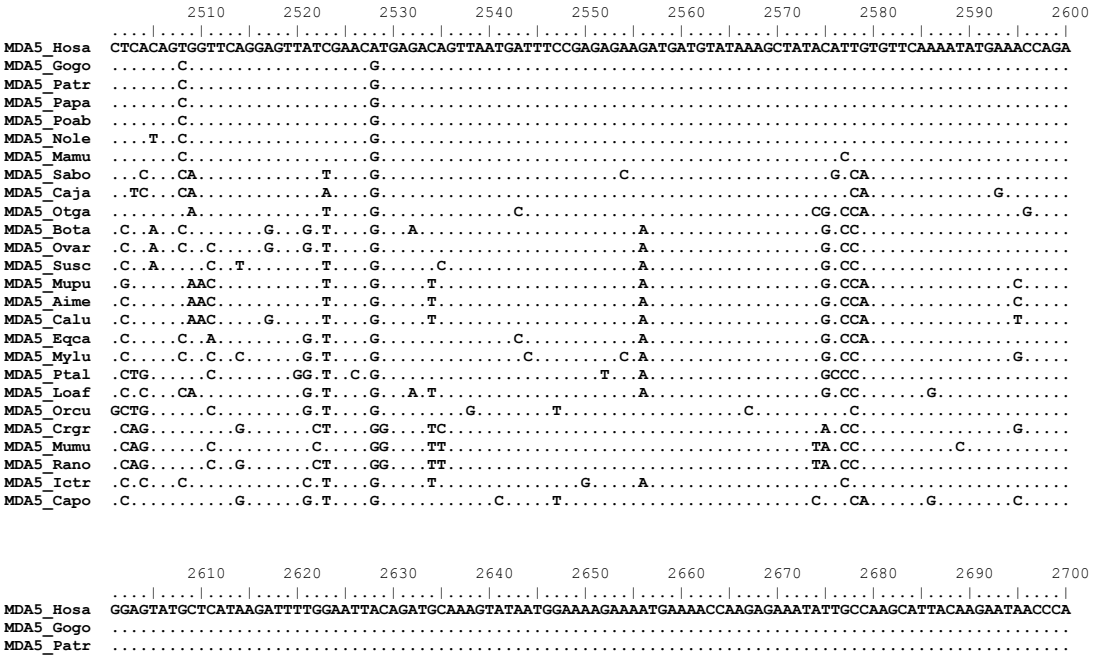

2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
MDA5_Hosa GAAACTGGATGAAACAGATAGATTTCTCATGACTTTATTTTTTGAAAACAATAAAATGTTGAAAAGGCTGGCTGAAAACCCAGAATATGAAAATGAAAAG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

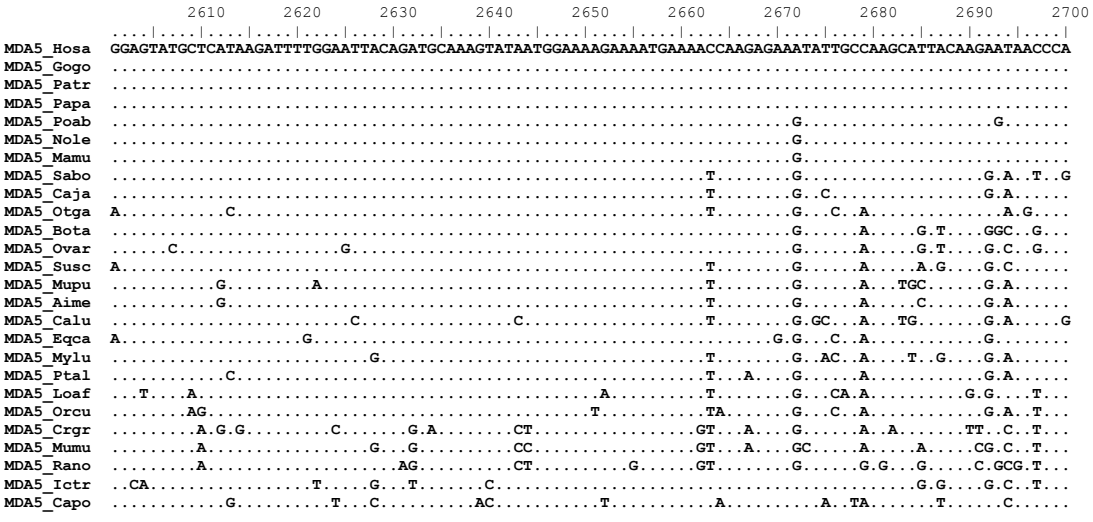

2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
MDA5_Hosa CTGACCAAATTAAGAAATACCATAATGGAGCAATATACTAGGACTGAGGAATCAGCACGAGGAATAATCTTTACAAAAACACGACAGAGTGCATATGCGC
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

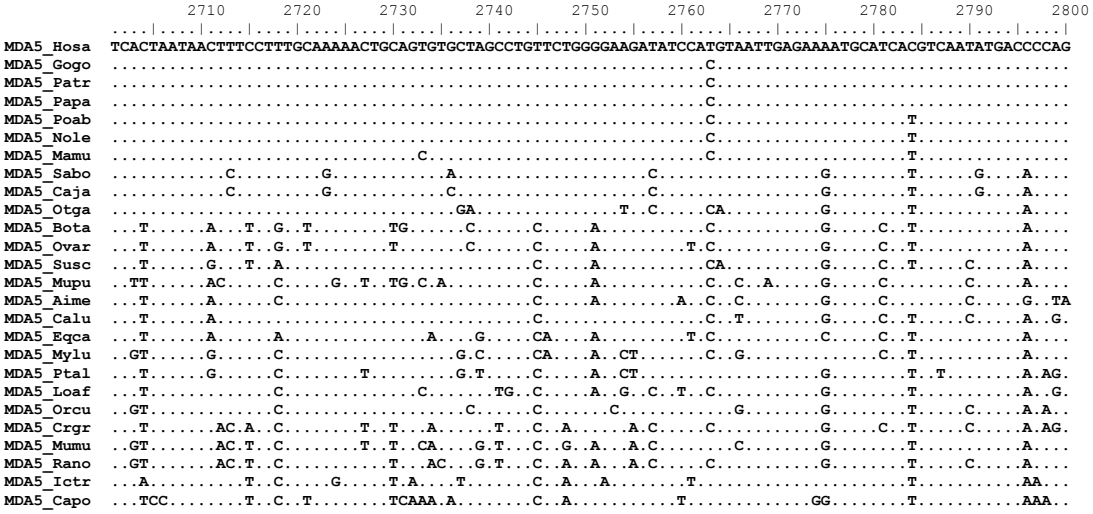

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
MDA5_Hosa TTTCCCAGTGGATTACTGAAAATGAAAAATTTGCTGAAGTAGGAGTCAAAGCCCACCATCTGATTGGAGCTGGACACAGCAGTGAGTTCAAACCCATGAC
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

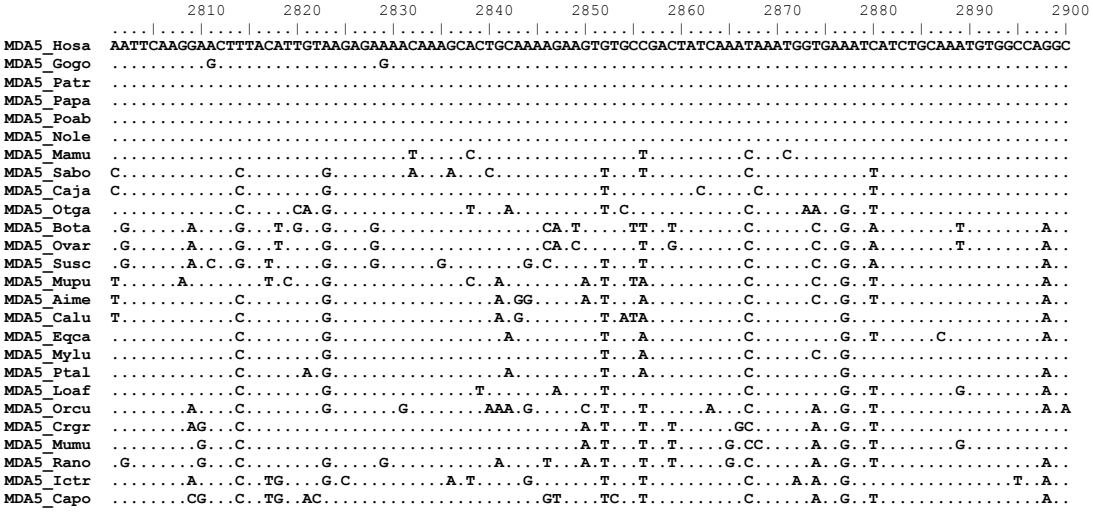

2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
MDA5_Hosa ACAGAATGAACAAAAAGAAGTCATTAGTAAATTTCGCACTGGAAAAATAAATCTGCTTATCGCTACCACAGTGGCAGAAGAAGGTCTGGATATTAAAGAA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

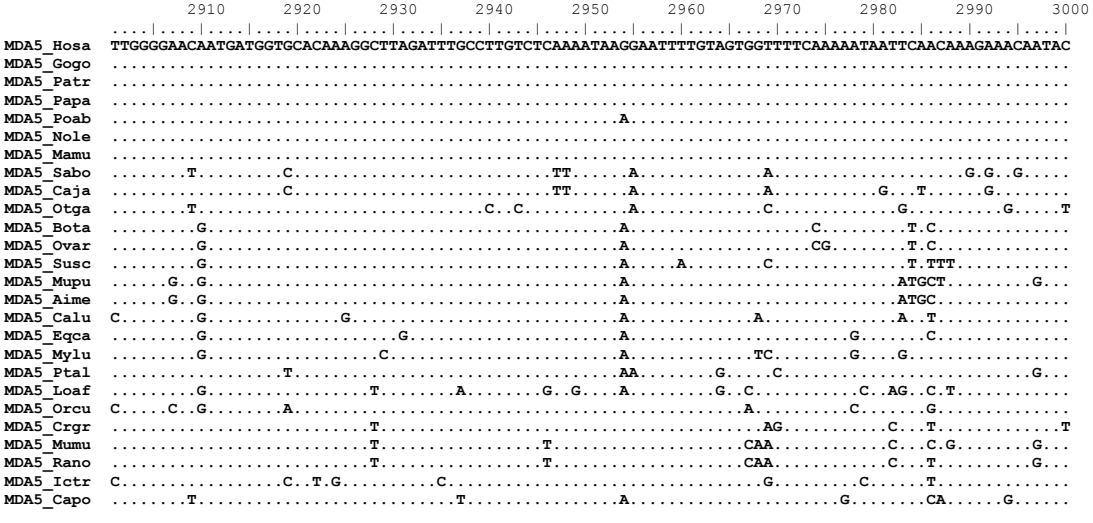

2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
MDA5_Hosa TGTAACATTGTTATCCGTTATGGTCTCGTCACCAATGAAATAGCCATGGTCCAGGCCCGTGGTCGAGCCAGAGCTGATGAGAGCACCTACGTCCTGGTTG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

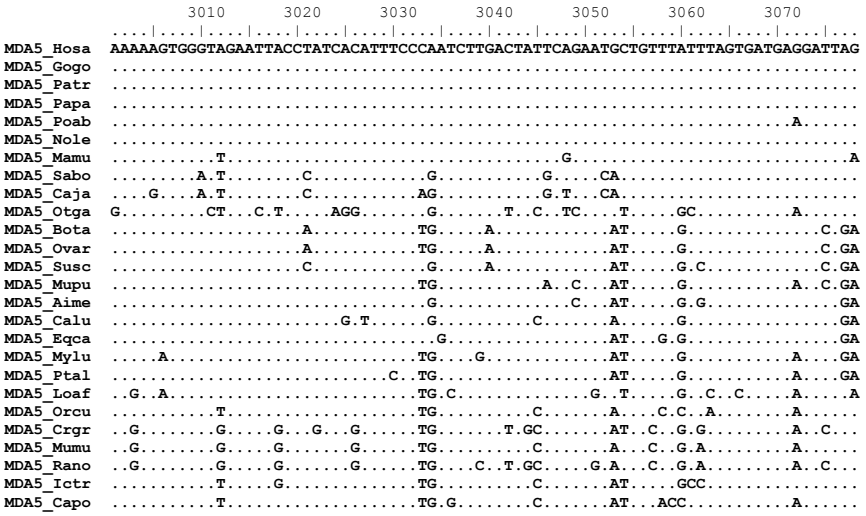

2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
MDA5_Hosa CTCACAGTGGTTCAGGAGTTATCGAACATGAGACAGTTAATGATTTCCGAGAGAAGATGATGTATAAAGCTATACATTGTGTTCAAAATATGAAACCAGA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo


2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
MDA5_Hosa GGAGTATGCTCATAAGATTTTGGAATTACAGATGCAAAGTATAATGGAAAAGAAAATGAAAACCAAGAGAAATATTGCCAAGCATTACAAGAATAACCCA
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

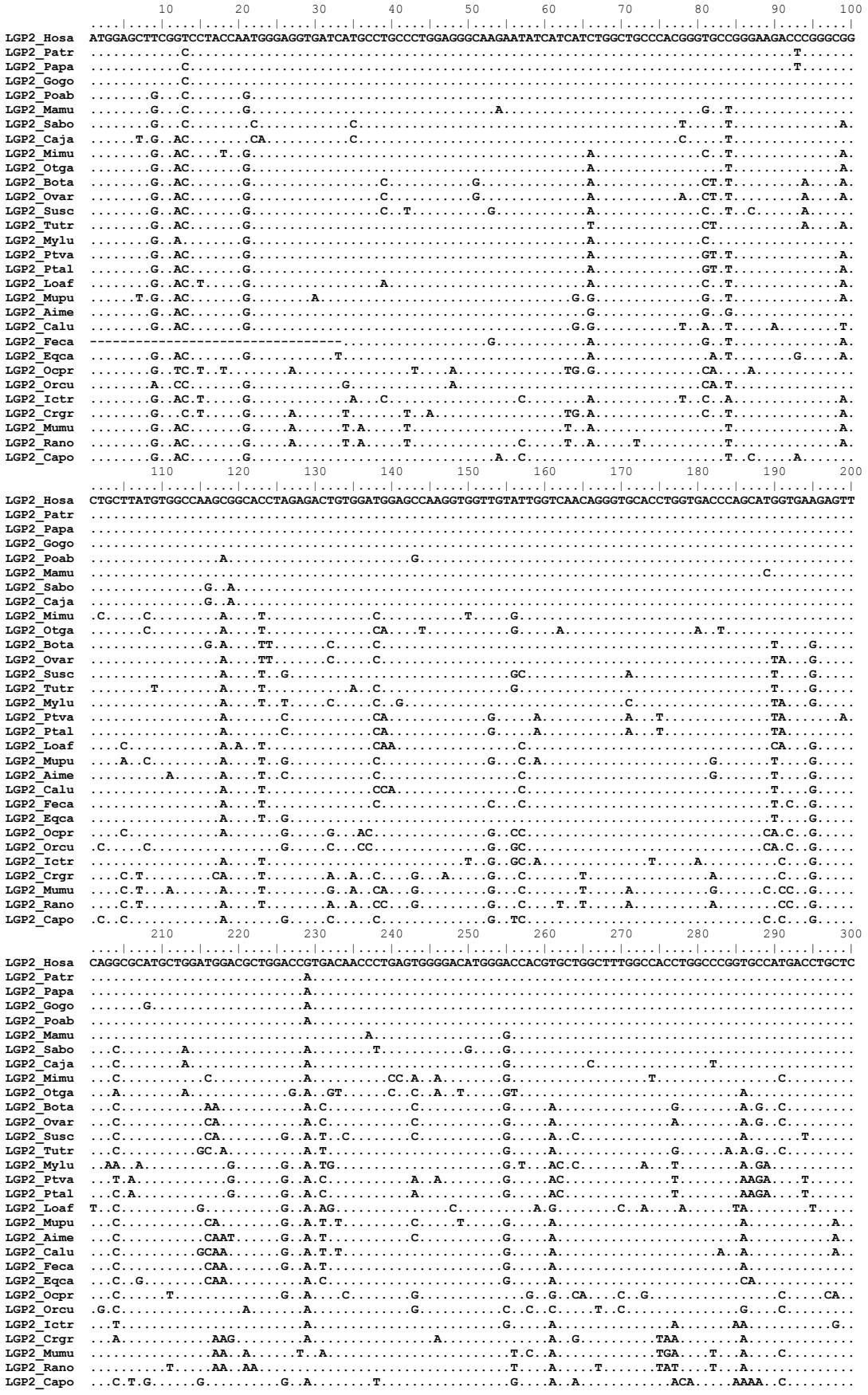

2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
MDA5_Hosa TCACTAATAACTTTCCTTTGCAAAAACTGCAGTGTGCTAGCCTGTTCTGGGGAAGATATCCATGTAATTGAGAAAATGCATCACGTCAATATGACCCCAG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

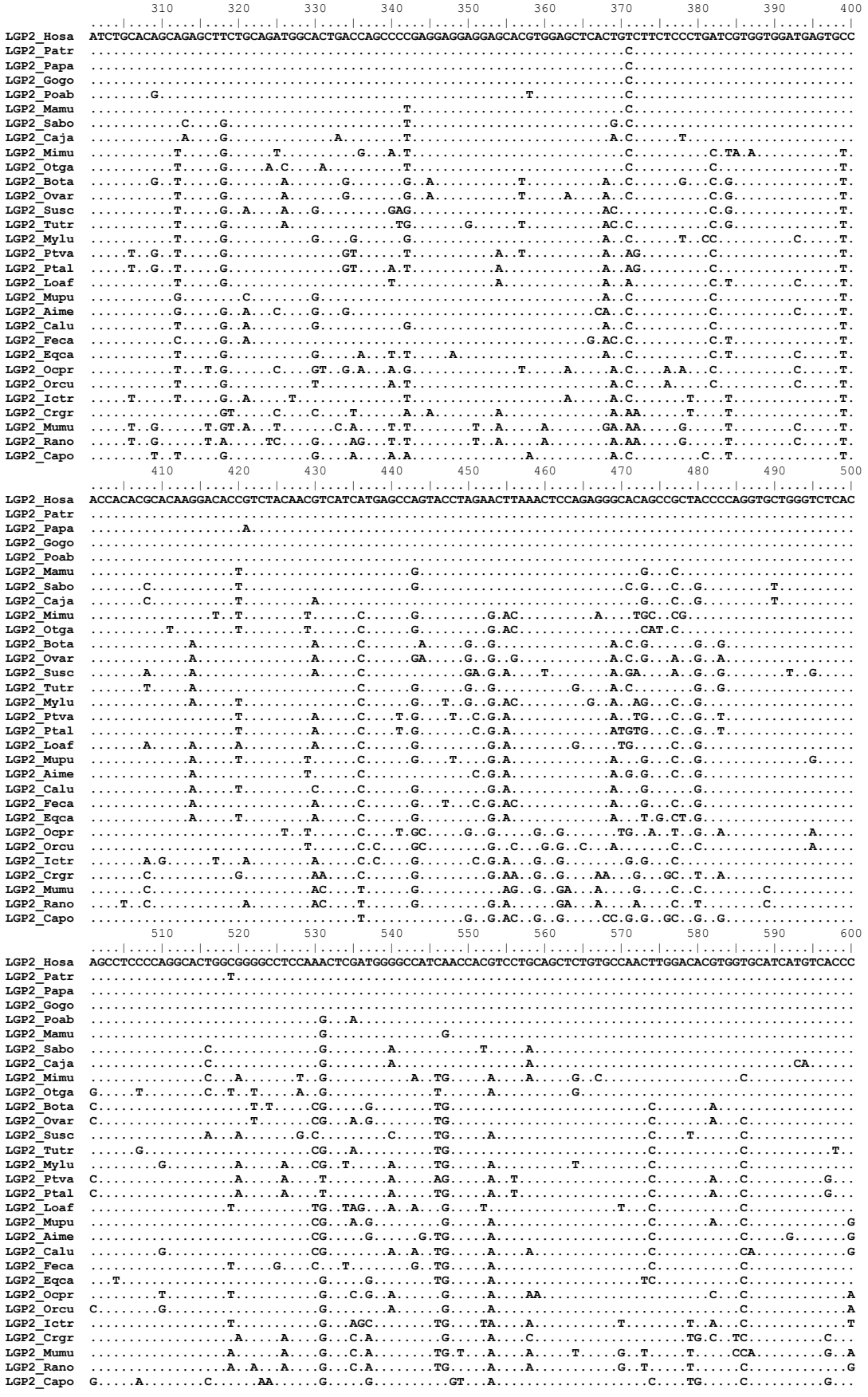

2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
MDA5_Hosa AATTCAAGGAACTTTACATTGTAAGAGAAAACAAAGCACTGCAAAAGAAGTGTGCCGACTATCAAATAAATGGTGAAATCATCTGCAAATGTGGCCAGGC
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

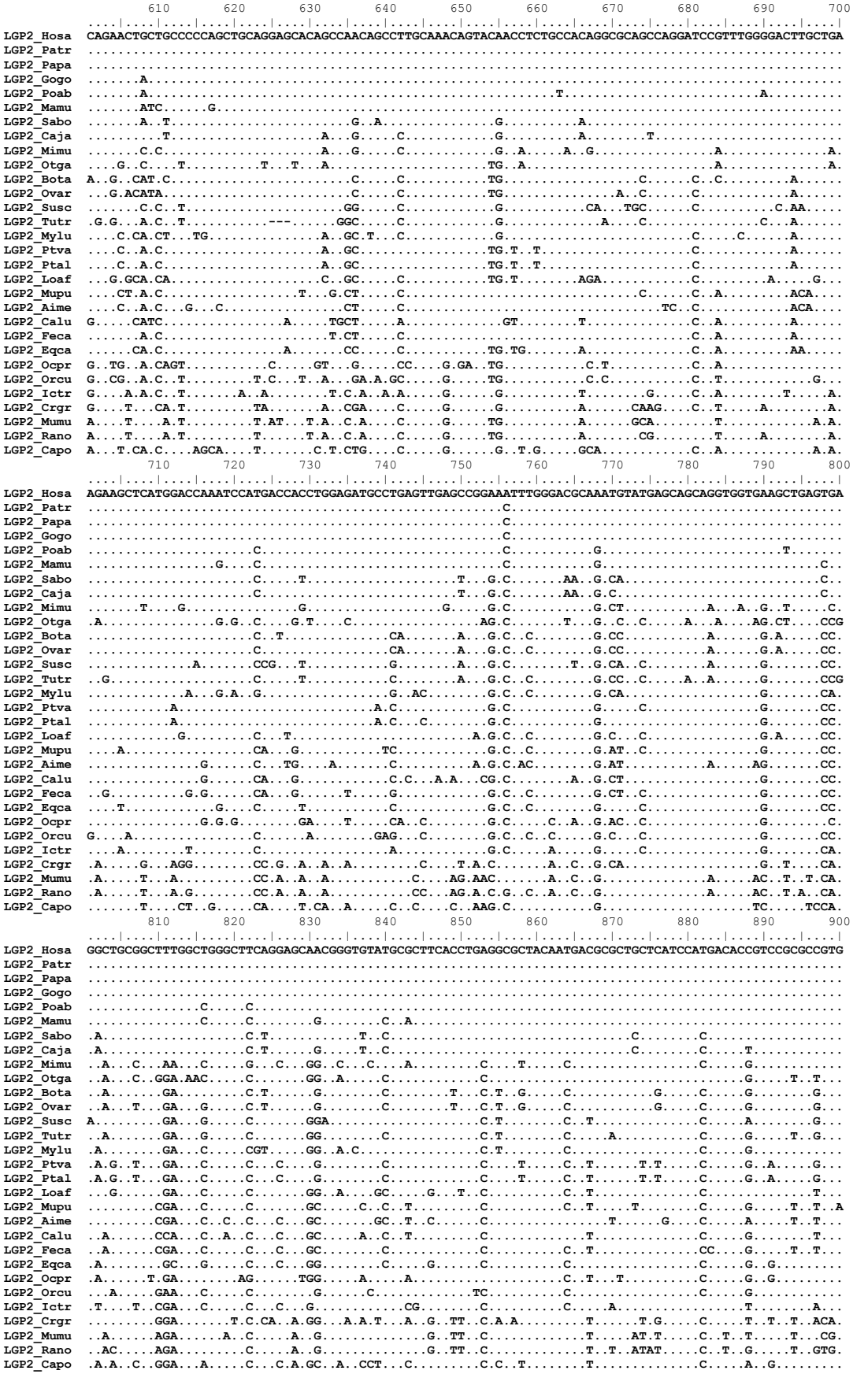

2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
MDA5_Hosa TTGGGGAACAATGATGGTGCACAAAGGCTTAGATTTGCCTTGTCTCAAAATAAGGAATTTTGTAGTGGTTTTCAAAAATAATTCAACAAAGAAACAATAC
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

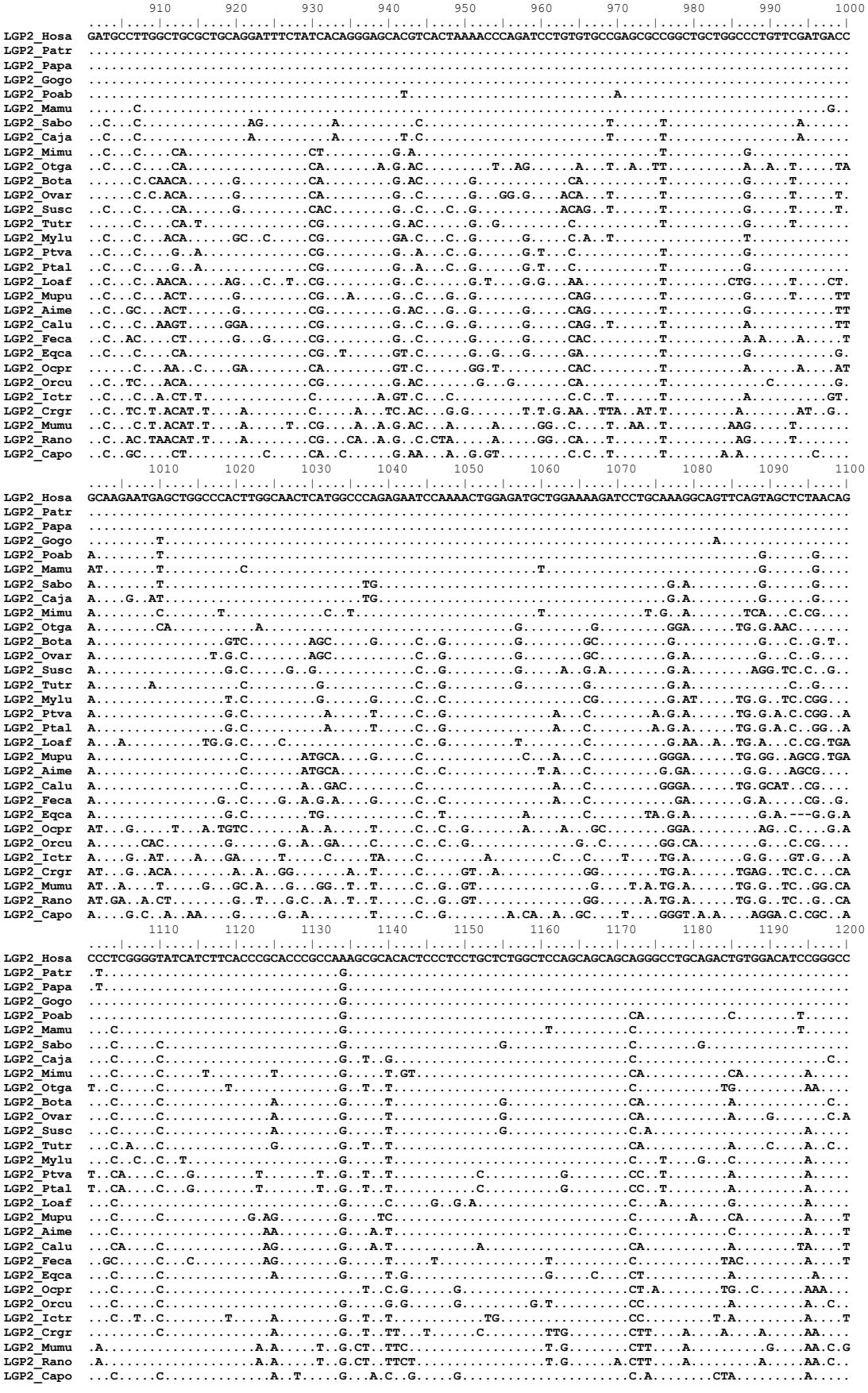

3010 3020 3030 3040 3050 3060 3070
MDA5_Hosa AAAAAGTGGGTAGAATTACCTATCACATTTCCCAATCTTGACTATTCAGAATGCTGTTTATTTAGTGATGAGGATTAG
MDA5_Gogo
MDA5_Patr
MDA5_Papa
MDA5_Poab
MDA5_Nole
MDA5_Mamu
MDA5_Sabo
MDA5_Caja
MDA5_Otga
MDA5_Bota
MDA5_Ovar
MDA5_Susc
MDA5_Mupu
MDA5_Aime
MDA5_Calu
MDA5_Eqca
MDA5_Mylu
MDA5_Ptal
MDA5_Loaf
MDA5_Orcu
MDA5_Crgr
MDA5_Mumu
MDA5_Rano
MDA5_Ictr
MDA5_Capo

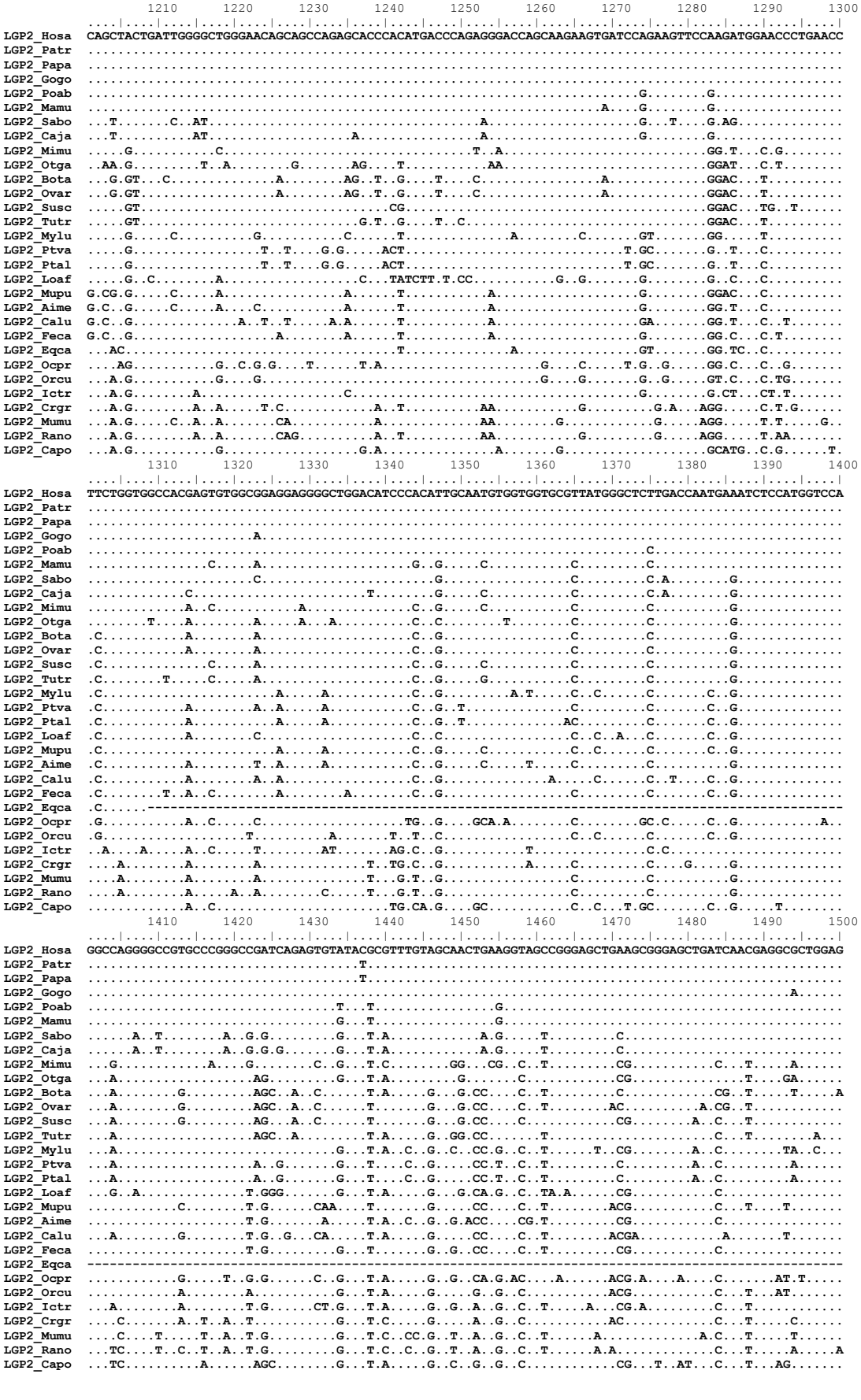
Supplementary Figure S9

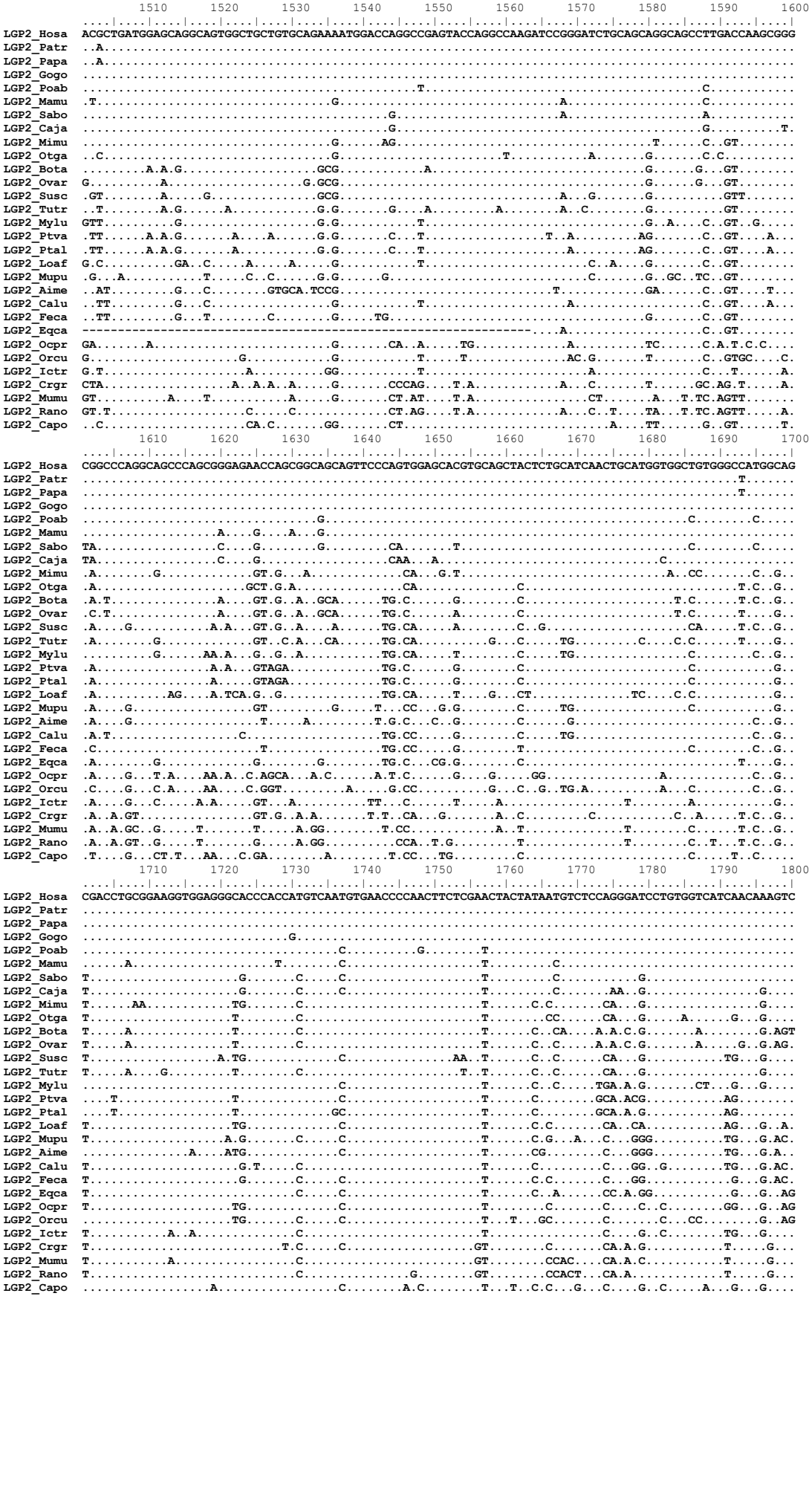


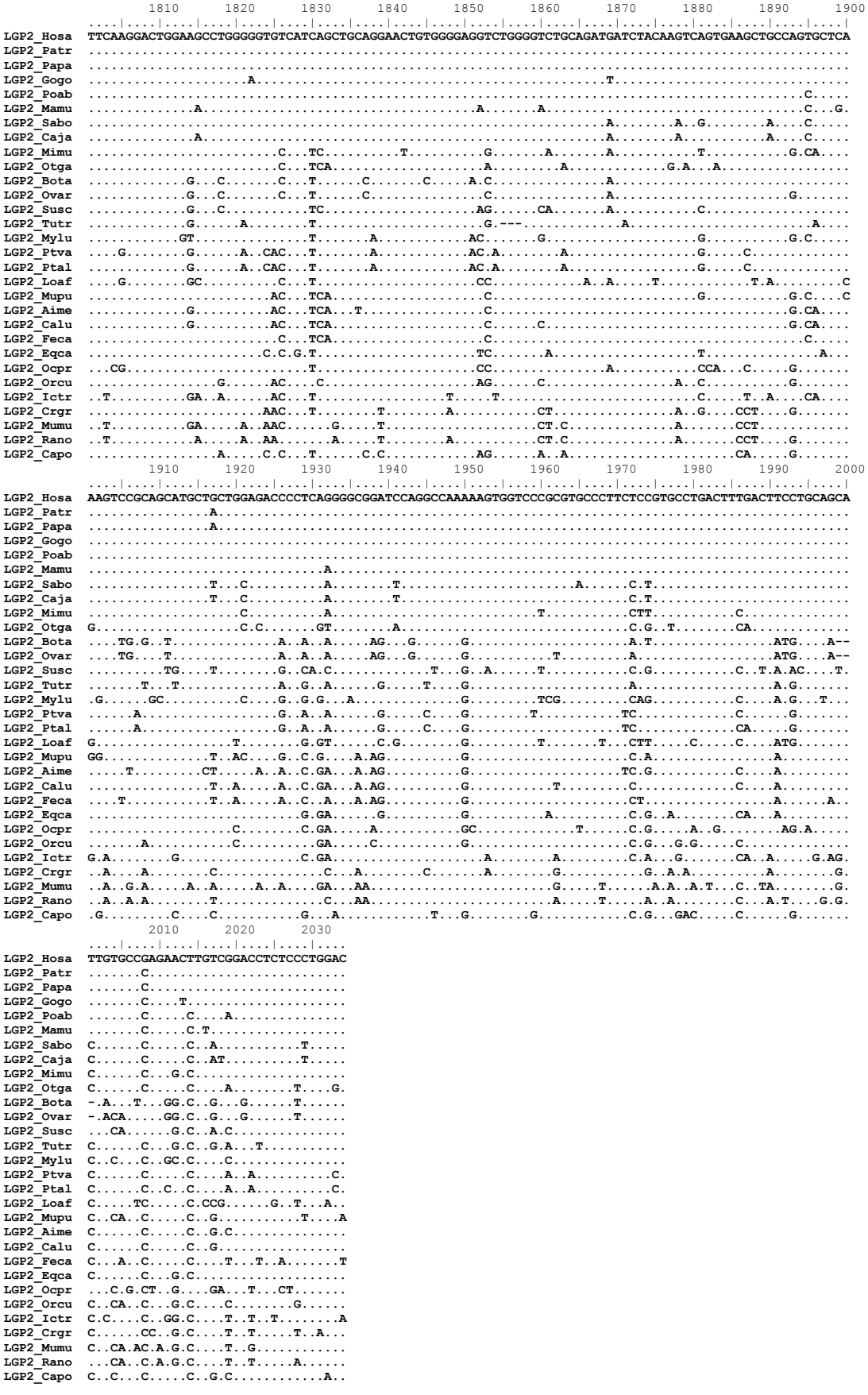

310 320 330 340 350 360 370 380 390 400
LGP2_Hosa ATCTGCACAGCAGAGCTTCTGCAGATGGCACTGACCAGCCCCGAGGAGGAGGAGCACGTGGAGCTCACTGTCTTCTCCCTGATCGTGGTGGATGAGTGCC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
410 420 430 440 450 460 470 480 490 500
LGP2_Hosa ACCACACGCACAAGGACACCGTCTACAACGTCATCATGAGCCAGTACCTAGAACTTAAACTCCAGAGGGCACAGCCGCTACCCCAGGTGCTGGGTCTCAC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
510 520 530 540 550 560 570 580 590 600
LGP2_Hosa AGCCTCCCCAGGCACTGGCGGGGCCTCCAAACTCGATGGGGCCATCAACCACGTCCTGCAGCTCTGTGCCAACTTGGACACGTGGTGCATCATGTCACCC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo


610 620 630 640 650 660 670 680 690 700
LGP2_Hosa CAGAACTGCTGCCCCCAGCTGCAGGAGCACAGCCAACAGCCTTGCAAACAGTACAACCTCTGCCACAGGCGCAGCCAGGATCCGTTTGGGGACTTGCTGA
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
710 720 730 740 750 760 770 780 790 800
LGP2_Hosa AGAAGCTCATGGACCAAATCCATGACCACCTGGAGATGCCTGAGTTGAGCCGGAAATTTGGGACGCAAATGTATGAGCAGCAGGTGGTGAAGCTGAGTGA
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
810 820 830 840 850 860 870 880 890 900
LGP2_Hosa GGCTGCGGCTTTGGCTGGGCTTCAGGAGCAACGGGTGTATGCGCTTCACCTGAGGCGCTACAATGACGCGCTGCTCATCCATGACACCGTCCGCGCCGTG
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo


910 920 930 940 950 960 970 980 990 1000
LGP2_Hosa GATGCCTTGGCTGCGCTGCAGGATTTCTATCACAGGGAGCACGTCACTAAAACCCAGATCCTGTGTGCCGAGCGCCGGCTGCTGGCCCTGTTCGATGACC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
LGP2_Hosa GCAAGAATGAGCTGGCCCACTTGGCAACTCATGGCCCAGAGAATCCAAAACTGGAGATGCTGGAAAAGATCCTGCAAAGGCAGTTCAGTAGCTCTAACAG
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
LGP2_Hosa CCCTCGGGGTATCATCTTCACCCGCACCCGCCAAAGCGCACACTCCCTCCTGCTCTGGCTCCAGCAGCAGCAGGGCCTGCAGACTGTGGACATCCGGGCC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo


1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
LGP2_Hosa CAGCTACTGATTGGGGCTGGGAACAGCAGCCAGAGCACCCACATGACCCAGAGGGACCAGCAAGAAGTGATCCAGAAGTTCCAAGATGGAACCCTGAACC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
LGP2_Hosa TTCTGGTGGCCACGAGTGTGGCGGAGGAGGGGCTGGACATCCCACATTGCAATGTGGTGGTGCGTTATGGGCTCTTGACCAATGAAATCTCCATGGTCCA
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
LGP2_Hosa GGCCAGGGGCCGTGCCCGGGCCGATCAGAGTGTATACGCGTTTGTAGCAACTGAAGGTAGCCGGGAGCTGAAGCGGGAGCTGATCAACGAGGCGCTGGAG
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo


1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
LGP2_Hosa ACGCTGATGGAGCAGGCAGTGGCTGCTGTGCAGAAAATGGACCAGGCCGAGTACCAGGCCAAGATCCGGGATCTGCAGCAGGCAGCCTTGACCAAGCGGG
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
LGP2_Hosa CGGCCCAGGCAGCCCAGCGGGAGAACCAGCGGCAGCAGTTCCCAGTGGAGCACGTGCAGCTACTCTGCATCAACTGCATGGTGGCTGTGGGCCATGGCAG
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
LGP2_Hosa CGACCTGCGGAAGGTGGAGGGCACCCACCATGTCAATGTGAACCCCAACTTCTCGAACTACTATAATGTCTCCAGGGATCCTGTGGTCATCAACAAAGTC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo


1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
LGP2_Hosa TTCAAGGACTGGAAGCCTGGGGGTGTCATCAGCTGCAGGAACTGTGGGGAGGTCTGGGGTCTGCAGATGATCTACAAGTCAGTGAAGCTGCCAGTGCTCA
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
LGP2_Hosa AAGTCCGCAGCATGCTGCTGGAGACCCCTCAGGGGCGGATCCAGGCCAAAAAGTGGTCCCGCGTGCCCTTCTCCGTGCCTGACTTTGACTTCCTGCAGCA
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo
2010 2020 2030
LGP2_Hosa TTGTGCCGAGAACTTGTCGGACCTCTCCCTGGAC
LGP2_Patr
LGP2_Papa
LGP2_Gogo
LGP2_Poab
LGP2_Mamu
LGP2_Sabo
LGP2_Caja
LGP2_Mimu
LGP2_Otga
LGP2_Bota
LGP2_Ovar
LGP2_Susc
LGP2_Tutr
LGP2_Mylu
LGP2_Ptva
LGP2_Ptal
LGP2_Loaf
LGP2_Mupu
LGP2_Aime
LGP2_Calu
LGP2_Feca
LGP2_Eqca
LGP2_Ocpr
LGP2_Orcu
LGP2_Ictr
LGP2_Crgr
LGP2_Mumu
LGP2_Rano
LGP2_Capo

Supplementary Table S2 - *RIG-I*, *MDA5* and *LGP2* likelihood ratio test (LRT) for PARRIS analysis from HyPhy software

PARRIS analysis	lnL_{null}[a]	$lnL_{alternative}$[a]	$2\Delta lnL$[b]	*p*-Value	M2 parameters estimate
RIG-I					ω_1 = 0.07 (p = 0.571)
$M1_{(no\ selection)}$ vs. $M2_{(selection)}$	-20727.6	-20727.6	0	n.s.	ω_2 = 1.00 (p = 0.429)
					ω_3 = 4.57 (p = 0.000)
MDA5					ω_1 = 0.06 (p = 0.641)
$M1_{(no\ selection)}$ vs. $M2_{(selection)}$	-21908.5	-21908.5	0	n.s.	ω_2 = 1.00 (p = 0.359)
					ω_3 = 4.43 (p = 0.000)
LGP2					ω_1 = 0.08 (p = 0.740)
$M1_{(no\ selection)}$ vs. $M2_{(selection)}$	-18390.7	-18390.7	0	n.s.	ω_2 = 1.00 (p = 0.260)
					ω_3 = 5.16 (p = 0.000)

[a] lnL: log-likelihood scores.
[b] 2ΔlnL: likelihood ratio test (LRT) to detect positive selection.
n.s. – non-significant.

Supplementary Table S3 - Positively-selected codon positions for RIG-I, MDA5 and LGP2 determined by six different methods

	RIG-I	MDA5	LGP2
PAML M8[a]	10*, 13*, 31, **62**, **80***, 107*, **59**, **846**, 1008*		**74**, **180***, 269*, 324*, 347*, **438****, **627***, 653*
	123, 142*, 183*, 196*, 201*, **224***, 241*, **245**, 528*, 534*, 585*, 601*, 610, 703, **783**, **800**, **808***, **816***, 871*, **874***, 920, 944*		
SLAC[b]	**250**, **337**, **783***, **800**, **808**, **816**, **874***	**11**, **53***, **59***, **212***, **218**, **305**, **395**, **533***, **673**, **684**, **871**, **909***, **1002**	**97**, **145***, **180**, **205***, **316****, **438**, 554, **618***, 674
FEL[b]	5, **35***, **61***, **62***, **80**, 97, **135***, 208, **224***, **240***, **245***, **250***, **337***, 467, 509*, 510, **516**, 535, 630*, 681, 684, 768, **783***, **800***, 805, **808****, **816***, 823*, 858, **874****, 878, 889	**11***, 17, **53****, **59****, **73****, 142, **144***, **149**, **156**, 172, **197***, 204, **212****, **218***, **242***, 252, **305***, 307, 348, 385, **395**, 486, 523, **533****, 672*, **673**, **684***, 713, **846***, **871***, 907*, **909****, **1002***	47, **97**, **145***, **180***, 197, **205***, 233, **312**, **316****, **438****, **618***, **627**, 658
REL[c]	**35**, **61**, **62**, **80**, **135**, **224**, **240**, **245**, 308, **337**, **516**, **783**, **800**, **808**, **816**, **874**	**11**, 17, 24, **53**, **59**, **73**, **144**, **149**, **156**, **197**, 204, **212**, **218**, **242**, 302, 303, **305**, 414, **533**, 649, 671, **673**, **684**, 688, 713, **846**, **871**, **909**, **1002**, 1008, 1033	**74**, **97**, **145**, **180**, **205**, **312**, **316**, **438**, **618**
MEME[b]	5, 21*, **35***, 44, **61***, **62***, **80**, 97*, 98, 121*, 123*, **135**, 150*, 195*, 196**, 198, 211, **224**, 232*, 236, **240****, 242**, **245***, 246**, **250****, 252*, 259*, 262**, **337***, 407*, 467*, 478**, 481, 492, 497, 500, 509*, 510, **516***, 528, 535*, 630, 655**, 656*, 667, 669**, 681, 684*, 696*, 768, 772, 779, **783***, **800***, 805, **808***, 814*, **816***, 818, 823, 849, 851, 858*, 865, 870*, 873**, **874***, 889*, 926, 940**, 942**	**11****, 26, 42, **53***, **59****, **73***, 142*, **144***, **156**, 195, **197**, **212****, **218****, 228, 237*, **242**, 303, **305***, 307*, 348*, 374, 385*, 386, **395**, 437*, 440*, 482*, **533***, 551**, 566, 589, 593*, 594**, 595, 599*, 641, **673***, **684***, 688, 691, **846***, **871***, 907, **909***, 967*, 995**, **1002**, 1004, 1032*	47*, **74***, **97**, 125*, 130, **145***, **180***, **205***, 233, 253*, 272, 276, 301, **312**, **316***, 318, 321, 327, 368, 401, 405*, 406*, 416, **438****, 519**, 520*, 569, 579*, 594**, 609**, **618***, **627**, 637*, 674
FUBAR[d]	**62**, **80**, **224**, **245**, **783**, **800**, **808***, **816**, **874***	**53***, **59***, **73***, **149**, **212**, **305**, **533***, 649, **673**, **846**, **909***, **1002***	**180***, **205**, **316**, **438****, **618**, **627***

[a] Sites with posterior probabilities >0.90 in the BEB (Bayes empirical Bayes) analyses (*: P>0.95; **: P>0.99).
[b] Sites with significance level <0.1 (*: p<0.05; **: p<0.01).
[c] Sites with Bayes Factor >50.
[d] Sites with posterior probabilities >0.90 (*: P>0.95; **: P>0.99).
Sites identified by three or more than three methods are in bold.

Paper 4

Evolution of viral sensing RIG-I-like receptor genes in Leporidae genera *Oryctolagus*, *Sylvilagus*, and *Lepus*

A. Lemos de Matos, G. McFadden & P. J. Esteves

1. Summary

One of the most severe European rabbit (*Oryctolagus cuniculus*) pathogens is myxoma virus (MYXV), a rabbit-specific leporipoxvirus that causes the highly lethal disease myxomatosis. Other leporid genera, *Sylvilagus* and *Lepus*, encompass species with variable susceptibilities to MYXV, but these do not develop the lethal form of the disease. The protective role of the retinoic acid-inducible gene-I (RIG-I/*DDX58*) in sensing MYXV in nonpermissive human myeloid cells prompted the study of the RIG-I-like receptor (RLR) family evolution in the three leporid genera. This viral-sensor family also includes the melanoma differentiation-associated factor 5 (MDA5/*IFIH1*) and the laboratory of genetics and physiology 2 (LGP2/*DHX58*). Considering specifically the MYXV susceptible host (European rabbit) and one of the virus natural long-term hosts (*Sylvilagus bachmani*, brush rabbit), the amino acid differences of positively selected sites in RIG-I between the two species were located in the protein region responsible for viral RNA recognition and binding, the repressor domain. Such differences might play a determinant role in how MYXV is sensed. When looking for episodic selection on MDA5 and LGP2 of the eastern cottontail (*Sylvilagus floridanus*), we also uncovered evidence of selective pressures that might be exerted by a species-specific leporipoxvirus, the Shope fibroma virus. Finally, a putative alternative splicing case was identified in *Oryctolagus* and *Lepus* MDA5 isoforms, corresponding to the deletion of one specific exon. This study provided the first insights into the evolution of the leporid RLR gene family that helps illuminate the origins of the species-specific innate responses to pathogens, and more specifically to MYXV.

2. Introduction

The rabbit-specific poxvirus myxoma virus (MYXV) is one of the most severe known European rabbit (*Oryctolagus cuniculus*) pathogens with major clinical implications in populations around the world [1-3]. MYXV is a member of the *Leporipoxvirus* genus in the *Poxviridae* family that causes a highly lethal disease in European rabbits, called myxomatosis, which is characterized by the systemic spread of the virus and the development of secondary skin lesions [4]. The complete virus replication cycle occurs in the cytoplasm of infected cells and a large spectrum of host-interactive immunomodulatory proteins are expressed by MYXV [1, 5]. Recently, the pattern recognition receptor (PRR) retinoic acid-inducible gene-I (RIG-I/*DDX58*) was described as a cytoplasmic sensor for MYXV infection in primary human macrophages, possibly mediated by the double-stranded RNA (dsRNA) generated by divergent transcription events for viral genes [6]. Human macrophages are nonpermissive for MYXV replication, and RIG-I sensing of the virus in these cells leads to the co-induction of type I interferon (IFN) and tumor necrosis factor (TNF), which rapidly leads to the abort of the virus replication cycle [6]. Thus, it is presumed that MYXV possesses the capacity to subvert RIG-I sensing in rabbit cells, by still-unknown mechanisms.

RIG-I is one of the elements of the RIG-I-like receptor (RLR) family, which also contains two other members: the melanoma differentiation-associated factor 5 (MDA5/ *IFIH1*) and the laboratory of genetics and physiology 2 (LGP2/*DHX58*) [7-11]. This family of pattern recognition receptors is responsible for recognizing "nonself" nucleic acids from actively replicating viruses in the cytoplasm of infected cells [12]. Both RIG-I and MDA5 detect several RNA and DNA viruses [8, 10], which consequently signal the production of anti-viral cytokines, such as the type I IFNs, and the induction of an anti-viral response [11].

The *Oryctolagus* genus, with European rabbit as the only extant representative species, belongs to the Leporidae family (Order Lagomorpha) and it is one of the best-studied leporid genera, together with *Sylvilagus* and *Lepus* genera which diverged ~10 and ~12 Mya, respectively, from *Oryctolagus* [13-15]. These three well-studied leporid genera have different susceptibilities to MYXV. Indeed, although hares (genus *Lepus*) very rarely develop generalized disease following exposure to MYXV, myxomatosis after natural infection has been reported in European brown hares (*Lepus europaeus*) [16]. On the other hand, American cottontail rabbits (genus *Sylvilagus*) tapeti (Sylvilagus brasiliensis) and brush rabbit (Sylvilagus bachmani) serve as the natural host for different MYXV strains and only exhibit local skin fibromas [1-3].

The protective role of RIG-I in sensing MYXV in nonpermissive human myeloid cells prompted us to study the genetic and evolutionary aspects of the RLR genes in the three leporid genera where the host-virus interaction dynamics of MYXV has been previously reported. Our intent was to examine whether MYXV might have evolved RLR-specific anti-host mechanisms that could be revealed by evidence of selective evolutionary pressures exerted on the rabbit RLR gene family. Special attention was given to RLR gene molecular evolution within the MYXV susceptible host (European rabbit) and also for one of the MYXV natural long-term hosts (brush rabbit). Finally, as known host innate sentinel proteins involved in the recognition of multiple viral pathogens, RLRs could be subjected to long-term positive selection pressures by many viruses, and evidence of positive selection could suggest specifics of those viruses capable of antagonizing RLR sensing functions.

3. Materials and Methods

3.1. Leporid RIG-I-like receptors gene and protein sequences

Total RNA preparation and cDNA synthesis were performed for the two European rabbit subspecies, *Oryctolagus cuniculus cuniculus* and *O. cuniculus algirus*, brush rabbit (*Sylvilagus bachmani*), eastern cottontail (*S. floridanus*), Iberian hare (*Lepus granatensis*) and European brown hare (*L. europaeus*) samples, and protocols are described elsewhere [17].

The European rabbit *RIG-I* (ENSOCUG00000004710), *MDA5* (ENSOCUG00000002863), and *LGP2* (ENSOCUG00000013278) gene sequences available in Ensembl database (http://www.ensembl.org/) were used as templates for primer design. Amplification by RT-PCR was performed for leporid *RIG-I* (F: 5'-ATGACCGCCGAGGAGCGG-3'; R: 5'-TCATTCAGACAGTTCTTCAGGATTAAATGGTATC-3'), *MDA5* (F: 5'-AAGATGTCGAATGGTTATTCCGCAGAC-3'; R: 5'-CTAATCTTCATCACTTA AGAGACAGTATTCTGAGTAG-3'), and *LGP2* (F: 5'-ATGGAGCTACGCCCCTACCAG-3'; R: 5'-TCAGTCCAGCGAGAGGTCGGAC-3'). RT-PCR amplification was performed using the Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Finnzymes, Waltham, MA, USA) and according to manufacturer's protocol. Amplified PCR products were cloned into the pGEM-T Easy vector (Promega, Madison, WI, USA) and sequenced by the Sanger Sequencing core of the Genomics division in the ICBR (Interdisciplinary Center

for Biotechnology Research), University of Florida. Nucleotide sequence data obtained in this study have been submitted to GenBank and have been assigned the following accession numbers: KF640607 to KF640636.

To predict the protein structure of the three RLR proteins for European rabbit (MYXV susceptible host) and brush rabbit (MYXV natural long-term host) we used the I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) [18]. Proteins structure visualization was performed with the UCSF Chimera package (http://www.cgl.ucsf.edu/chimera/) [19].

Mammalian RLR gene coding sequences available in NCBI (http://www.ncbi.nlm.nih.gov) and Ensembl databases were collected and aligned altogether with the leporid RLR sequences obtained in this study by using Clustal W [20] implemented in BioEdit v7.0.9 [21]. Translation into protein sequences (Supplementary Figure S1 to Supplementary Figure S3) was performed using also BioEdit [21].

3.2. Molecular evolutionary analyses

The nucleotide sequences alignment of each RLR gene was firstly tested for recombination, as this biological process can mislead molecular evolutionary analyses [22]. The software GARD [23] implemented in the Datamonkey Web server [24, 25] was applied to detect possible recombination breakpoints on each gene alignment. Nevertheless, no significant breakpoints were identified in the three alignments.

To look for signatures of natural selection operating in *RIG-I*, *MDA5*, and *LGP2* alignments, we used PAML package [26] and compared site-based models to determine if a model that allows positive selection (alternative model, M8) is a better fit to the data than a neutral model (null model, M7). Likelihood ratio test (LRT) was performed with two degrees of freedom to compare the fit of the two models by using the likelihood scores of the null neutral and alternative selection models. To perform the selection analyses in PAML package [26], the phylogeny and corresponding tree topology of each alignment had to be determined first. The nucleotide substitution model for each RLR gene phylogenetic tree was indicated by the Akaike Information Criterion (AIC) implemented in jModelTest v0.1.1 [27] and the phylogenetic Maximum Likelihood (ML) reconstruction was performed using GARLI v2.0 [28]. For both *RIG-I* and *MDA5* phylogenetic trees, GTR+G was retrieved as the best-fit model of nucleotide substitution, while the TPM2uf+I+G model of nucleotide substitution was best-fitted for the *LGP2* alignment.

Six different codon-based ML methods were applied to detect codons under positive selection on *RIG-I*, *MDA5*, and *LGP2* alignments. Model M8 from PAML package [26] was one of the methods used to detect codons under positive selection, and a Bayes empirical Bayes (BEB) approach was employed to detect codons with a posterior probability >90% of being under selection [29]. Five other methods, using HyPhy software implemented in the Datamonkey Web server [24, 25], were also applied to detect codons under selection for the three genes: the Single Likelihood Ancestor Counting (SLAC) method, the Fixed Effect Likelihood (FEL) method, the Random Effect Likelihood (REL) method [30] and the recently described Mixed Effects Model of Evolution (MEME) [31] and Fast Unbiased Bayesian AppRoximation (FUBAR) [32] methods. To avoid a high false-positive rate [30], codons with *p*-values <0.1 for SLAC, FEL, and MEME models, Bayes Factor >50 for REL model and a posterior probability >0.90 for FUBAR were accepted as candidates for selection.

PAML branch-site model A [33], which allows ω (d_N/d_S) ratios to vary both among tree lineages and amino acid sites, was performed for leporid branches on *RIG-I*, *MDA5*, and *LGP2* phylogenetic trees to detect evidence of episodic selection. Each individual leporid branch on a phylogenetic tree was analyzed as a foreground branch independently, and the remaining lineages were denominated as background branches. In branch-site model A, three ω ratios are assumed for foreground ($0 < \omega_0 < 1$, $\omega_1 = 1$, $\omega_2 > 1$) and two ω ratios for background ($0 < \omega_0 < 1$, $\omega_1 = 1$). The LRT performed for each branch was significant for $2\Delta\ln L > 3.84$ [26, 33]. The null model is the same as model A, but $\omega_2 = 1$ is fixed [33]. The BEB approach was also used to calculate the posterior probability of a specific codon site and to identify those most likely to be under positive selection (posterior probability, >90%) [29].

4. Results

4.1. RIG-I-like receptor genes prevalence in Leporidae family

For the three leporid RLR genes more than one allele was identified for some species and even different specimens revealed diverse allelic forms (Supplementary Figure S1 to Supplementary Figure S3). Yet, two aspects concerning the leporid RLR genes are worthy of note.

All the leporid sequences obtained for *RIG-I*, including for European rabbit subspecies, showed a noticeable difference in one specific region when compared with the two European rabbit RIG-I transcripts annotated in the Ensembl database

(ENSOCUG00000004710). According to the information available in this database, the dissimilar region matched exons 18, 19, and 20, corresponding to a total of 72 nucleotides. At this point, we assumed that there was a computational prediction error for the *RIG-I* gene in European rabbit genome. Next, we used the 72 nucleotide-portion of the European rabbit subspecies *Oryctolagus cuniculus cuniculus* obtained in this study (the same European rabbit subspecies which genome is annotated in Ensembl) to perform a BLAST analyses and verify if this sequence was present somewhere in *RIG-I* genomic DNA. Indeed, we obtained 100% similarity to the last 72 nucleotides of the intron sequence between exon 20 and 21. Therefore, the sequence annotated in Ensembl as exons 18, 19, and 20 is incorrect and should be replaced by the intronic sequence juxtaposed to the beginning of exon 21, reducing the total number of *RIG-I* exons from 21 to 18. The correct *RIG-I* gene coding sequence for the European rabbit subspecies *Oryctolagus cuniculus cuniculus* was submitted to GenBank and it has been assigned the accession number KF640625.

Another aspect was the observation of a 50-amino acid residues deletion in two specific alleles of *MDA5*: European rabbit subspecies *Oryctolagus cuniculus algirus* allele 2 and Iberian hare allele 2. Ensembl annotation on European rabbit *MDA5* gene allowed us to match the deleted region to exon 12.

4.2. Leporid RIG-I-like receptor genes molecular evolution

In the present study, we have used the entire coding sequence of RLR genes collected for different mammalian lineages and the data obtained for *Oryctolagus*, *Sylvilagus*, and *Lepus* species to help us identify the variation of functional sites of leporid *RIG-I*, *MDA5*, and *LGP2* genes. Additionally, it has been documented that the accuracy and power of the LRT in detecting positive selection is affected by sequence divergence, meaning the expected number of nucleotide substitutions per codon along the tree, which is given by the tree-length parameter [34]. Tree-length for each gene alignment comprising only leporid data presents very low values, between 0.133 and 0.185 (very short trees), corresponding to highly similar sequences. Sequences with such level of similarity carry little information, causing low power of the LRT [34]. Therefore, tree-length values for the mammalian alignment of the three genes fell into an intermediate and realistic level of sequence divergence, between 4.501 and 6.530, which confers power to the codon models indicated by the LRT scores [34].

Evidence for positive selection on *RIG-I*, *MDA5*, and *LGP2* gene alignments was detected using PAML package [26] site-specific models M7 versus M8. These models test at the codon level whether a hypothesis that allows for positive selection (M8) is a

better fit to the data when compared with a null neutral hypothesis (M7). Results on the LRT performed between the likelihood scores of the null neutral and alternative selection models for each gene is indicated in Table 1. Positive selection model M8 gave a significantly better fit to the data for the three RLR genes.

The total number of sites under selection for each RLR was determined using six different ML methods (Table 1), but only those identified by at least three of the six were considered to be under positive selection [35]. Sites numeration in Table 1 corresponds to the alignments represented in Supplementary Figure S1 to Supplementary Figure S3. Thirteen sites in RIG-I alignment (Supplementary Figure S1), twenty sites in MDA5 alignment (Supplementary Figure S2), and nine sites in LGP2 alignment (Supplementary Figure S3) were detected by ML methods as being under positive selection (Table 1).

From the identified sites under selection, the positions that exhibited amino acid differences between the European rabbit (the susceptible host of MYXV) and the brush rabbit (the natural long-term host of MYXV) were annotated in the predicted structure of each protein. Therefore, from the detected sites under selection, two of the three RIG-I amino acid differences between the two species (positions 805 and 816) are located in a known defined functional domain of this protein, the repressor domain (Figure 1a, b shadowed regions). For the MDA5 protein and based on the total number of sites under selection, the two amino acid differences between the European rabbit and the brush rabbit are both located in the N-terminal caspase activation and recruitment domain one (CARD1; Figure 1c, d). Considering the identified sites under selection for LGP2, there was only one amino acid difference between the two species, which is located in the C-terminal region of the helicase domain (Figure 1e, f).

When performing the branch-site model A analysis [26, 33] on the leporid cluster of each RLR phylogenetic tree, we were able to identify specific leporid branches under episodic selection. For *RIG-I*, only one specific sample and allele of Iberian hare (S3_Allele2; Figure 2) presented a significant likelihood ratio test ($2\Delta\ln L=9.04$; $p<0.01$) suggesting evidence of positive selection. The one positively selected site presented a posterior probability below the 90% confidence interval in the BEB analysis [29].

Table 1 - Positive selection analyses for RIG-I-like receptor genes

Gene	Test of selection				Sites under selection identified by at least three different methods						Total no. of sites
	lnL[a] M7 (neutral)	lnL[a] M8 (selection)	2ΔlnL[b]	*p*-Value	PAML M8[c]	SLAC[d]	FEL[d]	REL[e]	MEME[d]	FUBAR[f]	
RIG-I	-22393.34	-22347.17	92.34	<0.001	80**, 224**, 245, 808, 816**, 874*	250, 337, 783*, 805, 808, 816*, 874*	35*, 62**, 80*, 135*, 224*, 245*, 250*, 337*, 783*, 805*, 808**, 816**, 874**	35, 62, 80, 135, 224, 245, 783, 805, 808, 816, 874	35*, 62*, 80*, 135, 224, 245*, 250**, 337*, 783*, 805, 808*, 816**, 874*	62, 80, 224, 245, 783, 808, 816, 874	13[g]
MDA5	-23281.72	-23254.74	53.96	<0.001	59, 73*	11, 53*, 59*, 73, 144, 212*, 218, 305, 533*, 684, 871, 909*, 1002	11*, 17, 53**, 59**, 73**, 144*, 156, 197*, 204*, 212**, 218*, 242*, 305*, 533**, 684*, 846*, 871*, 909**, 1002*	11, 17, 53, 59, 73, 144, 156, 197, 204, 212, 218, 242, 305, 533, 673, 684, 846, 871, 909, 1002	11**, 17, 53**, 59**, 73**, 144*, 156, 197, 204, 212**, 218**, 242, 305*, 533*, 673*, 684*, 846*, 871*, 909*, 1002	53*, 59*, 73**, 197, 204, 212, 305, 533*, 673, 846, 909**, 1002*	20[h]
LGP2	-19561.92	-19541.36	41.12	<0.001	180*, 438**, 627	97, 145*, 180, 205*, 312, 316**, 438*, 618*, 627	97, 145*, 180**, 205*, 312, 316**, 438**, 618**, 627	145, 180, 205, 316, 438, 618	97, 145*, 180*, 205*, 312, 316*, 438**, 618*,	180*, 205, 316, 438**, 618, 627*	9[i]

[a] lnL: log-likelihood scores.
[b] 2ΔlnL: likelihood ratio test (LRT) to detect positive selection.
[c] Codons with posterior probabilities >90% in the BEB analyses (*: P>95%; **: P>99%).
[d] Codons with significance level <0.1 (*: p<0.05; **: p<0.01).
[e] Codons with Bayes Factor >50.
[f] Codons with posterior probabilities >0.90 (*: P>0.95; **: P>0.99).
[g] Sites: 35, 62, 80, 135, 224, 245, 250, 337, 783, 805, 808, 816, and 874.
[h] Sites: 11, 17, 53, 59, 73, 144, 156, 197, 204, 212, 218, 242, 305, 533, 673, 684, 846, 871, 909, and 1002.
[i] Sites: 97, 145, 180, 205, 312, 316, 438, 618, and 627.

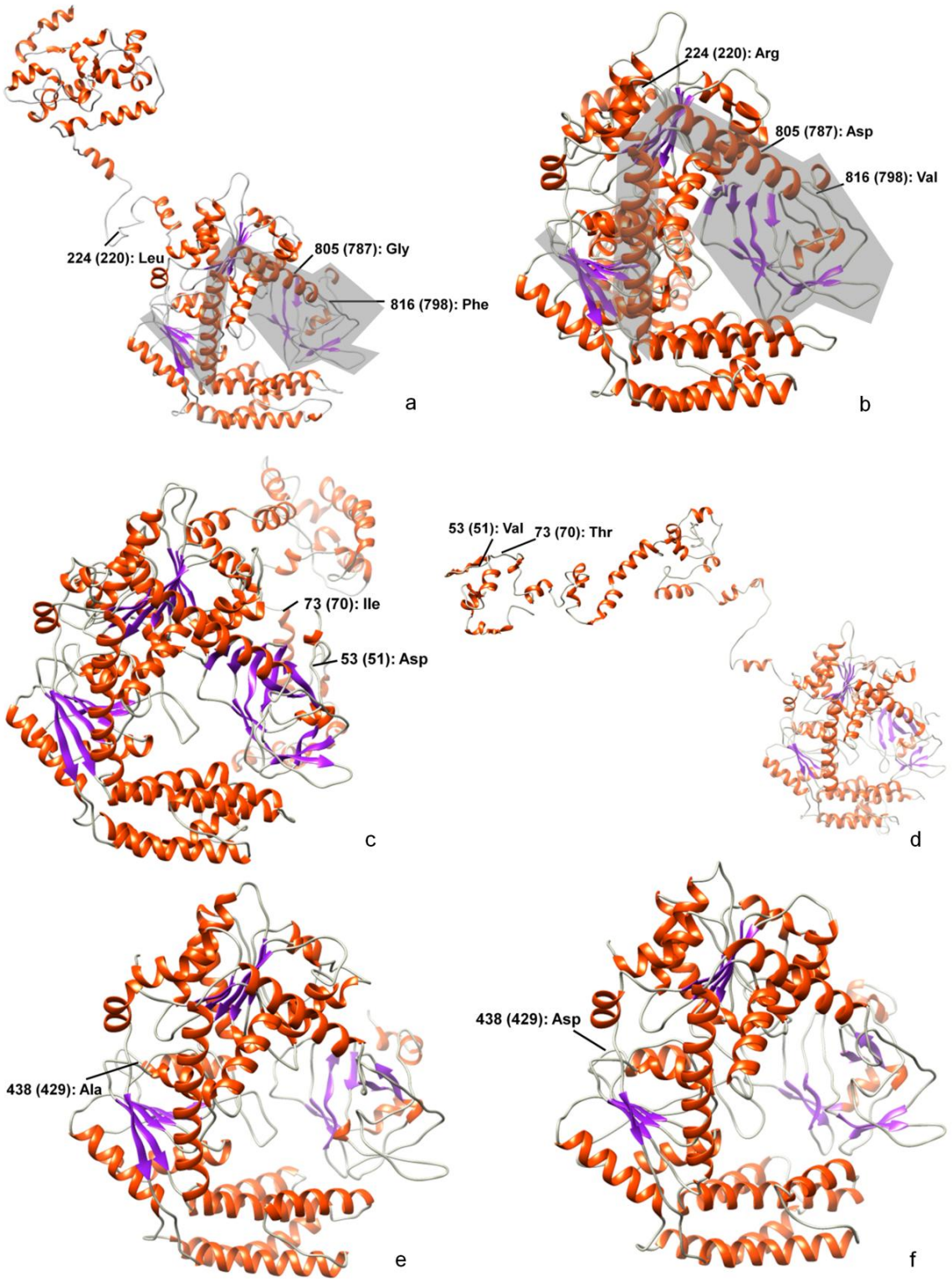

224 (220): Leu
805 (787): Gly
816 (798): Phe
a
224 (220): Arg
805 (787): Asp
816 (798): Val
b
73 (70): Ile
53 (51): Asp
c
53 (51): Val
73 (70): Thr
d
438 (429): Ala
e
438 (429): Asp
f

Figure 1 - Structural prediction of the three RIG-I-like receptors for the European rabbit and the brush rabbit.
The deduced protein sequences obtained for the two leporid species during this study were used to run in the I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/). The three dimensional protein structures a (RIG-I), c (MDA5) and e (LGP2) belong to the European rabbit, while structural predictions b (RIG-I), d (MDA5) and f (LGP2) refer to the brush rabbit. RIG-I (a and b) repressor domain is shadowed in grey. The amino acid differences of positively-selected sites in the three RIG-I-like receptors between the two species are represented. For each amino acid pinpointed, the first number matches its position in the corresponding mammalian alignment (Supplementary Figure S1 to Supplementary Figure S3), whereas the number within brackets is the amino acid "real" position in the species protein. Orange: helical structure; Purple: stranded structure; Grey: coiled structure.

The branch-site analysis and following likelihood ratio test performed on the eastern cottontail branches on both *MDA5* ($2\Delta\ln L=5.63$; $p<0.05$) and *LGP2* ($2\Delta\ln L=5.07$; $p<0.05$) phylogenetic reconstructions, cladograms in Figure 3 and Figure 4 respectively, suggested the action of episodic selection. Nevertheless, only *LGP2* presented a positively selected site with a posterior probability above 90% in the BEB analysis, the site 547 ($p=0.928$). Finally, also on the *LGP2* phylogenetic tree (Figure 3), the branch leading to the common ancestor of the three leporid genera exhibited striking evidence of positive selection ($2\Delta\ln L=8.13$; $p<0.01$). In this case, only one positively selected site with posterior probability above 90% in the BEB analysis was identified, the site 605 ($p=0.971$). It should be highlighted that the two referred sites are located in the defined LGP2 repressor domain.

5. Discussion

Many of the host innate sentinel proteins provide species-specific innate responses to pathogens, and they are also capable of mounting specific inhibitory strategies against specific host immune proteins. Studying this dynamic arms race between host and pathogen, and more specifically the long-term selective pressures imposed by pathogen on the host innate sentinel proteins, should provide new insights into these proteins biological functions. In this study, we were focused in documenting genetic and evolutionary striking aspects of the RLR family in three leporid genera with previous evidence of contact with a poxvirus, MYXV, known to express dozens of documented host-modulatory proteins [1, 5].

As the role of RIG-I in detecting MYXV in nonpermissive human myeloid cells has been previously described [6], the existence of three sites under selection with different amino acid residues between the susceptible and the natural long-term host of MYXV,

Figure 2 - *RIG-I* cladogram representing relationships among leporids and the remaining mammalian species.
The leporid cluster is shadowed in grey and leporid branches under episodic selection detected by branch-site model A analysis are colored in black, while the other branches in the diagram are represented in grey.

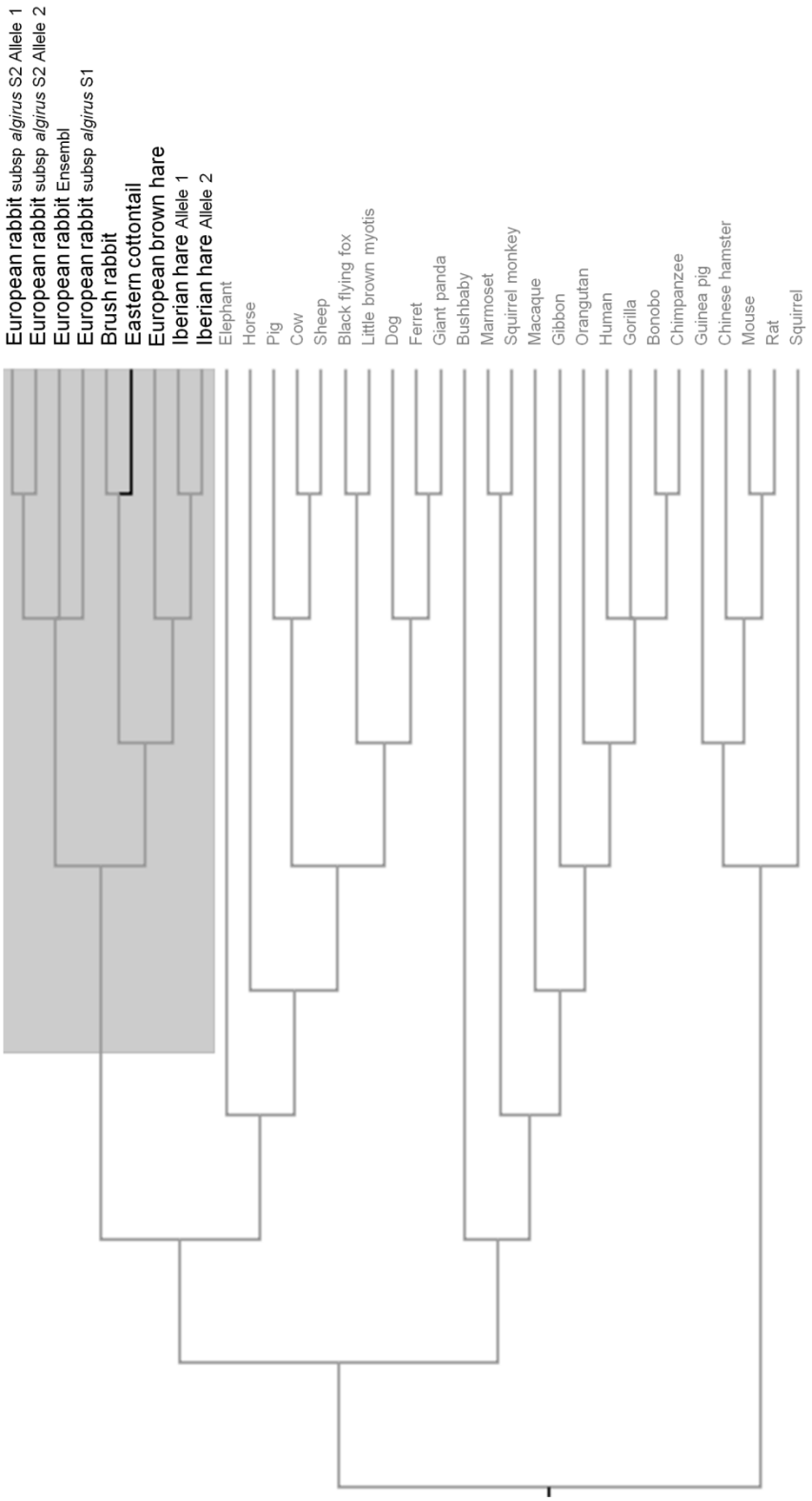


Figure 3 - *MDA5* cladogram representing relationships among leporids and the remaining mammalian species.
The leporid cluster is shadowed in grey and leporid branches under episodic selection detected by branch-site model A analysis are colored in black, while the other branches in the diagram are represented in grey.

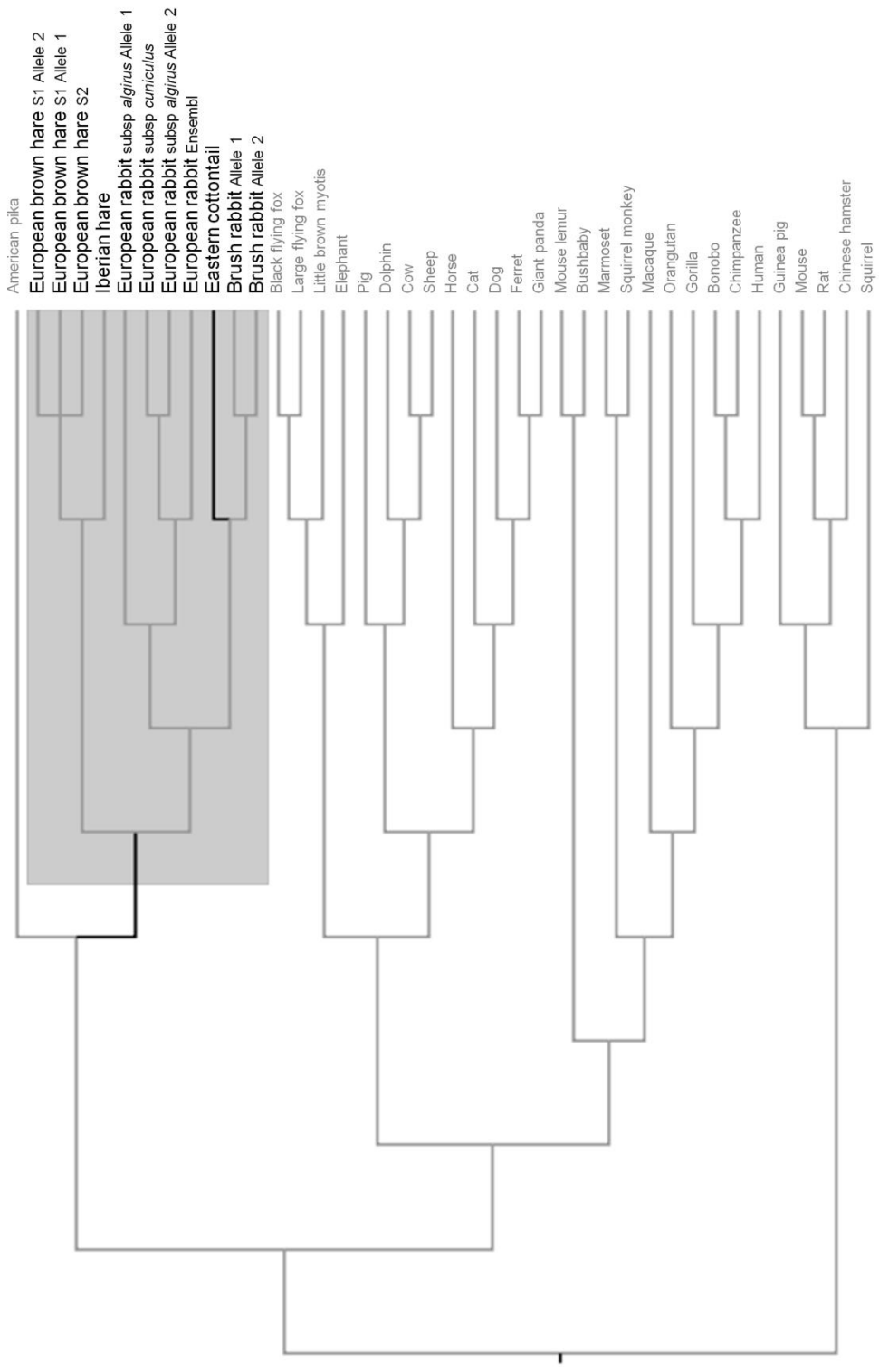


Figure 4 - *LGP2* cladogram representing relationships among leporids and the remaining mammalian species.
The leporid cluster is shadowed in grey and leporid branches under episodic selection detected by branch-site model A analysis are colored in black, while the other branches in the diagram are represented in grey.

European rabbit and brush rabbit respectively, is a significant result. Those amino acid alterations exhibited radical changes in their physicochemical properties, which can presumably mediate conformational changes in the protein structure [36]. Moreover, the location of two of those amino acid differences in the RIG-I repressor domain is revealing of the importance of these findings. The RIG-I repressor domain is responsible for recognizing and binding to its RNA substrates in a 5'-triphosphate (5'ppp)-dependent manner. In addition, binding studies clearly established that the 5'pppRNA binding site resides within the repressor domain [8, 37, 38]. We speculate that these differences might play a determinant role in how the MYXV is detected by RIG-I in the two hosts and in the following species-specific immune responses. We also propose that MYXV likely expresses one or more modulatory functions that specifically counteract RIG-I in rabbit cells, but this inhibition is somehow ineffective in human cells.

From the sites under selection identified for MDA5, the differences between MYXV host proteins are located in the CARD1 domain. The CARDs function as an interaction domain with other CARD-containing proteins, being fundamental for downstream MDA5 signaling [39, 40]. Therefore, as functional constraints are expected in this domain, the observed variability in European rabbit and brush rabbit might be explained as the existence of structural plasticity in the leporid MDA5 CARD1 domain. Some alleles of both *Oryctolagus* and *Lepus* MDA5 revealed a relevant genetic aspect, as they exhibited an isoform with a deletion that matched one specific exon. This genetic variation corresponds to a putative alternative splicing case that could be the result of an ancient polymorphism maintained across leporid genera. Studies on other immunological markers in leporid species demonstrated the preservation under balancing selection of trans-species polymorphisms among distantly related species [41-44].

The function of LGP2 in anti-viral immunity is still not clear, but in different studies its role as a negative and a positive regulator of RIG-I and MDA5 signaling has been described [45-47]. Consequently, it is hard to predict the possible consequences of the differences observed in the positive selection studies and of the results in the branch-site model A analysis.

Another leporipoxvirus closely related to MYXV, named Shope fibroma virus, is found exclusively in the eastern cottontail (*Sylvilagus floridanus*), causing cutaneous fibromas [16]. As the eastern cottontail branch on both MDA5 and LGP2 phylogenetic trees showed significant likelihood ratio test values, suggesting evidence of operating episodic selection, we can speculate that this naturally occurring viral pathogen might

be exerting long-term selective pressures on the two RLR genes in this specific species as well.

Further studies should be conducted to confirm the functional viability of the different alleles and isoforms obtained in this study for the leporid RLR family and to better understand the biological impact of the residues identified under selective pressures. Studies on the biomolecular structure of the three proteins from the European rabbit and the brush rabbit, in terms of their sensing capacities following MYXV infection, would also be very revealing.

6. Supplementary material

Supplementary material is appended to the present document by order of appearance in the main text.

Supplementary Figure S1 - RIG-I deduced protein sequences alignment for leporid and mammalian collected sequences. The mammalian RIG-I deduced protein sequences were retrieved from Ensembl and NCBI databases, and aligned with the leporid data using Clustal W implemented in BioEdit. The symbol "." represents the same codon as the reference sequence of human RIG-I protein, "?" symbolizes an undetermined codon and "-" represents a gap or deletion in the alignment.

Supplementary Figure S2 - MDA5 deduced protein sequences alignment for leporid and mammalian collected sequences. The mammalian MDA5 deduced protein sequences were retrieved from Ensembl and NCBI databases, and aligned with the leporid data using Clustal W implemented in BioEdit. The symbol "." represents the same codon as the reference sequence of human MDA5 protein, "?" symbolizes an undetermined codon and "-" represents a gap or deletion in the alignment.

Supplementary Figure S3 - LGP2 deduced protein sequences alignment for leporid and mammalian collected sequences. The mammalian LGP2 deduced protein sequences were retrieved from Ensembl and NCBI databases, and aligned with the leporid data using Clustal W implemented in BioEdit. The symbol "." represents the same codon as the reference sequence of human LGP2 protein and "-" represents a gap or deletion in the alignment.

7. References

1. Spiesschaert B, McFadden G, Hermans K, Nauwynck H, Van de Walle GR: The current status and future directions of myxoma virus, a master in immune evasion. *Veterinary Research* 2011, 42:76.
2. Kerr PJ: Myxomatosis in Australia and Europe: a model for emerging infectious diseases. *Antiviral Research* 2012, 93:387-415.
3. Kerr PJ, Donnelly TM: Viral infections of rabbits. *Veterinary Clinics of North America: Exotic Animal Practice* 2013, 16:437-468.
4. Best SM, Collins SV, Kerr PJ: Coevolution of host and virus: cellular localization of virus in myxoma virus infection of resistant and susceptible European rabbits. *Virology* 2000, 277:76-91.
5. Liu J, Wennier S, McFadden G: The immunoregulatory properties of oncolytic myxoma virus and their implications in therapeutics. *Microbes and Infection* 2010, 12:1144-1152.
6. Wang F, Gao X, Barrett JW, Shao Q, Bartee E, Mohamed MR, Rahman M, Werden S, Irvine T, Cao J, Dekaban GA, McFadden G: RIG-I mediates the co-induction of tumor necrosis factor and type I interferon elicited by myxoma virus in primary human macrophages. *PLoS Pathogens* 2008, 4:e1000099.
7. Bamming D, Horvath CM: Regulation of signal transduction by enzymatically inactive antiviral RNA helicase proteins MDA5, RIG-I, and LGP2. *The Journal of Biological Chemistry* 2009, 284:9700-9712.
8. Dixit E, Kagan JC: Intracellular pathogen detection by RIG-I-like receptors. *Advances in Immunology* 2013, 117:99-125.
9. Kato H, Takahasi K, Fujita T: RIG-I-like receptors: cytoplasmic sensors for non-self RNA. *Immunological Reviews* 2011, 243:91-98.
10. Loo YM, Gale M, Jr.: Immune signaling by RIG-I-like receptors. *Immunity* 2011, 34:680-692.
11. Yoneyama M, Kikuchi M, Matsumoto K, Imaizumi T, Miyagishi M, Taira K, Foy E, Loo YM, Gale M, Jr., Akira S, Yonehara S, Kato A, Fujita T: Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *The Journal of Immunology* 2005, 175:2851-2858.
12. Kawai T, Akira S: Innate immune recognition of viral infection. *Nature Immunology* 2006, 7:131-137.
13. Pinheiro A, Mera IF, Alves P, Gortázar C, Fuente J, Esteves P: Sequencing of modern *Lepus* VDJ genes shows that the usage of VHn genes has been retained

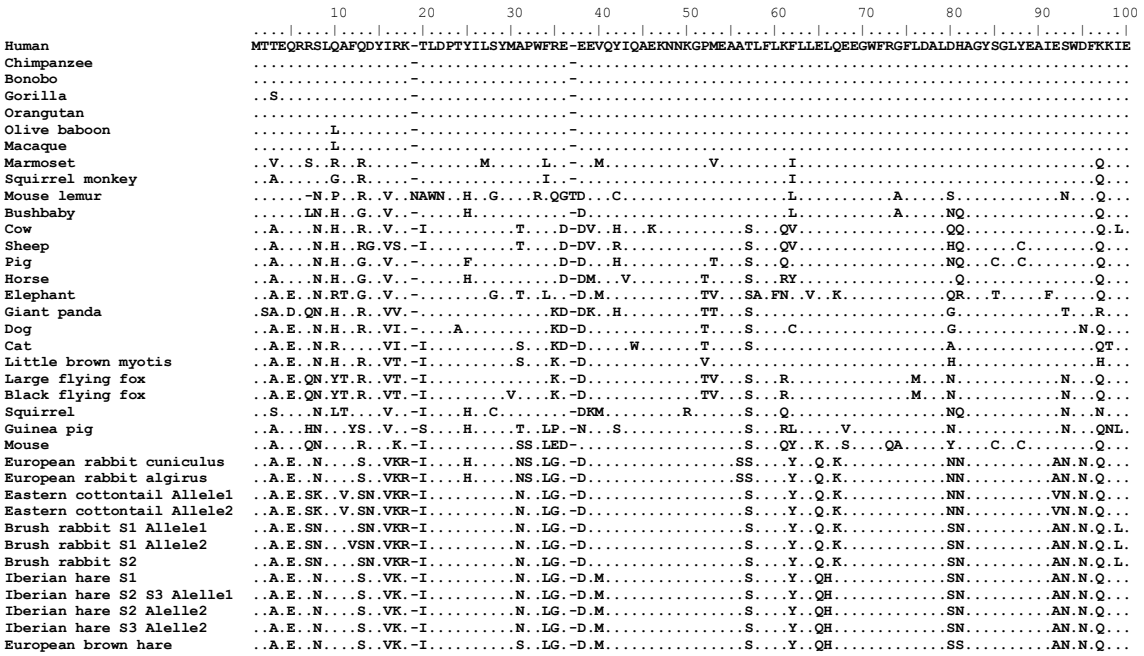
in both *Oryctolagus* and *Lepus* that diverged 12 million years ago. *Immunogenetics* 2013, 65:777-784.

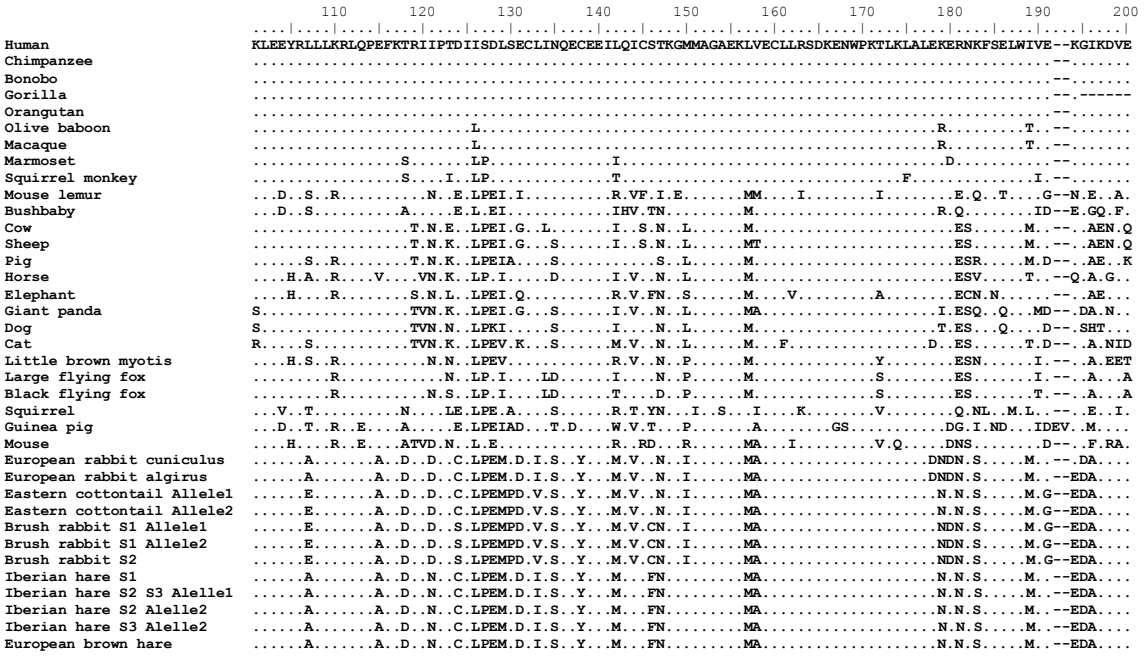
14. Matthee CA, Van Vuuren BJ, Bell D, Robinson TJ: A Molecular Supermatrix of the Rabbits and Hares (Leporidae) Allows for the Identification of Five Intercontinental Exchanges During the Miocene. *Systematic Biology* 2004, 53:433-447.
15. van der Loo W, Abrantes J, Esteves PJ: Sharing of endogenous lentiviral gene fragments among leporid lineages separated for more than 12 Million years. *Journal of Virology* 2009, 83:2386-2388.
16. Fenner FJ, Ratcliffe FN: *Myxomatosis.* Cambridge University Press; 1965.
17. de Matos AL, van der Loo W, Areal H, Lanning DK, Esteves PJ: Study of *Sylvilagus* rabbit TRIM5alpha species-specific domain: how ancient endoviruses could have shaped the antiviral repertoire in Lagomorpha. *BMC Evolutionary Biology* 2011, 11:294.
18. Zhang Y: I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 2008, 9:40.
19. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: UCSF Chimera - a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 2004, 25:1605-1612.
20. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 1994, 22:4673-4680.
21. Hall T: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 1999, 41:95-98.
22. Posada D, Crandall KA: The effect of recombination on the accuracy of phylogeny estimation. *Journal of Molecular Evolution* 2002, 54:396-402.
23. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD: GARD: a genetic algorithm for recombination detection. *Bioinformatics* 2006, 22:3096-3098.
24. Pond SL, Frost SD: Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 2005, 21:2531-2533.
25. Delport W, Poon AF, Frost SD, Kosakovsky Pond SL: Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 2010, 26:2455-2457.

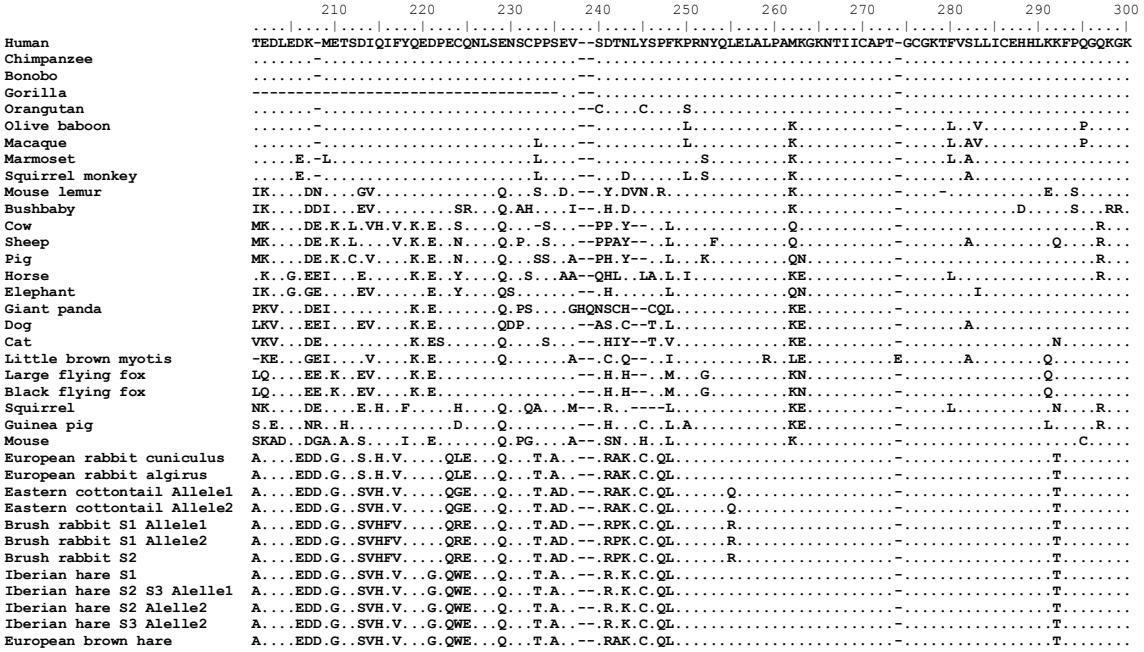
26. Yang Z: PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 2007, 24:1586-1591.
27. Posada D: jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 2008, 25:1253-1256.
28. Zwickl DJ: Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. *PhD Thesis.* University of Texas, 2006.
29. Yang Z, Wong WS, Nielsen R: Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* 2005, 22:1107-1118.
30. Kosakovsky Pond SL, Frost SD: Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* 2005, 22:1208-1222.
31. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL: Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics* 2012, 8:e1002764.
32. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K: FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution* 2013, 30:1196-1205.
33. Zhang J, Nielsen R, Yang Z: Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution* 2005, 22:2472-2479.
34. Anisimova M, Bielawski JP, Yang Z: Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution* 2001, 18:1585-1592.
35. Lemos de Matos A, Liu J, McFadden G, Esteves P: Evolution and divergence of the mammalian SAMD9/SAMD9L gene family. *BMC Evolutionary Biology* 2013, 13:121.
36. Majewski J, Ott J: Amino acid substitutions in the human genome: evolutionary implications of single nucleotide polymorphisms. *Gene* 2003, 305:167-173.
37. Baum A, Sachidanandam R, Garcia-Sastre A: Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107:16303-16308.
38. Cui S, Eisenacher K, Kirchhofer A, Brzozka K, Lammens A, Lammens K, Fujita T, Conzelmann KK, Krug A, Hopfner KP: The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I. *Molecular Cell* 2008, 29:169-179.

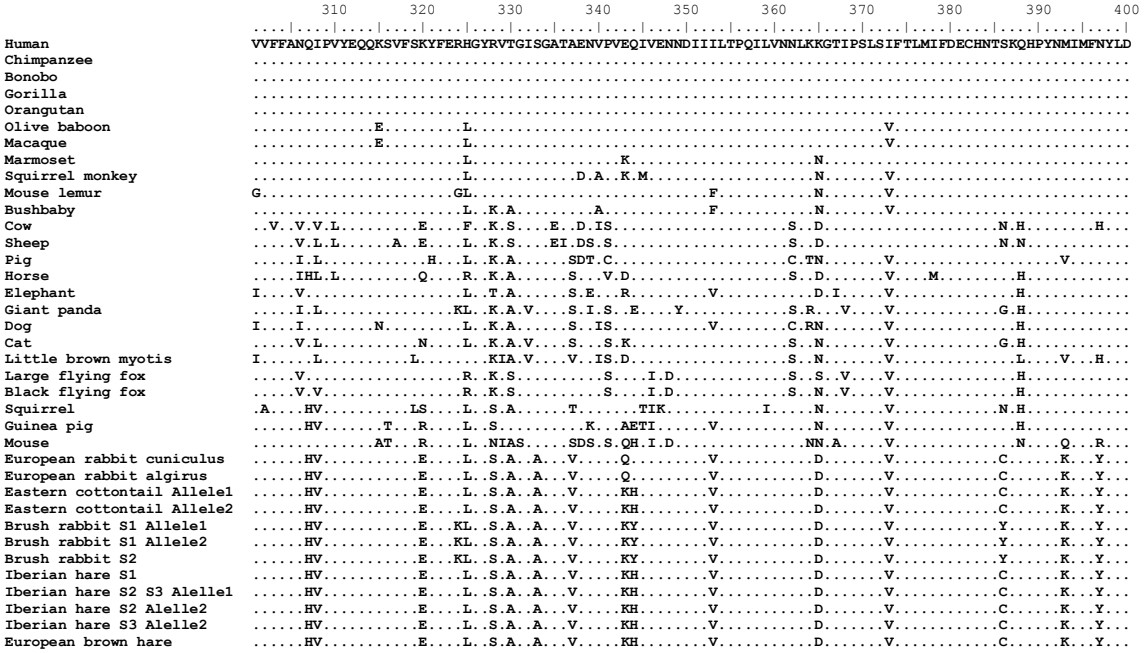
39. Bruns AM, Horvath CM: Activation of RIG-I-like receptor signal transduction. *Critical Reviews in Biochemistry and Molecular Biology* 2012, 47:194-206.
40. Schmidt A, Rothenfusser S, Hopfner KP: Sensing of viral nucleic acids by RIG-I: from translocation to translation. *European Journal of Cell Biology* 2012, 91:78-85.
41. Bouton C, van der Loo W: The trans-species nature of rabbit b locus polymorphism is supported by studies on the snow-shoe hare. *Immunogenetics* 1997, 45:444-446.
42. Esteves PJ, Lanning D, Ferrand N, Knight KL, Zhai SK, van der Loo W: The evolution of the immunoglobulin heavy chain variable region (IgVH) in Leporids: an unusual case of transspecies polymorphism. *Immunogenetics* 2005, 57:874-882.
43. Gouy de Bellocq J, Suchentrunk F, Baird SJ, Schaschl H: Evolutionary history of an MHC gene in two leporid species: characterisation of MHC-DQA in the European brown hare and comparison with the European rabbit. *Immunogenetics* 2009, 61:131-144.
44. Surridge AK, van der Loo W, Abrantes J, Carneiro M, Hewitt GM, Esteves PJ: Diversity and evolutionary history of the MHC DQA gene in leporids. *Immunogenetics* 2008, 60:515-525.
45. Rothenfusser S, Goutagny N, DiPerna G, Gong M, Monks BG, Schoenemeyer A, Yamamoto M, Akira S, Fitzgerald KA: The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I. *The Journal of Immunology* 2005, 175:5260-5268.
46. Satoh T, Kato H, Kumagai Y, Yoneyama M, Sato S, Matsushita K, Tsujimura T, Fujita T, Akira S, Takeuchi O: LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107:1512-1517.
47. Venkataraman T, Valdes M, Elsby R, Kakuta S, Caceres G, Saijo S, Iwakura Y, Barber GN: Loss of DExD/H box RNA helicase LGP2 manifests disparate antiviral responses. *The Journal of Immunology* 2007, 178:6444-6455.

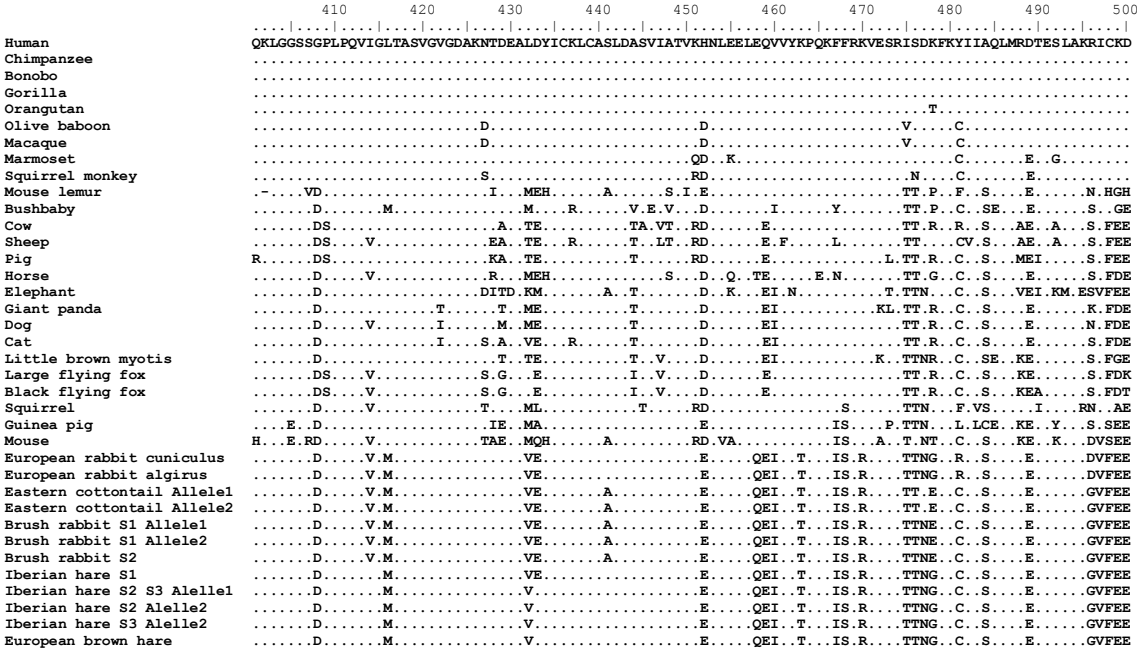
Supplementary Figure S1

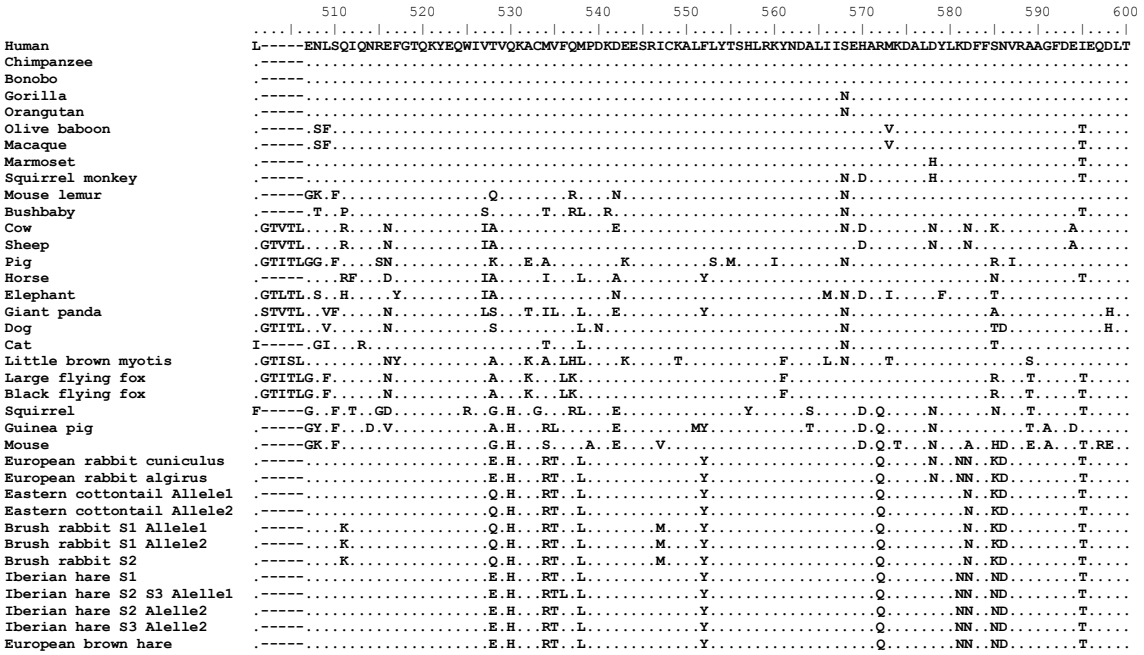


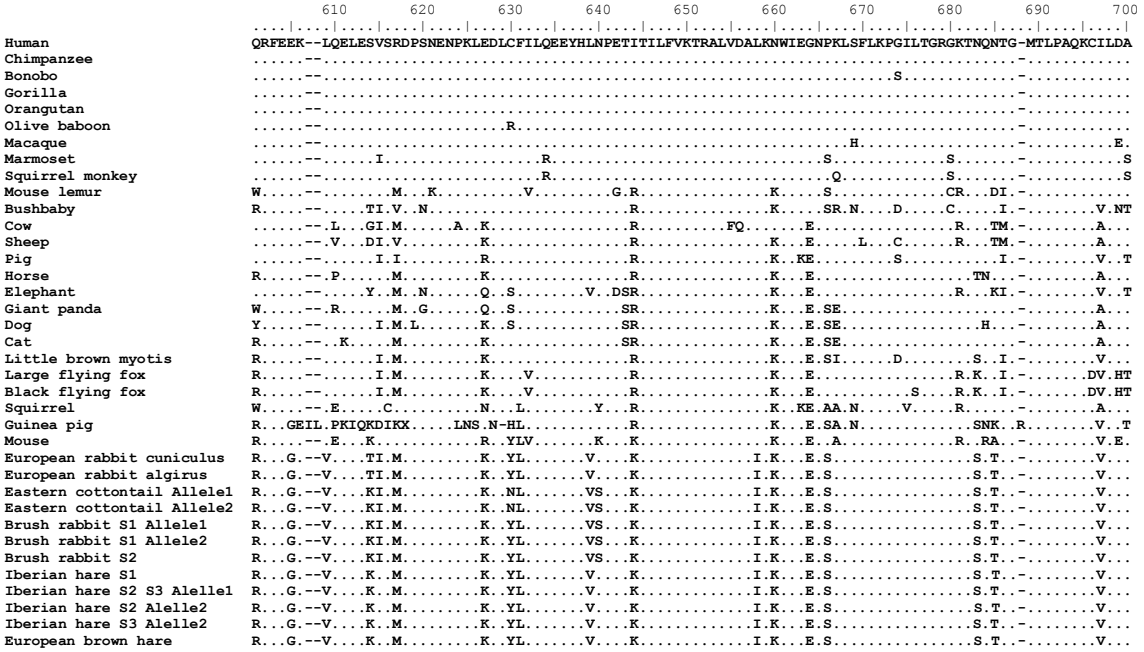


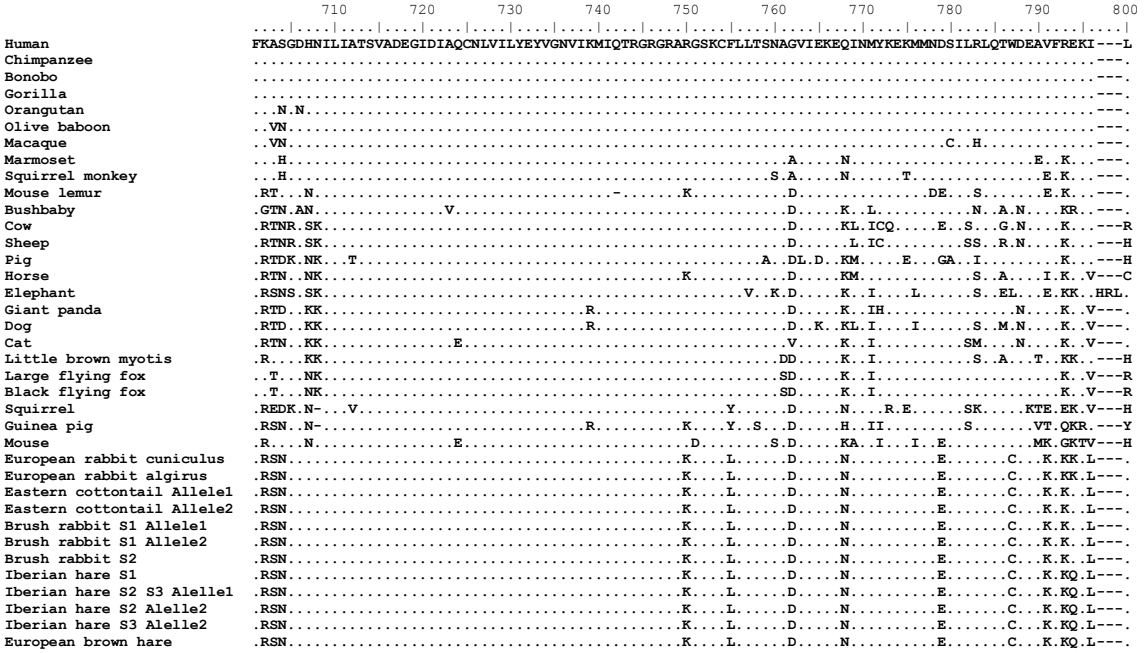

210 220 230 240 250 260 270 280 290 300
Human TEDLEDK-METSDIQIFYQEDPECQNLSENSCPPSEV--SDTNLYSPFKPRNYQLELALPAMKGKNTIICAPT-GCGKTFVSLLICEHHLKKFPQGQKGK
Chimpanzee
Bonobo
Gorilla
Orangutan
Olive baboon
Macaque
Marmoset
Squirrel monkey
Mouse lemur
Bushbaby
Cow
Sheep
Pig
Horse
Elephant
Giant panda
Dog
Cat
Little brown myotis
Large flying fox
Black flying fox
Squirrel
Guinea pig
Mouse
European rabbit cuniculus
European rabbit algirus
Eastern cottontail Allele1
Eastern cottontail Allele2
Brush rabbit S1 Allele1
Brush rabbit S1 Allele2
Brush rabbit S2
Iberian hare S1
Iberian hare S2 S3 Alelle1
Iberian hare S2 Alelle2
Iberian hare S3 Alelle2
European brown hare


310 320 330 340 350 360 370 380 390 400
Human VVFFANQIPVYEQQKSVFSKYFERHGYRVTGISGATAENVPVEQIVENNDIIILTPQILVNNLKKGTIPSLSIFTLMIFDECHNTSKQHPYNMIMFNYLD
Chimpanzee
Bonobo
Gorilla
Orangutan
Olive baboon
Macaque
Marmoset
Squirrel monkey
Mouse lemur
Bushbaby
Cow
Sheep
Pig
Horse
Elephant
Giant panda
Dog
Cat
Little brown myotis
Large flying fox
Black flying fox
Squirrel
Guinea pig
Mouse
European rabbit cuniculus
European rabbit algirus
Eastern cottontail Allele1
Eastern cottontail Allele2
Brush rabbit S1 Allele1
Brush rabbit S1 Allele2
Brush rabbit S2
Iberian hare S1
Iberian hare S2 S3 Alelle1
Iberian hare S2 Alelle2
Iberian hare S3 Alelle2
European brown hare


410 420 430 440 450 460 470 480 490 500
Human QKLGGSSGPLPQVIGLTASVGVGDAKNTDEALDYICKLCASLDASVIATVKHNLEELEQVVYKPQKFFRKVESRISDKFKYIIAQLMRDTESLAKRICKD
Chimpanzee
Bonobo
Gorilla
Orangutan
Olive baboon
Macaque
Marmoset
Squirrel monkey
Mouse lemur
Bushbaby
Cow
Sheep
Pig
Horse
Elephant
Giant panda
Dog
Cat
Little brown myotis
Large flying fox
Black flying fox
Squirrel
Guinea pig
Mouse
European rabbit cuniculus
European rabbit algirus
Eastern cottontail Allele1
Eastern cottontail Allele2
Brush rabbit S1 Allele1
Brush rabbit S1 Allele2
Brush rabbit S2
Iberian hare S1
Iberian hare S2 S3 Alelle1
Iberian hare S2 Alelle2
Iberian hare S3 Alelle2
European brown hare


510 520 530 540 550 560 570 580 590 600
Human L-----ENLSQIQNREFGTQKYEQWIVTVQKACMVFQMPDKDEESRICKALFLYTSHLRKYNDALIISEHARMKDALDYLKDFFSNVRAAGFDEIEQDLT
Chimpanzee
Bonobo
Gorilla
Orangutan
Olive baboon
Macaque
Marmoset
Squirrel monkey
Mouse lemur
Bushbaby
Cow
Sheep
Pig
Horse
Elephant
Giant panda
Dog
Cat
Little brown myotis
Large flying fox
Black flying fox
Squirrel
Guinea pig
Mouse
European rabbit cuniculus
European rabbit algirus
Eastern cottontail Allele1
Eastern cottontail Allele2
Brush rabbit S1 Allele1
Brush rabbit S1 Allele2
Brush rabbit S2
Iberian hare S1
Iberian hare S2 S3 Alelle1
Iberian hare S2 Alelle2
Iberian hare S3 Alelle2
European brown hare

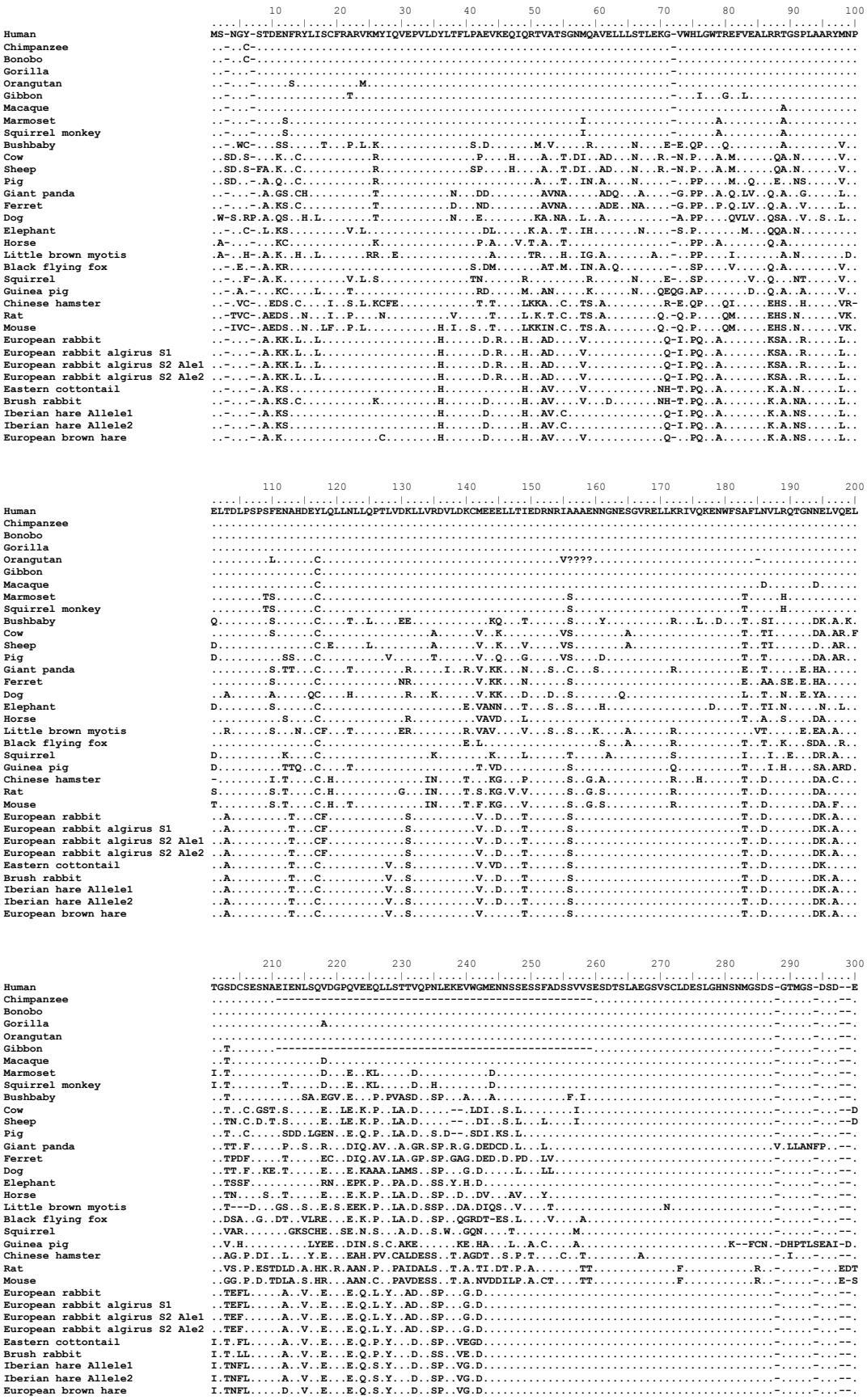

610 620 630 640 650 660 670 680 690 700
Human QRFEEK--LQELESVSRDPSNENPKLEDLCFILQEEYHLNPETITILFVKTRALVDALKNWIEGNPKLSFLKPGILTGRGKTNQNTG-MTLPAQKCILDA
Chimpanzee
Bonobo
Gorilla
Orangutan
Olive baboon
Macaque
Marmoset
Squirrel monkey
Mouse lemur
Bushbaby
Cow
Sheep
Pig
Horse
Elephant
Giant panda
Dog
Cat
Little brown myotis
Large flying fox
Black flying fox
Squirrel
Guinea pig
Mouse
European rabbit cuniculus
European rabbit algirus
Eastern cottontail Allele1
Eastern cottontail Allele2
Brush rabbit S1 Allele1
Brush rabbit S1 Allele2
Brush rabbit S2
Iberian hare S1
Iberian hare S2 S3 Alelle1
Iberian hare S2 Alelle2
Iberian hare S3 Alelle2
European brown hare

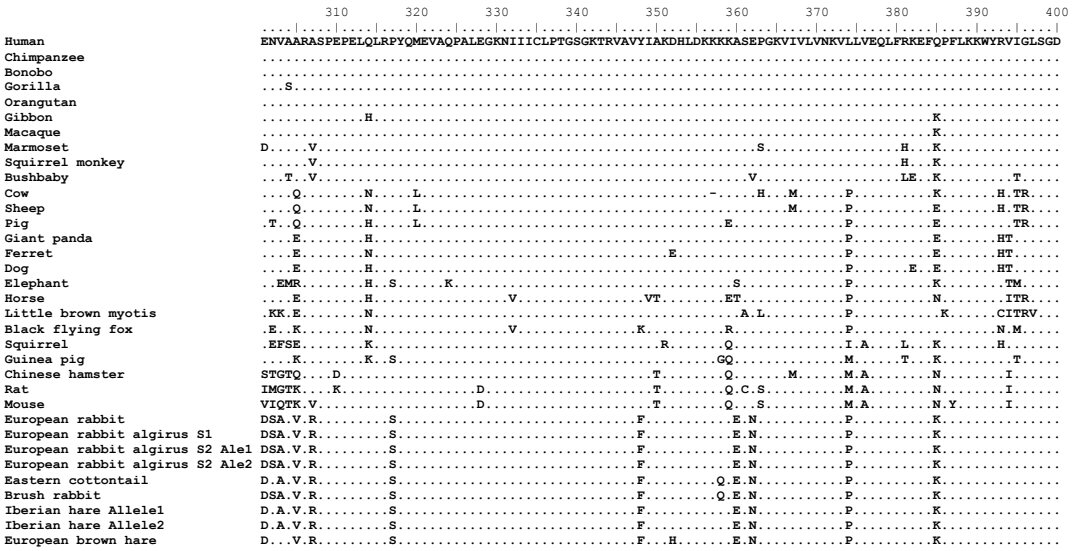

710 720 730 740 750 760 770 780 790 800
Human FKASGDHNILIATSVADEGIDIAQCNLVILYEYVGNVIKMIQTRGRGRARGSKCFLLTSNAGVIEKEQINMYKEKMMNDSILRLQTWDEAVFREKI---L
Chimpanzee
Bonobo
Gorilla
Orangutan
Olive baboon
Macaque
Marmoset
Squirrel monkey
Mouse lemur
Bushbaby
Cow
Sheep
Pig
Horse
Elephant
Giant panda
Dog
Cat
Little brown myotis
Large flying fox
Black flying fox
Squirrel
Guinea pig
Mouse
European rabbit cuniculus
European rabbit algirus
Eastern cottontail Allele1
Eastern cottontail Allele2
Brush rabbit S1 Allele1
Brush rabbit S1 Allele2
Brush rabbit S2
Iberian hare S1
Iberian hare S2 S3 Alelle1
Iberian hare S2 Alelle2
Iberian hare S3 Alelle2
European brown hare

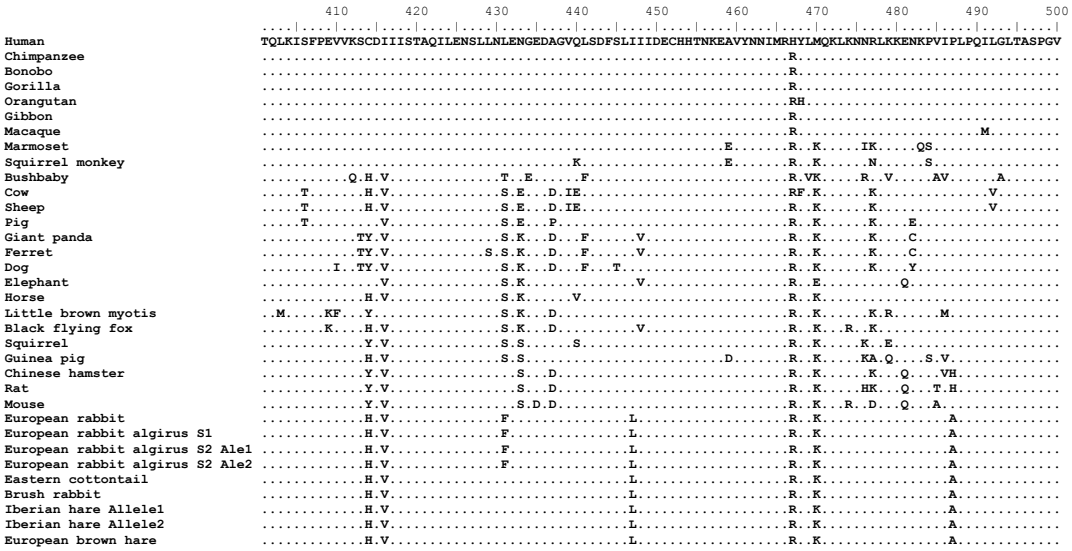
```
                          810       820       830       840       850       860       870       880       890       900
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Human                     HIQTHEKFIRDSQEKPKPVPDKENKKLLCRKCKALACYTADVRVIEECHYTVLGDAFKECFVSRPHPKPKQFSSFEKRAKIFCARQNCSHDWGIHVKYKT
Chimpanzee                ..................................................................................................
Bonobo                    ..................................................................................................
Gorilla                   ..................................................................................................
Orangutan                 ............N.....................I.........................NK.....................................
Olive baboon              .......L..................G.....V.....I...................R.......K.PN.D..E.....................
Macaque                   .......L..................G.....V.....I...................R.......K..N.D..E.............M....
Marmoset                  ...I.....................................F......I.....SQF.....................N.Y..........D..........
Squirrel monkey           ...I..........T.................F......I.....S.....................T...N.H.........RD.............
Mouse lemur               ...I.........G.A...L.............T.............DT....V.....K....K......K.CD.............D...........R.
Bushbaby                  D..I....L....A.....L..K......K...........I....DS....V......R...K......I..........Y....D.G...........
Cow                       Q..IQ..L.....G.V...V..K......G...TF......I..V....F..VR...R....TKL..R..K.G..D.K.......KD.L......M....
Sheep                     Q..VQ..L...N.G.V...V..N......G...TF......I..V....F..VR...R....TKS..S..K.GN...KS......KD.L......MT.R.
Pig                       Q..IR..I...N.G..E.....KT.....K....F......I.MV.N..F..V....R.R...KL.....S.GNI......Y...PD.......Y.R..A
Horse                     Q..I...S....G.A...L..K...........F......I..V.......V.N..R....C.S.....I.G....K...Y...ED.......C.....
Elephant                  I..NQ......NEG.A..E...K...........F....V.I..V..S....V....R.....K.....RTYG....K...H....G.......Y...R.
Giant panda               Q..IQ..L.....G.VER................GF......I..V.D.....V..D.RK.Y..KL.....S.GY...T.....S.ED.............
Dog                       Q..IQ..N.......VEL................F......I..V.......V....TK....KL.....S.GH...........R..G...........
Cat                       Q..IQ..V.....G.VE.................F....V.I..V.......V....RK....KL.....S.GY............P.............I
Little brown myotis       Q..I.........G.AE.AL..K...........F..H...I..........V.........TKA..R..SVGI...K..V.....................
Large flying fox          Q..I.........S.V..IL..............TF......I..V.......V.....K....KL.....NVG....K.........D.C...........
Black flying fox          Q..I.........S.V..IL..............F......I..V.D.....V.....K....KL.....N.G....K.........D.C...........
Squirrel                  S..I.D.......P.S......K......G...T...........V.T.....V..D...R..RK...R..NYGG...K......KES.R..........E
Guinea pig                ...NQ........P.S.F.......R...A...V.......I.....S....I......H...K.....LSYGN...KE.........G.......R...
Mouse                     R..VN..LL....H..Q.............G...NF......I..V.TS...........R..CK......IYDN...K......K.........F.R...
European rabbit cuniculus S...G..I...N.QNF..KL..Q......A.........SEI.MVKDS....V.ET..K...T....R..SCGD.D.IG.......D.......RA.VNA
European rabbit algirus   S...G..I...N.QNF..KL..Q......A.........SEI.MVKDS....V.ET..K...T....R..SCGD.D.IG.......D.......RA.VNA
Eastern cottontail Allele1 C...G......N.QNS..KL..Q......G.........SEI.TVKDS....V.ET..K...T....R..SCGDID.IG.......D.......RA.VNA
Eastern cottontail Allele2 C...G......N.QNS..KL..Q......G.........SEI.TVKDS....V.ET..N...T....R..SCGD.D.IG.......D.......RA.VNA
Brush rabbit S1 Allele1   C...D......N.QNV..KL..Q......G.........SEI.KLKDS....V.ET..K...T....R..SCGD.D.IG.......D.......RA.VNA
Brush rabbit S1 Allele2   C...D......N.QNV..KL..Q......G.........SEI.KLKDS....V.ET..K...T....R..SCGD.D.IG.......D.......RA.VNA
Brush rabbit S2           C...D......N.QNV..KL..Q......G.........SEI.KLKDS....V.ET..K...T....R..SCGD.D.IG.......D.......RA.VNA
Iberian hare S1           C..IG.......QNF..KL..Q......G.........S.I.MLKDS....V.ET..K...T....R..SCGD.D.IG.......D.......RA.VNA
Iberian hare S2 S3 Alelle1 C..IG.......QNF..KL..Q......G.........S.I.MLKDS....V.ET..K...T....R..SCGD.D.IG.......D.......RA.VNA
Iberian hare S2 Alelle2   C..IG.......QNF..KL..Q......G.........S.I.MLKDS....V.ET..K...T....R..SCGD.D.IG.......D.......RA.VNA
Iberian hare S3 Alelle2   C..IG.......QNF..KL..Q......G.........S.I.MLKDS....V.ET..K...T....R..SCGD.D.IG.......D.......RA.VNA
European brown hare       C..IG.......QNF..KL..Q......G.........S.I.MLKDS....V.ET..K...T....R..SCGD.D.IG.......D.......RA.VNA
```

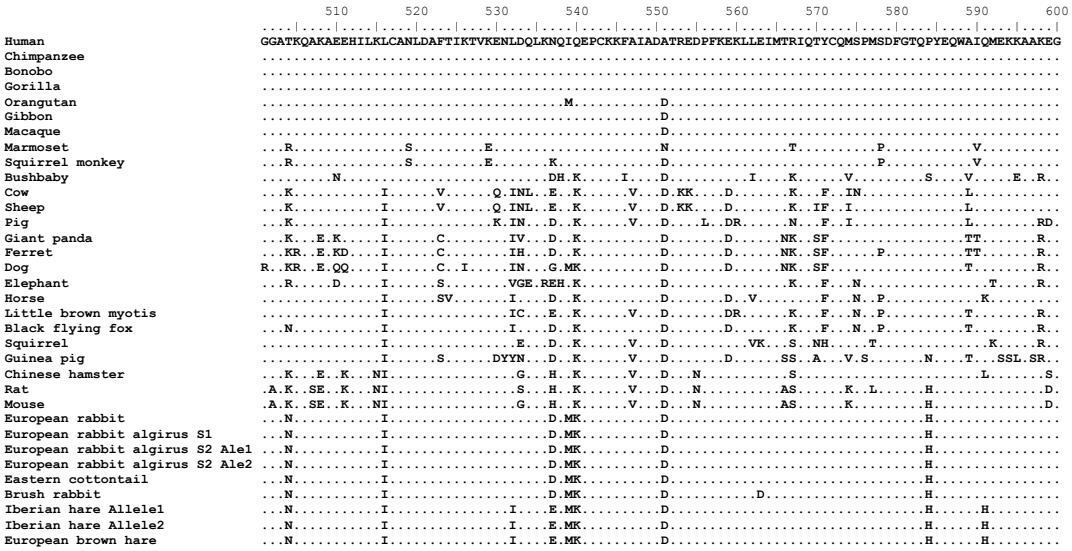
```
                          910       920       930       940       950       960       970       Database ID
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|.
Human                     FEIPVIKIESFVVEDIATGVQTLYSKWKDFHFEKIPFDPAEMSK------------------------------  ENST00000379883
Chimpanzee                ..........................................------------------------------  ENSPTRG00000020844
Bonobo                    ..........................................------------------------------  XM_003830100.1
Gorilla                   ..........................................------------------------------  ENSGGOT00000032102
Orangutan                 ..........................................------------------------------  XM_002819715.1
Olive baboon              .....................................A.....------------------------------  XM_003911615.1
Macaque                   .....................................A.....------------------------------  ENSMMUT00000017446
Marmoset                  ....................................A.....------------------------------  ENSCJAT00000015302
Squirrel monkey           .......M.............................A....TISEPQPSASGNE------------------  XM_003939729.1
Mouse lemur               ..............V.....R.............A....R------------------------------  ENSMICG00000006555
Bushbaby                  ..................A.R.....E.....T..HAV..Y.------------------------------  ENSOGAG00000010536
Cow                       ..............V...A.....A.....N.......A....PWAQDLNLQGVDGLE------------------  ENSBTAG00000003366
Sheep                     ........G......V...A..........N.......A....AGAQDLNLQAMNGLE------------------  XM_004005323.1
Pig                       ..-.F......AL.EY......VHAS....N...LS..A...AGGAQDMGLQGWATLSEGEIGLGLNHGSPVPLLR  ENSSSCG00000030408
Horse                     ..................R.....A..R..N.......A....N------------------------------  ENSECAG00000021989
Elephant                  ..................A.S..A.............A.....------------------------------  ENSLAFG00000005416
Giant panda               ..................A.K..A..R..PLQ.....A.....------------------------------  ENSAMEG00000003766
Dog                       ..................A.K..A.....P......GSP.IPE------------------------------  ENSCAFG00000001807
Cat                       ..................A.K..A.............AK..P.------------------------------  XM_003995540.1
Little brown myotis       ....I.............E.R..A..R......L.........------------------------------  ENSMLUG00000003041
Large flying fox          ....F...................A..R......L...A...P.------------------------------  ENSPVAG00000009207
Black flying fox          ....F.............A....A..R....G.L...A...P.------------------------------  JN031514.1
Squirrel                  .....................R............Q..A..YPNDFRA------------------------------  ENSSTOG00000006994
Guinea pig                ................V..L.KQ.PR..A.P.K..E..AT.....------------------------------  ENSCPOG00000001598
Mouse                     ..............VS...NRH.........R.Q.......V------------------------------  ENSMUST00000037907
European rabbit cuniculus ....L...Q.....TV..K.MCP......N......N.E.L.E------------------------------
European rabbit algirus   ....L...Q.....TV..K.MCP......N......N.E.L.E------------------------------
Eastern cottontail Allele1 ....L...Q.....TV..K.MCP....ALN.Q....N.E.L.E------------------------------
Eastern cottontail Allele2 ....L...Q.....TV..K.MCP......N......N.E.L.E------------------------------
Brush rabbit S1 Allele1   ....L...Q.....TV..K.MCP........QT...N.E.L.E------------------------------
Brush rabbit S1 Allele2   ....L...Q.....TV..K.MCP........QTT..N.E.L.E------------------------------
Brush rabbit S2           ....L...Q.....TV..K.MCP........QTT..N.E.L.E------------------------------
Iberian hare S1           ....L...Q.....TV..K.MCP......N......N.E.L.E------------------------------
Iberian hare S2 S3 Alelle1 ....L...Q.....TV..K.MCP......N......N.E.L.E------------------------------
Iberian hare S2 Alelle2   ....L...Q.....TV..K.MCP......N......N.E.L.E------------------------------
Iberian hare S3 Alelle2   ....L...Q.....TV..K.MCP......N......N.E.L.NH-----------------------------
European brown hare       ....L...Q.....TV..K.MCP......N......N.E.L.E------------------------------
```

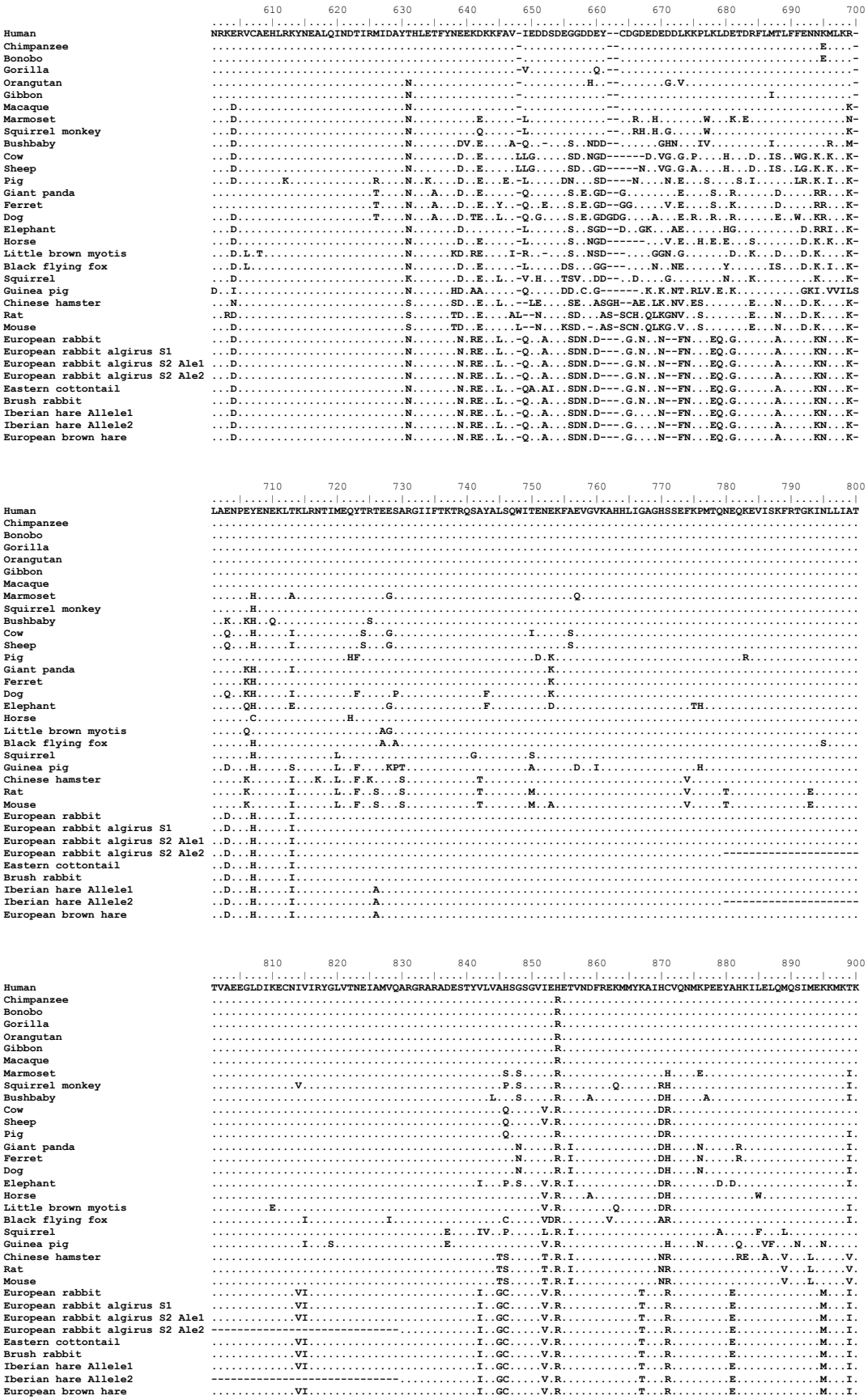
Supplementary Figure S2

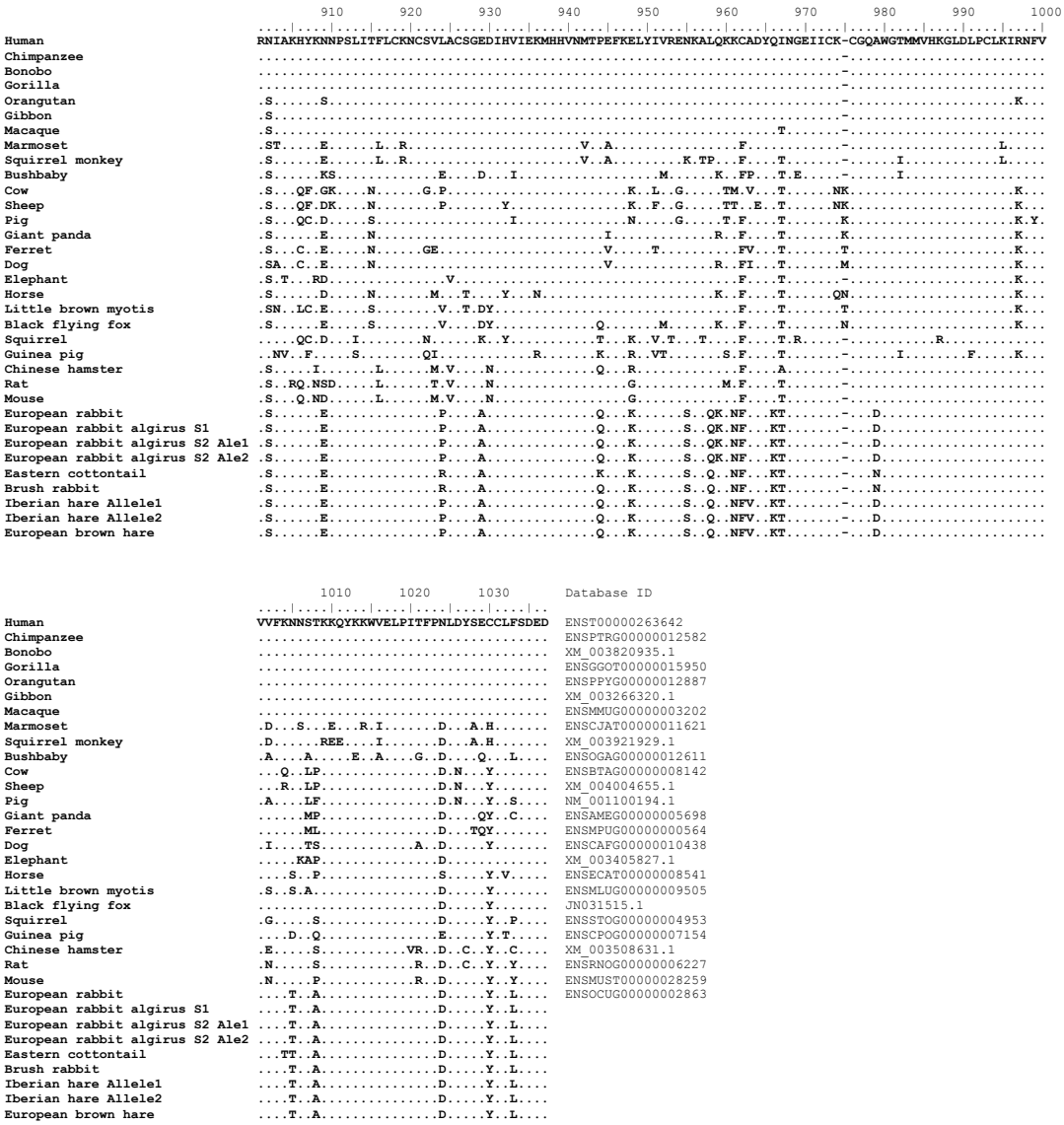



310 320 330 340 350 360 370 380 390 400
Human ENVAARASPEPELQLRPYQMEVAQPALEGKNIIICLPTGSGKTRVAVYIAKDHLDKKKKASEPGKVIVLVNKVLLVEQLFRKEFQPFLKKWYRVIGLSGD
Chimpanzee
Bonobo
Gorilla
Orangutan
Gibbon
Macaque
Marmoset
Squirrel monkey
Bushbaby
Cow
Sheep
Pig
Giant panda
Ferret
Dog
Elephant
Horse
Little brown myotis
Black flying fox
Squirrel
Guinea pig
Chinese hamster
Rat
Mouse
European rabbit
European rabbit algirus S1
European rabbit algirus S2 Ale1
European rabbit algirus S2 Ale2
Eastern cottontail
Brush rabbit
Iberian hare Allele1
Iberian hare Allele2
European brown hare

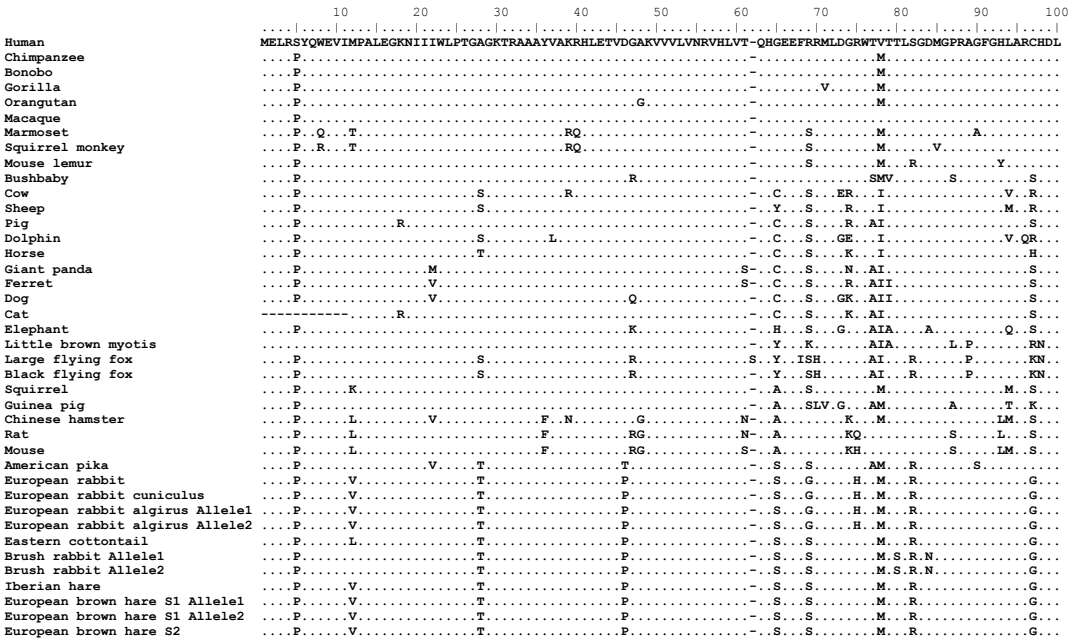

410 420 430 440 450 460 470 480 490 500
Human TQLKISFPEVVKSCDIIISTAQILENSLLNLENGEDAGVQLSDFSLIIIDECHHTNKEAVYNNIMRHYLMQKLKNNRLKKENKPVIPLPQILGLTASPGV
Chimpanzee
Bonobo
Gorilla
Orangutan
Gibbon
Macaque
Marmoset
Squirrel monkey
Bushbaby
Cow
Sheep
Pig
Giant panda
Ferret
Dog
Elephant
Horse
Little brown myotis
Black flying fox
Squirrel
Guinea pig
Chinese hamster
Rat
Mouse
European rabbit
European rabbit algirus S1
European rabbit algirus S2 Ale1
European rabbit algirus S2 Ale2
Eastern cottontail
Brush rabbit
Iberian hare Allele1
Iberian hare Allele2
European brown hare

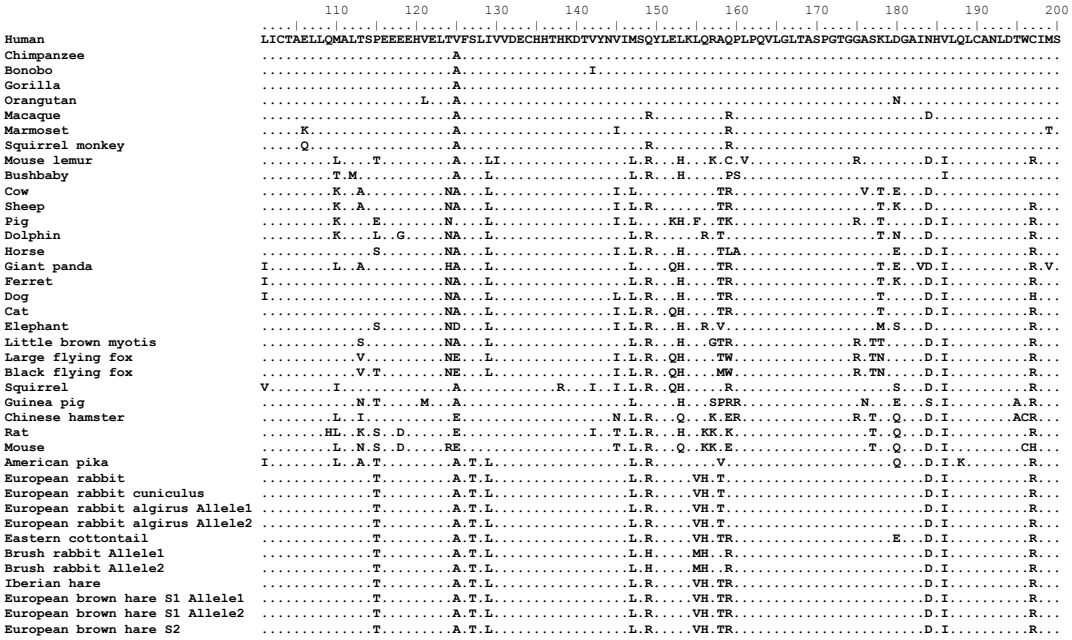

510 520 530 540 550 560 570 580 590 600
Human GGATKQAKAEEHILKLCANLDAFTIKTVKENLDQLKNQIQEPCKKFAIADATREDPFKEKLLEIMTRIQTYCQMSPMSDFGTQPYEQWAIQMEKKAAKEG
Chimpanzee
Bonobo
Gorilla
Orangutan
Gibbon
Macaque
Marmoset
Squirrel monkey
Bushbaby
Cow
Sheep
Pig
Giant panda
Ferret
Dog
Elephant
Horse
Little brown myotis
Black flying fox
Squirrel
Guinea pig
Chinese hamster
Rat
Mouse
European rabbit
European rabbit algirus S1
European rabbit algirus S2 Ale1
European rabbit algirus S2 Ale2
Eastern cottontail
Brush rabbit
Iberian hare Allele1
Iberian hare Allele2
European brown hare

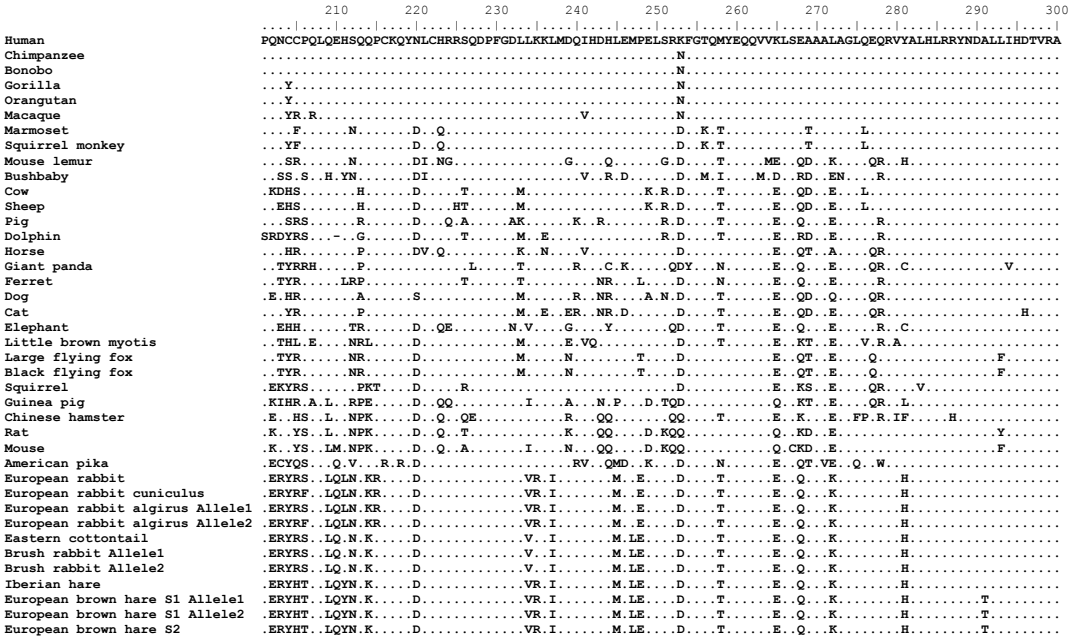

610 620 630 640 650 660 670 680 690 700
Human NRKERVCAEHLRKYNEALQINDTIRMIDAYTHLETFYNEEKDKKFAV-IEDDSDEGGDDEY--CDGDEDEDDLKKPLKLDETDRFLMTLFFENNKMLKR-
Chimpanzee
Bonobo
Gorilla
Orangutan
Gibbon
Macaque
Marmoset
Squirrel monkey
Bushbaby
Cow
Sheep
Pig
Giant panda
Ferret
Dog
Elephant
Horse
Little brown myotis
Black flying fox
Squirrel
Guinea pig
Chinese hamster
Rat
Mouse
European rabbit
European rabbit algirus S1
European rabbit algirus S2 Ale1
European rabbit algirus S2 Ale2
Eastern cottontail
Brush rabbit
Iberian hare Allele1
Iberian hare Allele2
European brown hare
710 720 730 740 750 760 770 780 790 800
Human LAENPEYENEKLTKLRNTIMEQYTRTEESARGIIFTKTRQSAYALSQWITENEKFAEVGVKAHHLIGAGHSSEFKPMTQNEQKEVISKFRTGKINLLIAT
Chimpanzee
Bonobo
Gorilla
Orangutan
Gibbon
Macaque
Marmoset
Squirrel monkey
Bushbaby
Cow
Sheep
Pig
Giant panda
Ferret
Dog
Elephant
Horse
Little brown myotis
Black flying fox
Squirrel
Guinea pig
Chinese hamster
Rat
Mouse
European rabbit
European rabbit algirus S1
European rabbit algirus S2 Ale1
European rabbit algirus S2 Ale2
Eastern cottontail
Brush rabbit
Iberian hare Allele1
Iberian hare Allele2
European brown hare
810 820 830 840 850 860 870 880 890 900
Human TVAEEGLDIKECNIVIRYGLVTNEIAMVQARGRARADESTYVLVAHSGSGVIEHETVNDFREKMMYKAIHCVQNMKPEEYAHKILELQMQSIMEKKMKTK
Chimpanzee
Bonobo
Gorilla
Orangutan
Gibbon
Macaque
Marmoset
Squirrel monkey
Bushbaby
Cow
Sheep
Pig
Giant panda
Ferret
Dog
Elephant
Horse
Little brown myotis
Black flying fox
Squirrel
Guinea pig
Chinese hamster
Rat
Mouse
European rabbit
European rabbit algirus S1
European rabbit algirus S2 Ale1
European rabbit algirus S2 Ale2
Eastern cottontail
Brush rabbit
Iberian hare Allele1
Iberian hare Allele2
European brown hare

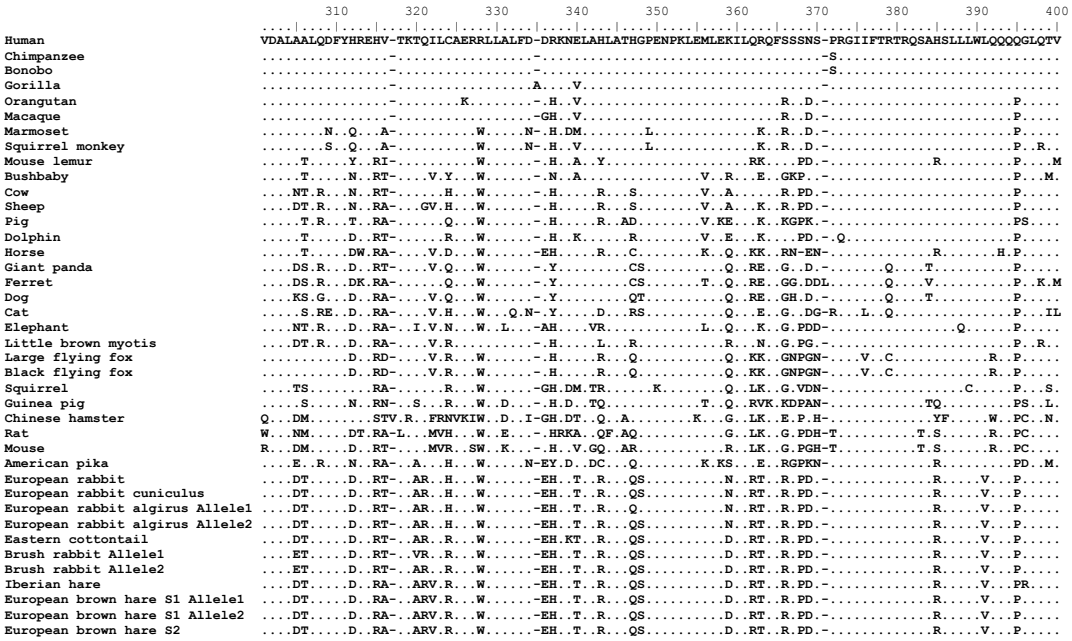

910 920 930 940 950 960 970 980 990 1000
Human RNIAKHYKNNPSLITFLCKNCSVLACSGEDIHVIEKMHHVNMTPEFKELYIVRENKALQKKCADYQINGEIICK-CGQAWGTMMVHKGLDLPCLKIRNFV
Chimpanzee
Bonobo
Gorilla
Orangutan
Gibbon
Macaque
Marmoset
Squirrel monkey
Bushbaby
Cow
Sheep
Pig
Giant panda
Ferret
Dog
Elephant
Horse
Little brown myotis
Black flying fox
Squirrel
Guinea pig
Chinese hamster
Rat
Mouse
European rabbit
European rabbit algirus S1
European rabbit algirus S2 Ale1
European rabbit algirus S2 Ale2
Eastern cottontail
Brush rabbit
Iberian hare Allele1
Iberian hare Allele2
European brown hare
1010 1020 1030 Database ID
Human VVFKNNSTKKQYKKWVELPITFPNLDYSECCLFSDED ENST00000263642
Chimpanzee ENSPTRG00000012582
Bonobo XM_003820935.1
Gorilla ENSGGOT00000015950
Orangutan ENSPPYG00000012887
Gibbon XM_003266320.1
Macaque ENSMMUG00000003202
Marmoset ENSCJAT00000011621
Squirrel monkey XM_003921929.1
Bushbaby ENSOGAG00000012611
Cow ENSBTAG00000008142
Sheep XM_004004655.1
Pig NM_001100194.1
Giant panda ENSAMEG00000005698
Ferret ENSMPUG00000000564
Dog ENSCAFG00000010438
Elephant XM_003405827.1
Horse ENSECAT00000008541
Little brown myotis ENSMLUG00000009505
Black flying fox JN031515.1
Squirrel ENSSTOG00000004953
Guinea pig ENSCPOG00000007154
Chinese hamster XM_003508631.1
Rat ENSRNOG00000006227
Mouse ENSMUST00000028259
European rabbit ENSOCUG00000002863
European rabbit algirus S1
European rabbit algirus S2 Ale1
European rabbit algirus S2 Ale2
Eastern cottontail
Brush rabbit
Iberian hare Allele1
Iberian hare Allele2
European brown hare

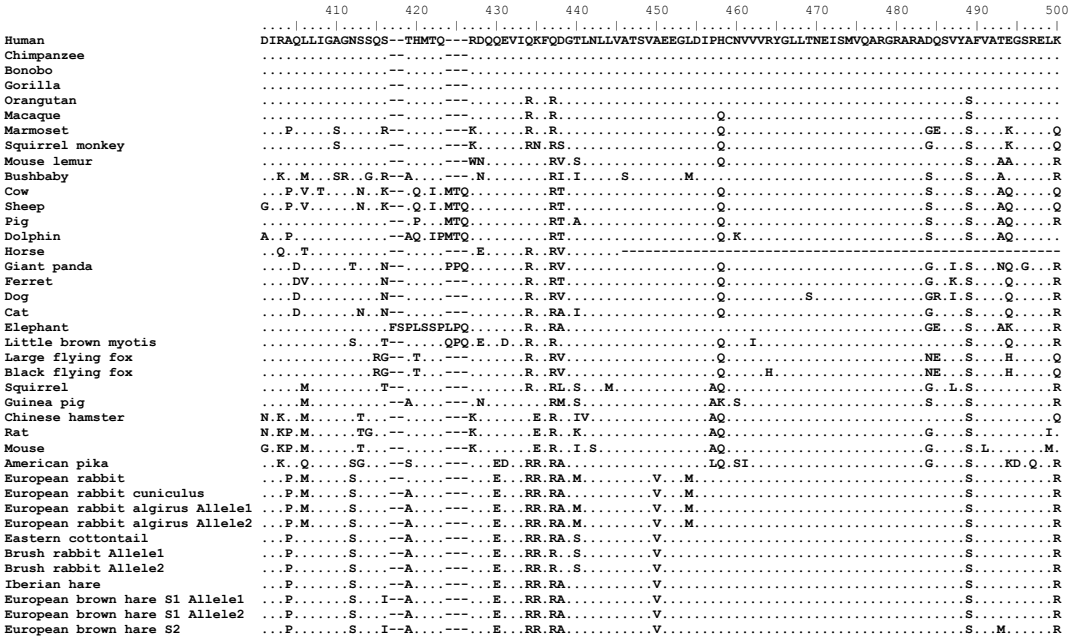
Supplementary Figure S3

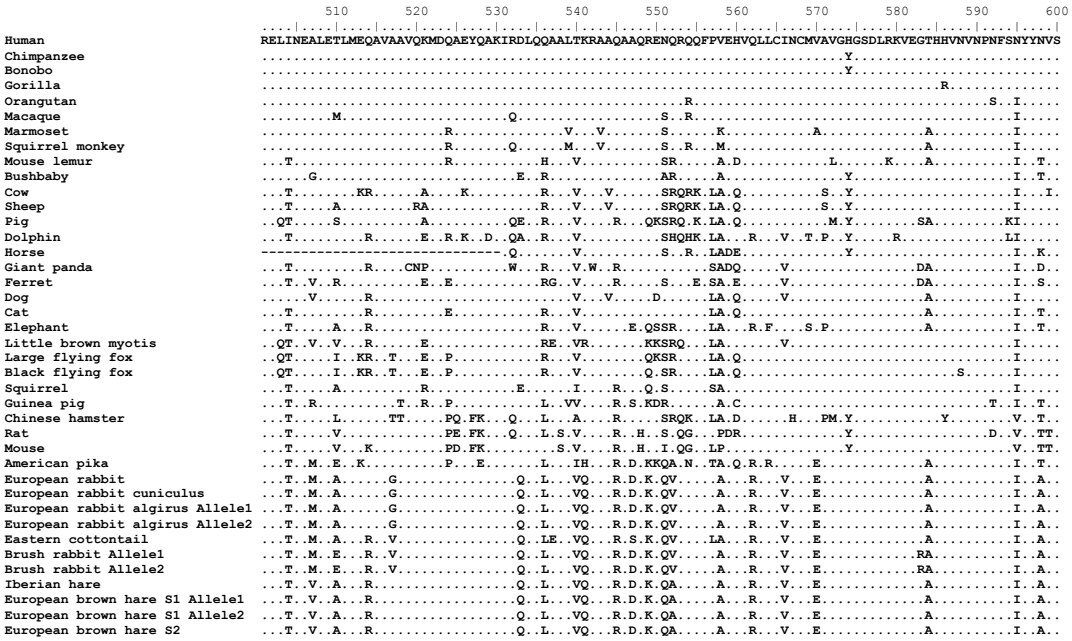





210 220 230 240 250 260 270 280 290 300
Human PQNCCPQLQEHSQQPCKQYNLCHRRSQDPFGDLLKKLMDQIHDHLEMPELSRKFGTQMYEQQVVKLSEAAALAGLQEQRVYALHLRRYNDALLIHDTVRA
Chimpanzee
Bonobo
Gorilla
Orangutan
Macaque
Marmoset
Squirrel monkey
Mouse lemur
Bushbaby
Cow
Sheep
Pig
Dolphin
Horse
Giant panda
Ferret
Dog
Cat
Elephant
Little brown myotis
Large flying fox
Black flying fox
Squirrel
Guinea pig
Chinese hamster
Rat
Mouse
American pika
European rabbit
European rabbit cuniculus
European rabbit algirus Allele1
European rabbit algirus Allele2
Eastern cottontail
Brush rabbit Allele1
Brush rabbit Allele2
Iberian hare
European brown hare S1 Allele1
European brown hare S1 Allele2
European brown hare S2


310 320 330 340 350 360 370 380 390 400
Human VDALAALQDFYHREHV-TKTQILCAERRLLALFD-DRKNELAHLATHGPENPKLEMLEKILQRQFSSSSNS-PRGIIFTRTRQSAHSLLLWLQQQQGLQTV
Chimpanzee
Bonobo
Gorilla
Orangutan
Macaque
Marmoset
Squirrel monkey
Mouse lemur
Bushbaby
Cow
Sheep
Pig
Dolphin
Horse
Giant panda
Ferret
Dog
Cat
Elephant
Little brown myotis
Large flying fox
Black flying fox
Squirrel
Guinea pig
Chinese hamster
Rat
Mouse
American pika
European rabbit
European rabbit cuniculus
European rabbit algirus Allele1
European rabbit algirus Allele2
Eastern cottontail
Brush rabbit Allele1
Brush rabbit Allele2
Iberian hare
European brown hare S1 Allele1
European brown hare S1 Allele2
European brown hare S2


410 420 430 440 450 460 470 480 490 500
Human DIRAQLLIGAGNSSQS--THMTQ---RDQQEVIQKFQDGTLNLLVATSVAEEGLDIPHCNVVVRYGLLTNEISMVQARGRARADQSVYAFVATEGSRELK
Chimpanzee
Bonobo
Gorilla
Orangutan
Macaque
Marmoset
Squirrel monkey
Mouse lemur
Bushbaby
Cow
Sheep
Pig
Dolphin
Horse
Giant panda
Ferret
Dog
Cat
Elephant
Little brown myotis
Large flying fox
Black flying fox
Squirrel
Guinea pig
Chinese hamster
Rat
Mouse
American pika
European rabbit
European rabbit cuniculus
European rabbit algirus Allele1
European rabbit algirus Allele2
Eastern cottontail
Brush rabbit Allele1
Brush rabbit Allele2
Iberian hare
European brown hare S1 Allele1
European brown hare S1 Allele2
European brown hare S2


510 520 530 540 550 560 570 580 590 600
Human RELINEALETLMEQAVAAVQKMDQAEYQAKIRDLQQAALTKRAAQAAQRENQRQQFPVEHVQLLCINCMVAVGHGSDLRKVEGTHHVNVNPNFSNYYNVS
Chimpanzee
Bonobo
Gorilla
Orangutan
Macaque
Marmoset
Squirrel monkey
Mouse lemur
Bushbaby
Cow
Sheep
Pig
Dolphin
Horse
Giant panda
Ferret
Dog
Cat
Elephant
Little brown myotis
Large flying fox
Black flying fox
Squirrel
Guinea pig
Chinese hamster
Rat
Mouse
American pika
European rabbit
European rabbit cuniculus
European rabbit algirus Allele1
European rabbit algirus Allele2
Eastern cottontail
Brush rabbit Allele1
Brush rabbit Allele2
Iberian hare
European brown hare S1 Allele1
European brown hare S1 Allele2
European brown hare S2

```
                                610       620       630       640       650       660       670       680
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|..
Human                  RDPVVINKVFKDWKPGGVISCRNCGEVWGLQMIYKSVKLPVLKVRSMLLETPQGRIQAKKWSRVPFSVPDFDFLQHCAENLSDLSLD
Chimpanzee             ...............................................................................Q........
Bonobo                 ...............................................................................Q........
Gorilla                ................S..............I.....................................................Q........
Orangutan              .....................................A.....................................Q........
Macaque                ....................I..............A.....................................Q........
Marmoset               KE.....R.....................I........A.........Q.................P...........Q..M.....
Squirrel monkey        .E..........................I........A.........Q.................P...........Q..T.....
Mouse lemur            QE.....R.........A.R......G..M.I........A.........Q.............C...L...........QS.......
Bushbaby               QE....DR.........A.H......D.......RTE.....R......H..R.............P....N......Q.......G
Cow                    KQ..D..RS....R...A........A....I.............G..........VR.........T.....YV.-Y..G.AG....
Sheep                  KQ..D.SRS....R...A........A....I.............G..........VR.........T.....YV.-YT.G.AG....
Pig                    QE....DR.....R.....R......S..M.I..............V.....N....V....C...P.....YT.Y.T.S.T.....
Dolphin                QE.....R.....R............G-....N..............L.........V.S.........T.....YV....QS.A.F...
Horse                  PKA...DRE.........AV.......S..M............I..........R..V......H...P....NY.....QS........
Giant panda            .GA...DRI....R...T.H......A..............A...C..P....R..V..........P.....Y.....Q..A.....
Ferret                 .GA...DRT........T.H......A..............A.R......Q..R..V..........P.....Y....TQ..A....E
Dog                    .G....DRT....R...T.H......A..............A...........R..V..........P.....Y.....Q..A.....
Cat                    .G....DRT........A.H......A..............A..............V..........L.....Y....TQ....FT..
Elephant               QA....DRD.R..S...A........P....I..........R.........R..L......C...L.....YV....Q..R....N
Little brown myotis    EK..A.DR.....V.......K....L..............A....A...Q..R.........S...Q.....YVL...QA.......
Large flying fox       QT....D...R..R...T...K....L..........N......H..........V.P........P.....V....Q...E...A
Black flying fox       QK....D......R...T...K....L..........N......H..........V.P........P....N.V....QH..E...A
Squirrel               .E....DR.....R...T.....................N..A.RI..V.....R..............P.A..NI.EDS.QG......E
Guinea pig             .E....DR.........A........S..........T........L......R...V........P.D....V....Q..A....N
Chinese hamster        QK..............T...S..............T.....I........H....P..........H...I..D..PS........
Rat                    QN..............I.R.S.......F.......T.....I........H.K..................I.RD.TQS.......
Mouse                  QN...........R...T.R.S.......F.......T.....IG.I.....R.K............I.V..I..D.TQS..E....
American pika          .....MDRE.R..............P....I....QN..............R......Q......P..E..LV...GLS.E..L..
European rabbit        ....P..RE........T.R......S.........................R..............P.A........TQS.......
European rabbit cuniculus ....P..RE........T........S.........................R..............P.A........TQS.......
European rabbit algirus Allele1 ....P..RE........T........S.........................R..............P.A..N.....TQS.......
European rabbit algirus Allele2 ....P..RE........T........S.........................R..............P.A........TQS.......
Eastern cottontail     ....P.DRE........T........S.........................R..............P..........TQS.......
Brush rabbit Allele1   ....P..RE........T........S........CL...A..........R..............P..........TQS.P.....
Brush rabbit Allele2   ....P..RE........T........S........CL...A..........R..............P..........TQS.P.....
Iberian hare           ....P..RE........T........S.........................R..............P..........TQS.......
European brown hare S1 Allele1 .E..P..RE........A........S.........................R..............P..........TQS.......
European brown hare S1 Allele2 .E..P..RE........A........S.........................R..............P..........TQS.......
European brown hare S2 .E..P..RE........A........S.........................R..............P..........TQS.......
```

	Database ID
Human	ENST00000251642
Chimpanzee	ENSPTRG00000009191
Bonobo	XM_003813868.1
Gorilla	ENSGGOG00000005458
Orangutan	NM_001131127.1
Macaque	ENSMMUG00000014755
Marmoset	ENSCJAT00000028952
Squirrel monkey	XM_003942770.1
Mouse lemur	ENSMICG00000011933
Bushbaby	ENSOGAG00000025321
Cow	ENSBTAT00000065738
Sheep	XM_004012918.1
Pig	ENSSSCT00000018960
Dolphin	ENSTTRG00000006650
Horse	XM_001495212.2
Giant panda	ENSAMEG00000005941
Ferret	ENSMPUG00000010568
Dog	ENSCAFG00000015720
Cat	XM_003996876.1
Elephant	ENSLAFG00000027301
Little brown myotis	ENSMLUG00000012001
Large flying fox	ENSPVAG00000003855
Black flying fox	JN031516.1
Squirrel	ENSSTOG00000006524
Guinea pig	ENSCPOG00000023721
Chinese hamster	XM_003504640.1
Rat	ENSRNOG00000018247
Mouse	ENSMUST00000017974
American pika	ENSOPRG00000000753
European rabbit	ENSOCUG00000013278

Chapter 5

Innate anti-viral factors – Genetic characterization of host restriction factor TRIM5α in Leporidae genera

Paper 5

Study of *Sylvilagus* rabbit TRIM5α species-specific domain: how ancient endoviruses could have shaped the anti-viral repertoire in Lagomorpha

A. Lemos de Matos, W. van der Loo, H. Areal, D. K. Lanning & P. J. Esteves

1. Summary

Since the first report of the anti-retroviral restriction factor TRIM5α in primates, several orthologs in other mammals have been described. Recent studies suggest that leporid retroviruses like RELIK, the first reported endogenous lentivirus ever, may have imposed positive selection in TRIM5α orthologs of the European rabbit and European brown hare. Considering that RELIK must already have been present in a common ancestor of the leporid genera *Lepus*, *Sylvilagus* and *Oryctolagus*, we extended the study of evolutionary patterns of TRIM5α to other members of the Leporidae family, particularly to the genus *Sylvilagus*. Therefore, we obtained the *TRIM5* nucleotide sequences of additional subspecies and species of the three leporid genera. We also compared lagomorph TRIM5α deduced protein sequences and established *TRIM5* gene and TRIM5α protein phylogenies.

The deduced protein sequence of Iberian hare TRIM5α was 89% identical to European rabbit TRIM5α, although high divergence was observed at the PRYSPRY v1 region between rabbit and the identified alleles from this hare species (allele 1: 50% divergence; allele 2: 53% divergence). A high identity was expected between the *Sylvilagus* and *Oryctolagus* TRIM5α proteins and, in fact, the *Sylvilagus* TRIM5α was 91% identical to the *Oryctolagus* protein. Nevertheless, the PRYSPRY v1 region was only 50% similar between these genera. Selection analysis of Lagomorpha TRIM5α proteins identified 25 positively-selected codons, 11 of which are located in the

PRYSPRY v1 region, responsible for species-specific differences in viral capsid recognition.

By extending Lagomorpha TRIM5α studies to an additional genus known to bear RELIK, we verified that the divergent species-specific pattern observed between the *Oryctolagus* and *Lepus* PRYSPRY-domains is also present in *Sylvilagus* TRIM5α. This work is one of the first known studies that compare the evolution of the anti-retroviral restriction factor TRIM5α in different mammalian groups, Lagomorpha and Primates.

2. Introduction

Retroviruses are RNA viruses that, when infecting a host cell, produce a viral reverse transcriptase and a viral integrase that make a DNA copy of the viral genome and integrate it into the host genome, respectively [1]. The family Retroviridae comprises a diverse range of animal viruses, including the viral genus *Lentivirus*. Lentiviruses have been isolated from primates, domestic and wild felids, and a variety of domestic ungulates (goat, sheep, cattle and horse) [2]. Until recently, all known lentiviruses were classified as exogenous (transmitted horizontally from host to host) [3]. However, in 2007, Katzourakis and colleagues [4] reported the first endogenous lentivirus identified in any species, the rabbit endogenous lentivirus type K (RELIK), present in the genome of the European rabbit (*Oryctolagus cuniculus*). RELIK has subsequently been reported in other leporid genera (*Lepus*, *Sylvilagus* and *Bunolagus*), establishing it as at least 12 million years (My) old [5, 6]. These striking observations demonstrate that lentiviruses are more widespread than previously thought, extending the host range to a different mammalian order, and demonstrate that lentiviruses can be endogenized [4-6].

The intense study of lentiviruses in the past 30 years, especially of human immunodeficiency viruses (HIV-1 and HIV-2), has been more recently accompanied by the study of anti-retroviral restriction factors, like the TRIM5α protein, one of the members of TRIM family [7-11]. TRIM proteins contain three domains, which together constitute the canonical TRIpartite Motif, including an N-terminal RING domain, one or two B-Box domains and a long Coiled-Coil (CC) domain [9-11]. TRIM5, like most TRIM proteins, also contains a C-terminal PRYSPRY domain, composed of four "variable loops" [9-11]. TRIM5α is the largest isoform encoded by the *TRIM5* gene and restricts infection by HIV-1 and other retroviruses, dependent on a species-specific sequence variation in the PRYSPRY domain, upon entry into the host cell cytoplasm and prior to reverse transcription [7, 8]. Each TRIM5α domain plays distinct roles in its anti-viral restriction activity. The RING domain has been shown to confer E3 ubiquitin ligase activity crucial for anti-HIV restriction [12,13]. The B-box 2 domain influences

recognition of the viral capsid by the C-terminal PRYSPRY domain [13-15]. The CC domain plays an important role in the restriction of viral infectivity and it is required for trimerization [11, 16]. Particularly for HIV-1 and N-tropic murine leukemia virus (N-MLV) retroviruses, restriction specificity has been mapped to the PRYSPRY domain for HIV-1, and N-MLV restriction specificity is determined by both the CC and PRYSPRY domains [9, 17, 18]. Human TRIM5α is not effective against HIV-1 but does inhibit N-MLV, while rhesus monkey TRIM5α restricts both [19-22]. However, a single amino acid change (R332P) in the human TRIM5α PRYSPRY domain causes it to behave like rhesus TRIM5α with regard to HIV-1 restriction [17, 23]. The PRYSPRY domain binds to the viral capsid, and the domain sequence variation determines the restriction specificity [17, 18, 24-26]. Recently, evidence from several studies began elucidating the detailed mechanism of TRIM5α activity. Also, additional activities linked to viral restriction have been described, including a role in signal transduction, the promotion of innate immune signaling and recognition of the retroviral capsid lattice [27, 28]. It has been suggested that direct binding of TRIM5α to the viral capsid leads to disruption of specific inter-hexamer interfaces, causing structural damage to the capsid. TRIM5α spontaneously forms a hexagonal lattice complementary to the capsid lattice, a molecular signature of retroviruses, which greatly stimulates TRIM5α lattice formation [29, 30].

Evolutionary studies of primate TRIM5α revealed a high ratio of non-synonymous to synonymous changes in the PRYSPRY domain [25, 31-33]. The distribution of positively selected residues is not random, but falls in a very tight cluster at the beginning of the domain in a 13 amino acid "patch", essential for retroviral restriction and responsible, in part, for the species-specific restriction activity [31]. The same domain has also undergone length variation and segmental duplications in different primate lineages [25]. However, polymorphisms found in the TRIM5α coding sequence for multiple individuals from two divergent lineages of Old World monkeys (rhesus macaque and sooty mangabey), indicated that specificity varies not only between different species but also within species [34]. Despite the geographic separation and the divergence time (> 8My), both species presented a highly similar pattern of polymorphisms, which constitutes compelling evidence for long-term balancing selection at the *TRIM5* locus [34]. Similar evidence of selection has been recently reported for the first intron of human *TRIM5* gene, which may affect transcription factor-binding sites and *TRIM5* transcriptional activity [35].

Although evolutionary and functional studies of TRIM5α anti-retroviral restriction activity have primarily focused on the primate lineage, *TRIM5* orthologs have been

reported in other mammalian genomes, e.g. mouse, rat, cow and European rabbit [36-39]. Active TRIM5α was identified in the European rabbit and the ability to restrict the replication of multiple unrelated retroviruses was also described [38]. Besides this leporid, Fletcher and co-workers [40] reported the restriction of divergent retroviruses by European brown hare (*Lepus europaeus*) TRIM5α and also significant differences between both leporids' TRIM5α PRYSPRY domains. These authors suggested that retroviruses like RELIK may have driven the speciation of the Old World rabbit and hare TRIM5α orthologs. The order Lagomorpha is divided into two families, Ochotonidae and Leporidae, which diverged around 40 My ago [41-43]. Ochotonidae contains only one extant genus, *Ochotona*, while the family Leporidae includes 11 genera where, *Lepus*, *Sylvilagus* and *Oryctolagus*, the most well-studied leporid genera, diverged around 12 My ago [41, 42]. It has been suggested that the global development of temperate grasslands (7 to 5 My ago) and the formation of the west Antarctic ice sheet (6.5 My ago) promoted the development of land bridges and consequent dispersal of the genus *Lepus* from North America through Asia and into Africa [41]. The New World *Sylvilagus* lineage initially remained in North America from which it more recently colonized South America [41]. The genus *Oryctolagus* is the only leporid genus native to Europe and consists of two subspecies, *O. cuniculus cuniculus* and *O. cuniculus algirus*, which diverged around 2 My ago. The subspecies *O. c. algirus* is restricted to the southwest region of the Iberian Peninsula and a few Atlantic islands, whereas *O. c. cuniculus* has essentially a man-made worldwide distribution and includes all domestic breeds [44].

As previously suggested, considering that RELIK must already have been present in a common ancestor of these leporid genera, the hypothesis that lentiviruses might have been the driving force of the leporid TRIM5α conserved anti-retroviral activity can be challenged by extending the study of evolutionary patterns of TRIM5α to other members of the Leporidae family, particularly to the New World *Sylvilagus* lineage. Therefore, in this study we have examined the *TRIM5* gene of additional subspecies and species of the three leporid genera: European rabbit subspecies (*Oryctolagus cuniculus algirus*), the Old World Iberian hare (*Lepus granatensis*) and European brown hare, and the New World brush rabbit (*Sylvilagus bachmani*).

3. Materials and Methods

3.1. Samples, RNA extraction and cDNA synthesis

Liver samples from one European rabbit (subspecies *Oryctolagus cuniculus algirus*; Orcual), two European brown hares (*Lepus europaeus*; Leeu) and two Iberian hares (*L.*

granatensis; Legr) were supplied by CIBIO, Vairão, Portugal. In addition, two brush rabbit (*Sylvilagus bachmani*; Syba) spleen samples were provided by the Blue Oak Ranch Reserve, University of California, USA. During this study, no experimental research on animals was conducted.

Total RNA was prepared using the guanidinium thiocyanate-phenol-chloroform extraction method (TRIzol) according to manufacturer's instructions (Molecular Research Center, Inc., Cincinnati, OH, USA). First-strand cDNA was prepared from 5 µg of total RNA, using oligo(dT) primers [45] and the SuperScript™ III First-Strand Synthesis System (Invitrogen, Carlsbad, CA, USA).

3.2. *TRIM5* amplification, cloning and sequencing

The sequence of the European rabbit subspecies *O. c. cuniculus* (Orcucu) *TRIM5* used in this study was taken from GenBank (accession number NM_001105673) [38]. A previously reported [40] American pika (*Ochotona princeps*; Ocpr) *TRIM5* nucleotide sequence retrieved from the whole genome shotgun (WGS) project (contig132533, Locus AAYZ01132534) was also included.

PCR primers were designed from the available sequence for European rabbit *TRIM5* cDNA (Forward 5'-TGTCTTGCAGAAATCTGTGAGCAAAAG-3'and Reverse 5'-AAGAGATGTACCCCAGGGTAAGAG-3'), generating an approximately 1.5 kb PCR product corresponding to the full-length CDS. The PCR thermal profile used was the following: initial denaturation (98 ºC for 30 s); 40 cycles of denaturation (98 ºC for 10 s), annealing (60 ºC for 30 s) and extension (72 ºC for 1 min); and a final extension (72 ºC for 10 min). Phusion® High-Fidelity DNA Polymerase (Finnzymes, Espoo, Finland) was used. Finally, an additional extension step (72 ºC for 10 min) with Taq polymerase (GoTaq, Promega; Madison, WI, USA) was performed.

The PCR products were cloned into the pGEM-T Easy vector (Promega, Madison, WI, USA). At least seven independent clones were sequenced per allele. Sequencing was performed with an ABI PRISM 3130 Genetic Analyser (PE Applied Biosystems), following the ABI PRISM BigDye Terminator Cycle sequencing protocol. PCR products were sequenced in both directions and also with an internal primer (5' CCAACAGGAGATAACTTCCTGGAA 3'). Nucleotide sequence data obtained in this study have been submitted to GenBank and have been assigned the following accession numbers: JN541226, JN541227, JN541228, JN541229 and JN541230.

3.3. Phylogenetic analyses

In order to infer the Lagomorpha TRIM5α phylogeny, based on nucleotide and deduced amino acid sequences, a Maximum Likelihood method implemented on GARLI 1.0 (Genetic Algorithm for Rapid Likelihood Inference) [46] was used. The analyses were performed with 1,000,000 generations and 1,000 bootstrap searches. The Model TIM3 +G for nucleotide substitution estimation was used as indicated by the jModelTest 0.1.1 [47]. The JTT [48] mutation model applied to the amino acid deduced sequences was used with a rate variation among sites with 4 rate categories (+G), as indicated by the program ProtTest 2.4 [49-51].

3.4. Codon-based analysis of positive selection

Most proteins appear to be under strong purifying selection most of the time, whereas positive selection is relegated to small regions of the molecule, meaning that structural and functional domains are likely to evolve at different rates [52]. A common approach to detect selective pressures involves estimating the rates of non-synonymous (d_N) and synonymous (d_S) substitutions [53]. The random effect likelihood (REL) model proposed by Kosakovsky and colleagues [53] and implemented in the Datamonkey web server [54] was used to identify TRIM5α codons under positive selection. REL involves fitting a distribution of substitution rates across sites and then inferring the rate for individual sites [53]. The REL method under the MG94xHKY85 model of evolution was used. Normalized posterior mean of the d_N-d_S difference and the Bayesian posterior probability for positive selection (d_N>d_S) for each codon position were obtained. A Bayes factor of greater than 50 suggests that a site is positively selected.

Estimation of leporid TRIM5α synonymous and non-synonymous substitution rates, using the Nei-Gojobori method [55], was performed on MEGA5 [56].

3.5. Sliding-window analysis

An alternative approach to determine nucleotide substitution rate variation among different genomic regions is to plot differences as averages by sliding a window along a sequence alignment [57]. A sliding-window analysis was performed using DnaSP version 5.10 [58]. A window length of 90 nucleotides and a step size of 10 were chosen for this analysis. Nucleotide replacements per site between European rabbit/European brown hare, European rabbit/Iberian hare, European rabbit/brush rabbit, European rabbit/American pika, human/chimpanzee and human/rhesus monkey *TRIM5* were analyzed.

4. Results and Discussion

4.1. TRIM5α divergence and phylogeny in Lagomorpha

In this study, all the deduced TRIM5α protein sequences obtained from leporids were aligned and compared to that previously described for the European rabbit (subspecies *O. cuniculus cuniculus*) [38] (Figure 1). To more completely assess the *TRIM5* gene in the Lagomorpha order, we also included American pika (*Ochotona princeps*) *TRIM5* nucleotide sequence retrieved from the whole genome shotgun (WGS) project (contig132533, Locus AAYZ01132534). During this work, we were able to bridge the American pika protein sequence gaps previously reported [40] (Figure 1).

The 2 My divergence between European rabbit subspecies apparently allowed the accumulation of one private residue in the *O. c. algirus* PRYSPRY v1 region (351H), different from *O. c. cuniculus* (Figure 1 and Figure 2). Amino acid position 327 was described as being polymorphic in *O. c. cuniculus* [40], but *O. c. algirus* clones presented the same residue (327Q). In addition, two other polymorphisms were observed in *O. c. algirus* clones, one in the BB2 domain (98R/C) and the other in the CC domain (148N/K) (GenBank accession numbers JN541226 and JN541227).

The sequencing of European brown hare *TRIM5* confirmed the previously reported findings [40]. However, our Lagomorpha TRIM5α alignment was somewhat different from that of Fletcher and colleagues [40] due to the inclusion of new species and a more complete American pika protein sequence. Out of the 30 PRYSPRY v1 region residues, we identified 15 differing positions between the European brown hare and the European rabbit, resulting in 50% identity among v1 regions (Figure 1 and Figure 3). *TRIM5* from another species of hare, Iberian hare, was also sequenced in this study. With fourteen clones from two individuals, we identified two alleles differing at three positions, each of them located within the PRYSPRY variable region (332F/L, 391I/T and 427C/S) (GenBank accession numbers JN541228 and JN541229) (Figure 1 and Figure 2). As expected, the overall identity between the TRIM5α protein sequence of European rabbit and the TRIM5α deduced protein sequence of the Iberian hare (89%) was the same as that previously reported for the European brown hare (89%) [40] which contrasted with the high divergence at the PRYSPRY v1 region (allele 1: 50% divergence; allele 2: 53% divergence) (Figure 3). The six amino acid positions previously identified as being polymorphic residues in the European brown hare [40] apparently were not polymorphic in the Iberian hare TRIM5α sequenced clones. In fact,

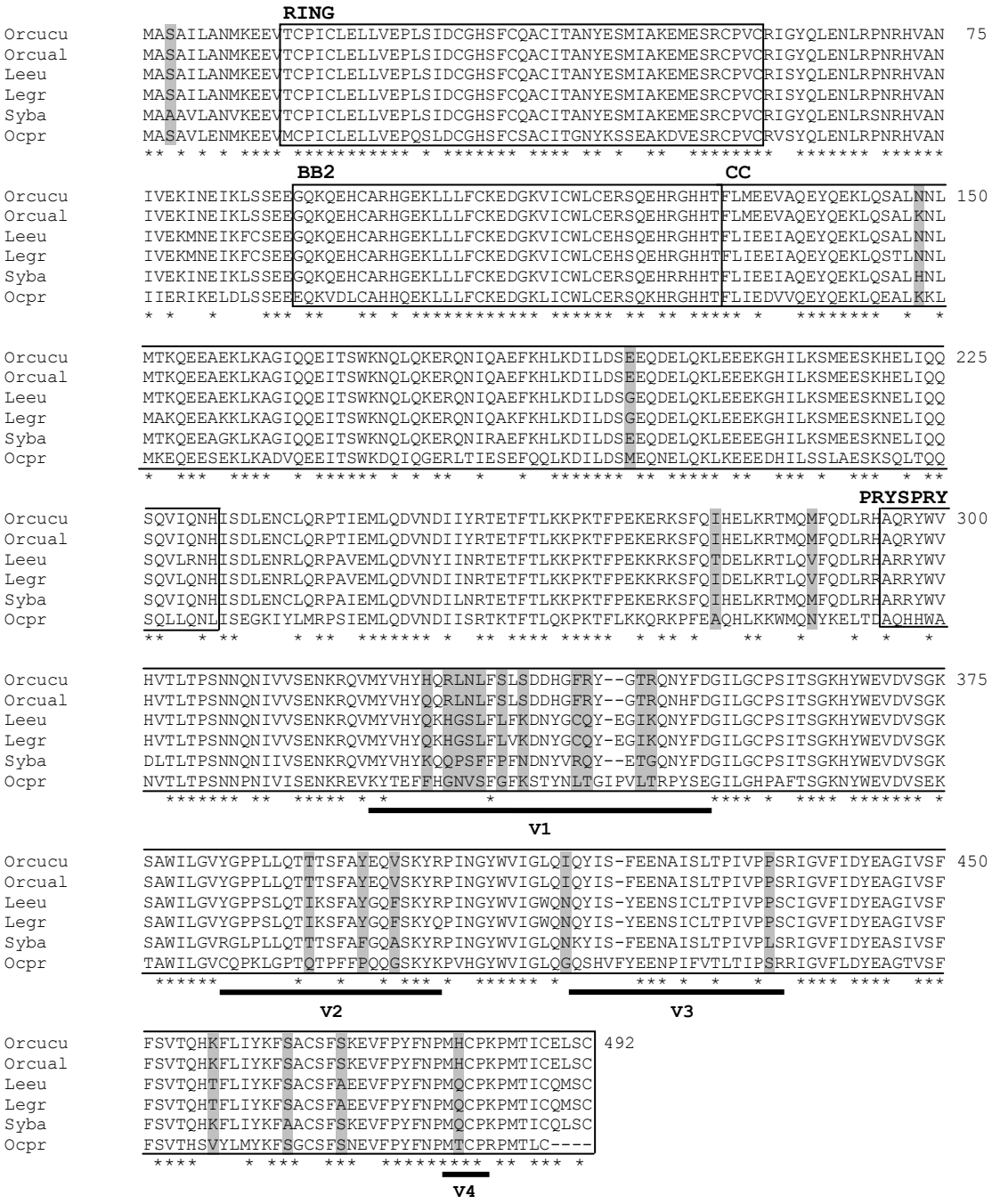


Figure 1 - Lagomorpha TRIM5α deduced protein sequences.

Alignment of TRIM5α deduced protein sequences from European rabbit, *Oryctolagus cuniculus cuniculus* (Orcucu) and *Oryctolagus cuniculus algirus* (Orcual) subspecies, European brown hare (Leeu – allele 1), Iberian hare (Legr – allele 1), brush rabbit (Syba) and American pika (Ocpr). Only allele 1 for Iberian hare is represented, as all the differences between both alleles are reported in the main text. European brown hare sequenced alleles were similar to those previously reported (Genbank accession numbers HM768824, HM768825) [40]. RING domain, B-box type 2 (BB2) domain, Coiled Coil (CC) domain and PRYSPRY domain, with its variable regions (v1, v2, v3 and v4), are indicated. Positively-selected codon positions are shaded; asterisk (*), identical residue between all species.

all the Iberian hare clones presented the same residues as the European brown hare allele 2 in these six positions (146T, 152A, 158K, 230L, 335V and 403Q). However, four positions of all the Iberian hare TRIM5α sequenced clones (185K, 253D, 279I and 294R) differed from both European brown hare alleles (185E, 253Y, 279T and 294H). None of them occurred in the PRYSPRY v1 region, which was identical among Lepus species, except for two polymorphic positions, one in the European brown hare allele 1 (335F) and one in the Iberian hare allele 2 (332F).

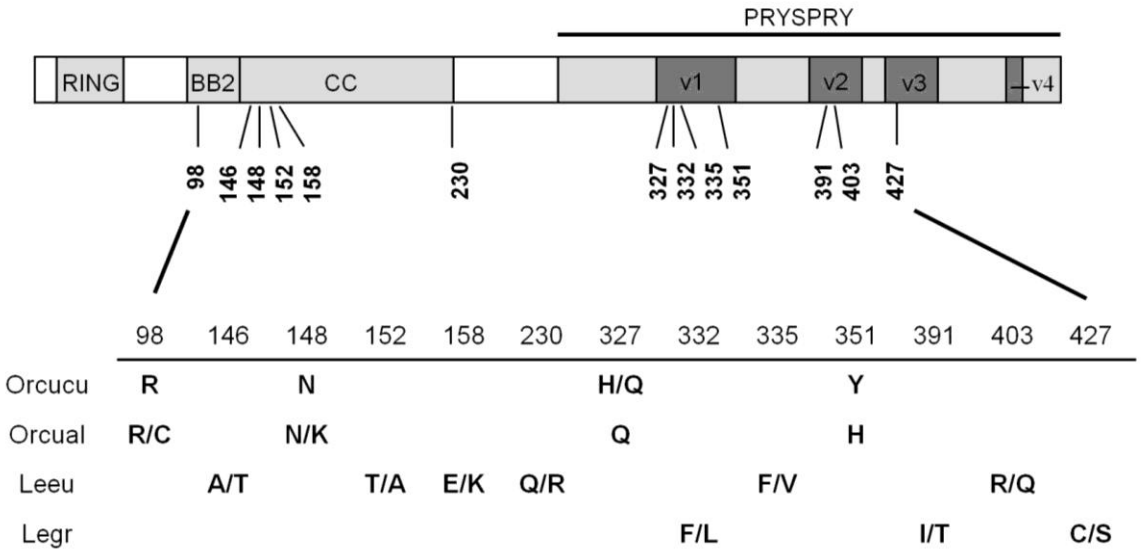


Figure 2 - Schematic representation of the polymorphisms within European rabbit subspecies and *Lepus* species TRIM5α.

Polymorphic sites between the European rabbit subspecies *Oryctolagus cuniculus cuniculus* (Orcucu) and *Oryctolagus cuniculus algirus* (Orcual) TRIM5α are represented. The polymorphic sites from two alleles for each *Lepus* species, European brown hare (Leeu) and Iberian hare (Legr), are also identified. Residues are numbered as in Figure 1.

The proposed leporid phylogeny estimates that the *Lepus* lineage diverged 12.80 My ago from the *Oryctolagus/Sylvilagus* clade, and that the *Sylvilagus* and *Oryctolagus* genera diverged 10 My ago [41, 42]. To extend the study of leporid TRIM5α, we determined the *TRIM5* nucleotide sequence from brush rabbit (GenBank accession number JN541230). As the *Sylvilagus* and *Oryctolagus* genera are closely related, a higher identity between their TRIM5α proteins was expected and, indeed, the brush rabbit TRIM5α deduced protein sequence was 91% identical to the European rabbit protein (Figure 1). However, such increase in similarity was not observed at the

PRYSPRY v1 region (50% similar between these species; Figure 3). From the seven clones obtained for this species, no polymorphisms were observed.

The high divergence obtained in the PRYSPRY v1 region could be explained by gene conversion with adjacent genes. Gene conversion has been reported in other mammalian genes. For example, in leporids, a gene conversion event was observed between the two chromosomally adjacent genes *CCR2* and *CCR5*, where the sequence motif $_{194}$QTLKMT$_{199}$ of the CCR5 protein was replaced by the HTIMRN motif which is characteristic of CCR2 [59, 60]. In the present study, none of the chromosomally adjacent genes showed clear evidence of gene conversion with *TRIM5*, making this explanation unlikely. Furthermore, no significant BLAST matches were obtained during searches of the mammalian NCBI database. No evidence of recombination between alleles was observed.

To obtain evidence of within-species variation, the number of individuals *per* species should be significant. In this study, the number of individuals was low, limiting the accuracy of the observed polymorphisms. However, it cannot be ruled out that some of the sites that appeared to be variable between species are polymorphisms present within species, especially when considering the two closely related *Lepus* species. A case of *trans*-species polymorphism was reported in a study of the evolution of the immunoglobulin heavy chain variable region in *Oryctolagus* and *Lepus* [61]. In light of the previously described long-term balancing selection on primate TRIM5α [34, 35], this scenario cannot be excluded in leporids.

Lagomorpha TRIM5α phylogenetic trees, based on nucleotide, including all described alleles, and amino acid deduced sequences, were obtained with the Maximum Likelihood method (Figure 4). TRIM5α nucleotide and amino acid sequences of three primates (human (*Homo sapiens*), chimpanzee (*Pan troglodytes*) and rhesus monkey (*Macaca mulatta*)), and TRIM6 nucleotide and amino acid sequences of European rabbit and human were also included. The trees typology was coincident with the known species tree [41, 42, 62, 63], where TRIM6 sequences represented an outgroup, and primate and lagomorph TRIM5α formed two orthologs groups. Due to the identical typology between the two sets of data, only the tree based on nucleotide sequences is represented in Figure 4.

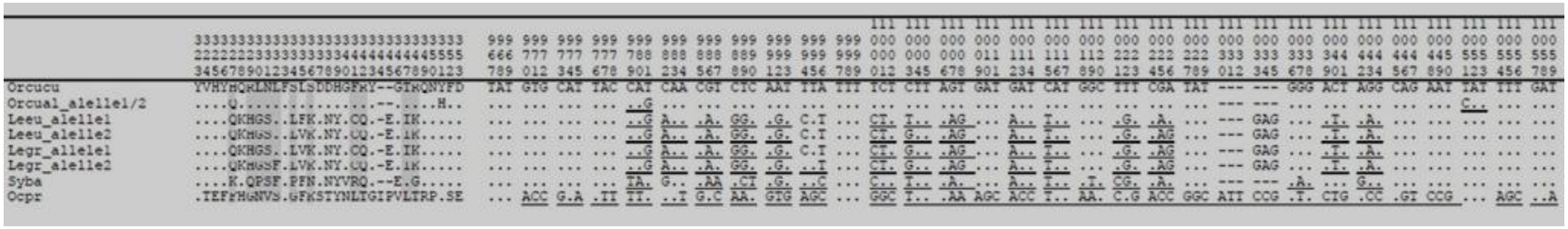


Figure 3 - Amino acid and nucleotide sequences of Lagomorpha PRYSPRY v1 region.

Amino acid and nucleotide sequences of PRYSPRY v1 region are represented for European rabbit subspecies *Oryctolagus cuniculus cuniculus* (Orcucu) and *Oryctolagus cuniculus algirus* (Orcual), European brown hare (Leeu – allele 1 and 2), Iberian hare (Legr – allele 1 and 2), brush rabbit (Syba) and American pika (Ocpr). The shadowed region on the amino acid representation corresponds to positively-selected sites obtained by REL analysis. Non-synonymous substitutions are underlined on the nucleotide sequences of Lagomorpha PRYSPRY v1 region. Residues are numbered as in Figure 1.

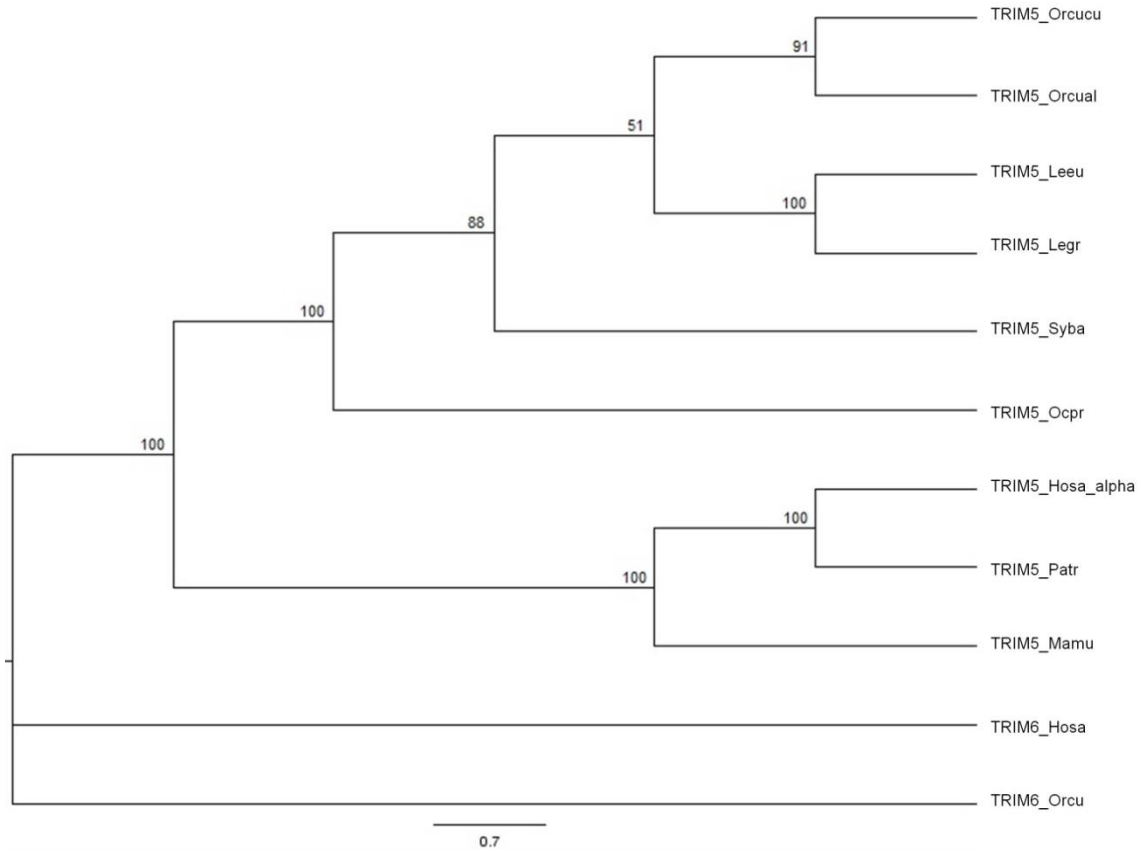


Figure 4 - Maximum Likelihood phylogenetic tree of lagomorph and primate *TRIM5* and *TRIM6* nucleotide sequences. *TRIM5* nucleotide sequences from Lagomorpha and three primates (human (Hosa), chimpanzee (Patr) and rhesus monkey (Mamu)), and *TRIM6* nucleotide sequences from European rabbit and human were used to construct a Maximum Likelihood phylogenetic tree. The analyses were performed with 1,000,000 generations and 1,000 bootstrap searches. The bootstrap values are indicated on the branches.

4.2. Inference of positive selection in Lagomorpha TRIM5α protein

To identify a specific pattern of nucleotide substitution in the leporid TRIM5α protein, synonymous and non-synonymous substitution rates were estimated using the Nei-Gojobori method [55] and a non-synonymous to synonymous substitution ratio (d_N/d_S) was calculated (Table 1). Under neutrality, coding sequences are expected to present a ratio of non-synonymous substitutions (d_N) over synonymous substitutions (d_S) that does not significantly deviate from 1 ($\omega = d_N/d_S = 1$), while significant deviations may be interpreted as either the result of positive selection ($\omega >> 1$) or of negative selection ($\omega << 1$). This simple analysis showed that the ratios obtained between genera ranged from 2.8 to 4.3, clearly higher than 1, suggesting that TRIM5α is under strong positive selection.

The high variability of the PRYSPRY domain and the positive selection of TRIM5α described in primates [18, 25, 31-33] prompted Fletcher and colleagues (2010) [40] to perform a codon-based selection analysis using the random effect likelihood (REL) model. With this analysis, the authors identified 11 positively-selected codons, 4 of which were located in the PRYSPRY v1 region [40]. Using the same methodology and parameters, we identified 25 positively-selected codons, 20 of which are located in the PRYSPRY domain and, more specifically, 11 in the v1 region (Table 2). In Figure 3 the amino acid and nucleotide sequences of Lagomorpha PRYSPRY v1 region are represented and the non-synonymous substitutions are marked. It should be pointed out that the PRYSPRY domain was the site of the most intense positive selection in primates [31-33]; its v1 region was identified as the major determinant of anti-HIV-1 potency distinguishing the human and rhesus monkey TRIM5α proteins [16-18, 23-25, 32]. The proposed evolutionary model for primate TRIM5α in which a history of virus-host interactions led to species-specific adaptations [31] can be considered also for leporids. But again, the *trans*-species scenario between leporid species cannot be ruled out to explain the species-specific variations.

Table 1 - Leporid *TRIM5* estimation of non-synonymous to synonymous substitution ratio (d_N/d_S)

	Orcucu	Orcual allele1	Orcual allele2	Leeu allele1	Leeu allele2	Legr allele1	Legr allele2
Orcucu							
Orcual_allele1	1.5						
Orcual_allele2	1.5	1.0					
Leeu_allele1	3.0	3.1	3.4				
Leeu_allele2	2.9	2.9	3.2	3.0			
Legr_allele1	3.2	3.2	3.2	2.5	2.0		
Legr_allele2	3.2	3.2	3.2	3.0	3.0	n. a.	
Syba	2.8	2.9	2.9	4.1	3.8	4.3	4.1

Oryctolagus cuniculus cuniculus (Orcucu), *Oryctolagus cuniculus algirus* (Orcual), European brown hare (Leeu), Iberian hare (Legr) and brush rabbit (Syba).
n.a.- not applicable (no non-synonymous substitutions were observed).

Our striking observation was visually reinforced by sliding-window analysis of *TRIM5* nucleotide divergence between species (Figure 5). Comparing *Oryctolagus* with other leporid genera that diverged about 12 My ago, the nucleotide differences throughout the gene occurred primarily around 0.00-0.10 and 0.00-0.05 nucleotide replacements per site for *Lepus* and *Sylvilagus* species, respectively (Figure 5A, 5B and 5C). Nevertheless, the nucleotide differences peak (~0.20) was observed around

Table 2 - Positively-selected codon positions in the Lagomorpha TRIM5α deduced protein sequences

Codon[a]	Normalized E [dN-dS]	Posterior Probability	Bayes Factor	Region	Orcu[b]	Leeu	Legr	Syba	Ocpr
3	0.30	1.00	795	-	S^0	S^0	S^0	$\underline{A}^0$	S^0
148	0.29	0.99	224	CC	N^0/K^+	N^0	N^0	$H^{0/+}$	K^+
196	0.29	0.99	179	CC	E^-	$\underline{G}^0$	$\underline{G}^0$	E^-	$\underline{M}^0$
279	0.30	1.00	1913	-	$\underline{I}^0$	T^0	$\underline{I}^0$	$\underline{I}^0$	$\underline{A}^0$
288	0.29	0.99	178	-	$\underline{M}^0$	$\underline{V}^0$	$\underline{V}^0$	$\underline{M}^0$	N^0
327	0.30	1.00	6611	v1	$H^{0/+}/Q^0$	Q^0	Q^0	K^+	$\underline{F}^0$
329	0.30	1.00	578	v1	R^+	$H^{0/+}$	$H^{0/+}$	Q^0	$\underline{G}^0$
330	0.30	1.00	35572	v1	$\underline{L}^0$	$\underline{G}^0$	$\underline{G}^0$	$\underline{P}^0$	N^0
331	0.30	0.99	405	v1	N^0	S^0	S^0	S^0	$\underline{V}^0$
332	0.30	1.00	33711	v1	$\underline{L}^0$	$\underline{L}^0$	$\underline{L}^0$	$\underline{F}^0$	S^0
334	0.30	1.00	5055	v1	S^0	$\underline{L}^0$	$\underline{L}^0$	$\underline{P}^0$	$\underline{G}^0$
336	0.30	0.99	294	v1	S^0	K^+	K^+	N^0	K^+
341	0.30	1.00	483	v1	$\underline{F}^0$	$\underline{C}^0$	$\underline{C}^0$	R^+	$\underline{L}^0$
342	0.30	1.00	574	v1	R^+	Q^0	Q^0	Q^0	T^0
347	0.30	0.99	320	v1	T^0	$\underline{I}^0$	$\underline{I}^0$	T^0	$\underline{L}^0$
348	0.30	1.00	691	v1	R^+	K^+	K^+	$\underline{G}^0$	T^0
391	0.30	1.00	3822	v2	T^0	$\underline{I}^0$	$\underline{I}^0$	T^0	Q^0
396	0.30	0.99	438	v2	Y^0	Y^0	Y^0	$\underline{F}^0$	$\underline{P}^0$
399	0.30	0.99	368	v2	$\underline{V}^0$	$\underline{F}^0$	$\underline{F}^0$	$\underline{A}^0$	$\underline{G}^0$
415	0.28	0.98	134	v3	$\underline{I}^0$	N^0	N^0	N^0	$\underline{G}^0$
434	0.30	0.99	432	v3	$\underline{P}^0$	$\underline{P}^0$	$\underline{P}^0$	$\underline{L}^0$	S^0
457	0.29	0.99	217	-	K^+	T^0	T^0	K^+	$\underline{V}^0$
464	0.30	0.99	416	-	S^0	S^0	S^0	$\underline{A}^0$	S^0
469	0.30	0.99	418	-	S^0	$\underline{A}^0$	$\underline{A}^0$	S^0	S^0
480	0.30	1.00	639	-	$H^{0/+}$	Q^0	Q^0	Q^0	T^0

European rabbit (Orcu), European brown hare (Leeu), Iberian hare (Legr), brush rabbit (Syba) and American pika (Ocpr).

[a] Codon positions are numbered according to the alignment in Figure 1.

[b] Single amino acid = same for both Orcu subspecies, *Oryctolagus cuniculus cuniculus* (Orcucu) and *Oryctolagus cuniculus algirus* (Orcual); Two amino acids divided by "/" = Orcucu amino acid/Orcual amino acid.

Codon physical-chemical properties are also represented: Underlined codon = non-polar amino acid; non-underlined codon = polar amino acid; $^+$ = positive amino acid; $^-$ = negative amino acid; ^{0} = neutral amino-acid.

nucleotide position 1000, where the PRYSPRY v1 region is located (positions 967-1059). The 40 My separation between *Ochotona* and the leporids is apparent in the 0.30 to 0.60 nucleotide replacements per site when comparing American pika to European rabbit (Figure 5D). However, the peak in the v1 region is still clearly defined. These observations can be compared to the nucleotide differences among primates. The approximately 4.5-6 My of evolutionary divergence between human and chimpanzee [62-64] resulted in a nucleotide difference of 0.00-0.03 nucleotide replacements per site, although some peaks are still defined, including one around the nucleotide position 1000 (Figure 5E). Comparison of the human and rhesus monkey also revealed a peak in the PRYSPRY v1 region (Figure 5F), variance similar to that observed among leporid genera (~0.20), although the average nucleotide differences are higher than between leporids (0.00-0.15 nucleotide replacements per site), which is consistent with the < 31 My divergence time between human and rhesus monkey [18, 31].

The interest in studying ancient extinct viruses (paleoviruses) in primate genomes has increased in the past few years. However, using sequences of "modern" viruses to identify paleoviruses has been a problem and some new strategies began to be applied. The approach broadly used consists in looking for signatures of evolutionary adaptation in anti-viral genes [65]. Several primate anti-viral genes have already been studied and positive selection was inferred, including the focus of this paper, TRIM5α [e.g. 31, 33, 66, 67]. The detection of extensive diversity in primate TRIM5α led the scientific community to speculate that endogenous retroviruses and/or exogenous lentiviral pathogens may have exerted selective pressure on this host restriction factor and, in the specific case of human TRIM5α, that the acquisition of resistance to specific ancient endogenous retroviruses may be responsible for our susceptibility to HIV-1 in the present-day [31, 32, 68].

Assuming that selective pressure acts on the TRIM5α region that recognizes variation in the capsid of retroviruses, it was predicted that the PRYSPRY v1 region represents the interface with the capsid [18, 25, 31, 32]. Recent studies showed that RELIK is highly similar structurally to modern-day exogenous lentiviruses and that the capacity of the capsid to form a protein-protein complex with CypA is maintained [69]. CypA is a host cell peptidyl proline isomerase that binds to the retroviral capsid [70, 71]. With the absence of known exogenous lentiviruses affecting leporids, endogenous retroviruses such as RELIK were suggested to dominate leporid TRIM5α evolution after host germline infection [40]. Our prediction was that the TRIM5α protein from the third leporid genus known to harbor RELIK, the New World genus *Sylvilagus*, should also

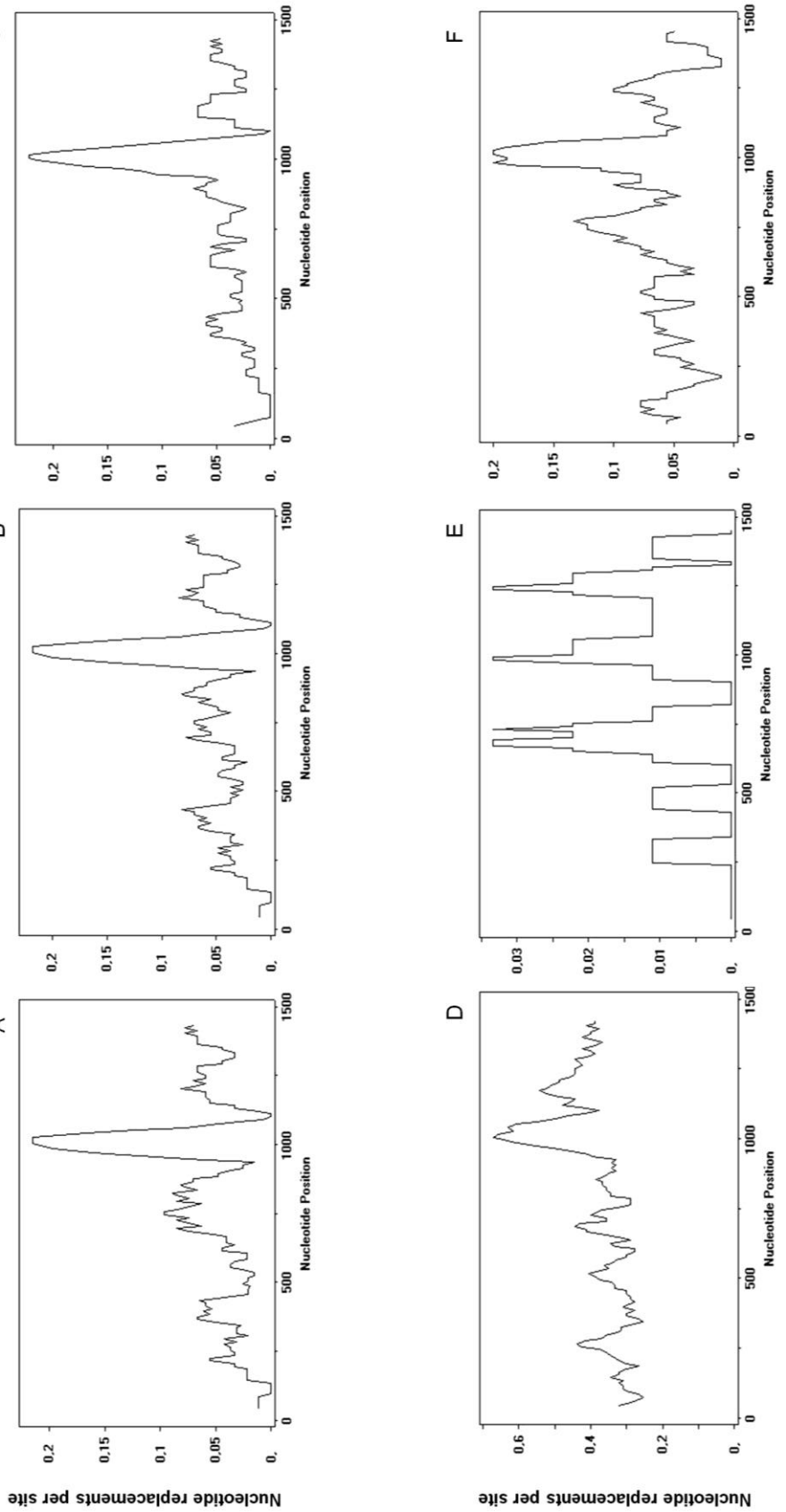


Figure 5 - Sliding-window analysis to detect nucleotide differences between *TRIM5* genes from different species.
(A), (B), (C) and (D) represent the nucleotide differences between the lagomorphs European rabbit/European brown hare, European rabbit/Iberian hare, European rabbit/brush rabbit and European rabbit/American pika, respectively. (E) and (F) represent the differences in nucleotide replacements per site between the primates human/chimpanzee and human/rhesus monkey, respectively.

reflect selective pressure patterns in specific regions previously reported for the Old World *Oryctolagus* and *Lepus* genera. In fact, despite *Sylvilagus* being more closely related to *Oryctolagus*, the sequence divergence in TRIM5α is comparable to that found between European brown hare, Iberian hare and the two *O. cuniculus* subspecies, particularly in the PRYSPRY v1 region where the majority of positively-selected codons is concentrated.

The current release of both European rabbit and American pika genomes has increased the opportunities to identify other Lagomorpha endogenous retroviruses, remnants of ancient extinct viruses. At the same time, while comparing sequence data from several orthologs of lagomorph anti-viral genes, signatures of evolutionary change in these anti-viral genes could date when different ancient viral pathogens acted. Of course, it is not correct to fully assume that only one ancient retrovirus was responsible for selectively pressuring a specific anti-viral gene and vice versa, or that an exogenous lentiviruses did not play a preponderant role in TRIM5α Lagomorpha evolution. However, until other endogenous retroviruses are identified in leporid or even lagomorph genomes and due to the absence of known exogenous lentiviruses infecting leporids, we speculate that endogenous retroviruses like RELIK could have acted as evolutionary forces on leporid TRIM5α.

5. Conclusions

This evolutionary study on Lagomorpha *TRIM5* gene shows a remarkable differentiation in the PRYSPRY v1 region suggesting that this gene has evolved under a high selective pressure within the Lagomorpha order. With the exception of studies on the primate lineage, this is one of the first comprehensive and detailed evolutionary studies of the anti-retroviral restriction factor TRIM5α. Furthermore, the similarities observed in the species split within primates and lagomorphs allow the establishment of comparisons of the evolutionary patterns observed in *TRIM5* gene.

6. References

1. Boeke JD, Stoye JP: Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In *Retroviruses.* Edited by Coffin JM, Hughes SH, Varmus HE. New York: CSHL Press; 1997: 343-435

2. Petropoulos C: Retroviral taxonomy, protein structures, sequences, and genetic maps. In *Retroviruses.* Edited by Coffin JM, Hughes SH, Varmus HE. New York: CSHL Press; 1997: 757-805
3. Gifford R, Tristem M: The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* 2003, 26:291-315.
4. Katzourakis A, Tristem M, Pybus OG, Gifford RJ: Discovery and analysis of the first endogenous lentivirus. *Proceedings of the National Academy of Sciences of the United States of America* 2007, 104:6261-6265.
5. Keckesova Z, Ylinen LM, Towers GJ, Gifford RJ, Katzourakis A: Identification of a RELIK orthologue in the European hare (*Lepus europaeus*) reveals a minimum age of 12 million years for the lagomorph lentiviruses. *Virology* 2009, 384:7-11.
6. van der Loo W, Abrantes J, Esteves PJ: Sharing of endogenous lentiviral gene fragments among leporid lineages separated for more than 12 million years. *Journal of Virology* 2009, 83:2386-2388.
7. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J: The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* 2004, 427:848-853.
8. Sayah DM, Sokolskaja E, Berthoux L, Luban J: Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* 2004, 430:569-573.
9. Nisole S, Stoye JP, Saib A: TRIM family proteins: retroviral restriction and antiviral defence. *Nature Reviews Immunology* 2005, 3:799-808.
10. Ozato K, Shin DM, Chang TH, Morse HC, 3rd: TRIM family proteins and their emerging roles in innate immunity. *Nature Reviews Immunology* 2008, 8:849-860.
11. Reymond A, Meroni G, Fantozzi A, Merla G, Cairo S, Luzi L, Riganelli D, Zanaria E, Messali S, Cainarca S, Guffanti A, Minucci S, Pelicci PG, Ballabio A: The tripartite motif family identifies cell compartments. *The EMBO Journal* 2001, 20:2140-2151.
12. Yamauchi K, Wada K, Tanji K, Tanaka M, Kamitani T: Ubiquitination of E3 ubiquitin ligase TRIM5 alpha and its potential role. *FEBS Journal* 2008, 275:1540-1555.
13. Javanbakht H, Diaz-Griffero F, Stremlau M, Si Z, Sodroski J: The contribution of RING and B-box 2 domains to retroviral restriction mediated by monkey TRIM5alpha. *The Journal of Biological Chemistry* 2005, 280:26933-26940.
14. Li X, Song B, Xiang SH, Sodroski J: Functional interplay between the B-box 2 and the B30.2(SPRY) domains of TRIM5alpha. *Virology* 2007, 366:234-244.

15. Li X, Sodroski J: The TRIM5alpha B-box 2 domain promotes cooperative binding to the retroviral capsid by mediating higher-order self-association. *Journal of Virology* 2008, 82:11495-11502.
16. Javanbakht H, Yuan W, Yeung DF, Song B, Diaz-Griffero F, Li Y, Li X, Stremlau M, Sodroski J: Characterization of TRIM5alpha trimerization and its contribution to human immunodeficiency virus capsid binding. *Virology* 2006, 353:234-246.
17. Yap MW, Nisole S, Stoye JP: A single amino acid change in the SPRY domain of human Trim5alpha leads to HIV-1 restriction. *Current Biology* 2005, 15:73-78.
18. Stremlau M, Perron M, Welikala S, Sodroski J: Species-specific variation in the B30.2(SPRY) domain of TRIM5alpha determines the potency of human immunodeficiency virus restriction. *Journal of Virology* 2005, 79:3139-3145.
19. Hatziioannou T, Perez-Caballero D, Yang A, Cowan S, Bieniasz PD: Retrovirus resistance factors Ref1 and Lv1 are species-specific variants of TRIM5alpha. *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101:10774-10779.
20. Keckesova Z, Ylinen LM, Towers GJ: The human and African green monkey TRIM5alpha genes encode Ref1 and Lv1 retroviral restriction factor activities. *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101:10780-10785.
21. Yap MW, Nisole S, Lynch C, Stoye JP: Trim5alpha protein restricts both HIV-1 and murine leukemia virus. *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101:10786-10791.
22. Perron MJ, Stremlau M, Song B, Ulm W, Mulligan RC, Sodroski J: TRIM5alpha mediates the postentry block to N-tropic murine leukemia viruses in human cells. *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101:11827-11832.
23. Li Y, Li X, Stremlau M, Lee M, Sodroski J: Removal of arginine 332 allows human TRIM5alpha to bind human immunodeficiency virus capsids and to restrict infection. *Journal of Virology* 2006, 80:6738-6744.
24. Sebastian S, Luban J: TRIM5alpha selectively binds a restriction-sensitive retroviral capsid. *Retrovirology* 2005, 2:40.
25. Song B, Gold B, O'Huigin C, Javanbakht H, Li X, Stremlau M, Winkler C, Dean M, Sodroski J: The B30.2(SPRY) domain of the retroviral restriction factor TRIM5alpha exhibits lineage-specific length and sequence variation in primates. *Journal of Virology* 2005, 79:6111-6121.

26. James LC, Keeble AH, Khan Z, Rhodes DA, Trowsdale J: Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proceedings of the National Academy of Sciences of the United States of America* 2007, 104:6200-6205.
27. Tareen SU, Emerman M: Human Trim5alpha has additional activities that are uncoupled from retroviral capsid recognition. *Virology* 2011, 409:113-120.
28. Pertel T, Hausmann S, Morger D, Zuger S, Guerra J, Lascano J, Reinhard C, Santoni FA, Uchil PD, Chatel L, Bisiaux A, Albert ML, Strambio-De-Castillia C, Mothes W, Pizzato M, Grutter MG, Luban J: TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* 2011, 472:361-365.
29. Ganser-Pornillos BK, Chandrasekaran V, Pornillos O, Sodroski JG, Sundquist WI, Yeager M: Hexagonal assembly of a restricting TRIM5alpha protein. *Proceedings of the National Academy of Sciences of the United States of America* 2011, 108:534-539.
30. Zhao G, Ke D, Vu T, Ahn J, Shah VB, Yang R, Aiken C, Charlton LM, Gronenborn AM, Zhang P: Rhesus TRIM5alpha disrupts the HIV-1 capsid at the inter-hexamer interfaces. *PLoS Pathogens* 2011, 7:e1002009.
31. Sawyer SL, Wu LI, Emerman M, Malik HS: Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102:2832-2837.
32. Johnson WE, Sawyer SL: Molecular evolution of the antiretroviral TRIM5 gene. *Immunogenetics* 2009, 61:163-176.
33. Soares EA, Menezes AN, Schrago CG, Moreira MA, Bonvicino CR, Soares MA, Seuanez HN: Evolution of TRIM5alpha B30.2 (SPRY) domain in New World primates. *Infection, Genetics and Evolution* 2010, 10:246-253.
34. Newman RM, Hall L, Connole M, Chen GL, Sato S, Yuste E, Diehl W, Hunter E, Kaur A, Miller GM, Johnson WE: Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5alpha. *Proceedings of the National Academy of Sciences of the United States of America* 2006, 103:19134-19139.
35. Cagliani R, Fumagalli M, Biasin M, Piacentini L, Riva S, Pozzoli U, Bonaglia MC, Bresolin N, Clerici M, Sironi M: Long-term balancing selection maintains trans-specific polymorphisms in the human TRIM5 gene. *Human Genetics* 2010, 128:577-588.

36. Si Z, Vandegraaff N, O'Huigin C, Song B, Yuan W, Xu C, Perron M, Li X, Marasco WA, Engelman A, Dean M, Sodroski J: Evolution of a cytoplasmic tripartite motif (TRIM) protein in cows that restricts retroviral infection. *Proceedings of the National Academy of Sciences of the United States of America* 2006, 103:7454-7459.
37. Sawyer SL, Emerman M, Malik HS: Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. *PLoS Pathogens* 2007, 3:e197.
38. Schaller T, Hue S, Towers GJ: An active TRIM5 protein in rabbits indicates a common antiviral ancestor for mammalian TRIM5 proteins. *Journal of Virology* 2007, 81:11713-11721.
39. Tareen SU, Sawyer SL, Malik HS, Emerman M: An expanded clade of rodent Trim5 genes. *Virology* 2009, 385:473-483.
40. Fletcher AJ, Hue S, Schaller T, Pillay D, Towers GJ: Hare TRIM5alpha restricts divergent retroviruses and exhibits significant sequence variation from closely related lagomorpha TRIM5 genes. *Journal of Virology* 2010, 84:12463-12468.
41. Matthee CA, van Vuuren BJ, Bell D, Robinson TJ: A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. *Systematic Biology* 2004, 53:433-447.
42. Matthee CA: Pikas, hares and rabbits (Lagomorpha). In *The timetree of life.* Edited by Hedges SB, Kumar S. New York: Oxford University Press; 2009: 487–489
43. Lanier HC, Olson LE: Inferring divergence times within pikas (*Ochotona* spp.) using mtDNA and relaxed molecular dating techniques. *Molecular Phylogenetics and Evolution* 2009, 53:1-12.
44. Ferrand N, Branco M: The evolutionary history of the European rabbit (*Oryctolagus cuniculus*): major patterns of population differentiation and geographic expansion inferred from protein polymorphism. In *Phylogeography of Southern European Refugia.* Edited by Weiss S, Ferrand N. The Netherlands: Springer; 2007: 207-235
45. Krug MS, Berger SL: First-strand cDNA synthesis primed with oligo(dT). *Methods in Enzymology* 1987, 152:316-325.
46. Zwickl DJ: Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. University of Texas, 2006.

47. Posada D: jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 2008, 25:1253-1256.
48. Jones DT, Taylor WR, Thornton JM: The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* 1992, 8:275-282.
49. Abascal F, Zardoya R, Posada D: ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 2005, 21:2104-2105.
50. Drummond A, Strimmer K: PAL: an object-oriented programming library for molecular evolution and phylogenetics. *Bioinformatics* 2001, 17:662-663.
51. Guindon S, Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 2003, 52:696-704.
52. Fares MA, Elena SF, Ortiz J, Moya A, Barrio E: A sliding window-based method to detect selective constraints in protein-coding genes and its application to RNA viruses. *Journal of Molecular Evolution* 2002, 55:509-521.
53. Kosakovsky Pond SL, Frost SD: Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* 2005, 22:1208-1222.
54. Pond SL, Frost SD: Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 2005, 21:2531-2533.
55. Nei M, Gojobori T: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* 1986, 3:418-426.
56. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 2011, 28:2731-2739.
57. Tajima F: Determination of window size for analyzing DNA sequences. *Journal of Molecular Evolution* 1991, 33:470-473.
58. Librado P, Rozas J: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009, 25:1451-1452.
59. Carmo CR, Esteves PJ, Ferrand N, van der Loo W: Genetic variation at chemokine receptor CCR5 in leporids: alteration at the 2nd extracellular domain by gene conversion with CCR2 in *Oryctolagus*, but not in *Sylvilagus* and *Lepus* species. *Immunogenetics* 2006, 58:494-501.
60. Abrantes J, Carmo C, Matthee C, Yamada F, Loo W, Esteves P: A shared unusual genetic change at the chemokine receptor type 5 between *Oryctolagus*, *Bunolagus* and *Pentalagus*. *Conservation Genetics* 2011, 12:325-330.

61. Esteves PJ, Lanning D, Ferrand N, Knight KL, Zhai SK, van der Loo W: The evolution of the immunoglobulin heavy chain variable region (IgVH) in Leporids: an unusual case of transspecies polymorphism. *Immunogenetics* 2005, 57:874-882.
62. Siepel A: Phylogenomics of primates and their ancestral populations. *Genome Research* 2009, 19:1929-1941.
63. Takahata N, Satta Y: Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. *Proceedings of the National Academy of Sciences of the United States of America* 1997, 94:4811-4815.
64. Johnson WE, Coffin JM: Constructing primate phylogenies from ancient retrovirus sequences. *Proceedings of the National Academy of Sciences of the United States of America* 1999, 96:10254-10260.
65. Emerman M, Malik HS: Paleovirology - modern consequences of ancient viruses. *PLoS Biology* 2010, 8:e1000301.
66. Kerns JA, Emerman M, Malik HS: Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. *PLoS Genetics* 2008, 4:e21.
67. Elde NC, Child SJ, Geballe AP, Malik HS: Protein kinase R reveals an evolutionary model for defeating viral mimicry. *Nature* 2009, 457:485-489.
68. Kaiser SM, Malik HS, Emerman M: Restriction of an extinct retrovirus by the human TRIM5alpha antiviral protein. *Science* 2007, 316:1756-1758.
69. Goldstone DC, Yap MW, Robertson LE, Haire LF, Taylor WR, Katzourakis A, Stoye JP, Taylor IA: Structural and functional analysis of prehistoric lentiviruses uncovers an ancient molecular interface. *Cell Host & Microbe* 2010, 8:248-259.
70. Gamble TR, Vajdos FF, Yoo S, Worthylake DK, Houseweart M, Sundquist WI, Hill CP: Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* 1996, 87:1285-1294.
71. Luban J, Bossolt KL, Franke EK, Kalpana GV, Goff SP: Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell* 1993, 73:1067-1078.

Chapter 6

Final considerations

1. Co-evolution between myxoma virus and Leporidae host immune genes

The successful replication and propagation of myxoma virus (MYXV) within its natural long-term *Sylvilagus* hosts and in its novel and susceptible host, the European rabbit (*Oryctolagus cuniculus*), requires the evasion and manipulation of the potent and complex host immune defenses by a remarkable repertoire of viral proteins [1-5]. The consequences of this dynamic battle between MYXV and its hosts are expected to be manifested by the presence of selective pressure hallmarks in their genome, with particular incidence in the host immune genes.

Once MYXV successfully crossed the species barrier, it had already partially adapted to its natural long-term hosts, the tapeti (*S. brasiliensis*) and the brush rabbit (*S. bachmani*). It is still unclear why MYXV infection has such a strikingly different outcome in the natural hosts and in the susceptible European rabbit; yet, it is likely that the severity of the effects of viral proteins depends on the genotype of the host [5, 6]. Besides, the genetic basis of resistance of certain European rabbit populations, namely in Australia, is also still not understood, but might be due to a single gene with multiple alleles encoding various levels of resistance or, instead, to multiple genes [5]. Unfortunately, most of the studies on MYXV infection have been performed on the susceptible laboratory European rabbit.

To enhance our knowledge of host immune genes likely to play a crucial role in fighting, restricting and inhibiting viral infection and replication, we studied the evolutionary and genetic aspects of some of these key genes. The analyses performed on the sterile alpha motif domain-containing protein 9 (*SAMD9*), SAMD9-like (*SAMD9L*), retinoic acid-inducible gene-I (*RIG-I*), melanoma differentiation-associated factor protein 5 (*MDA5*), laboratory of genetics and physiology 2 (*LGP2*) and tripartite motif-containing protein 5 (*TRIM5*) genes identified distinct genetic characteristics between mammals, and also between leporid species, that can be further tested with functional studies. In addition, the evolutionary signatures of genetic conflict imposed by viral selective pressures allowed the identification of positively selected sites that can also be functionally tested. On the other hand, the C-C motif chemokine ligands 3, 4 and 5 (CCL3, CCL4 and CCL5) might have been structurally constrained, as they exhibited strong evidence of negative selection.

Chemokines are chemotactic cytokines that activate specific G protein-coupled receptors [7-8]. Cells produce inflammatory chemokines during infection or in the

presence of a pro-inflammatory stimulus. These chemokines are responsible for the recruitment and migration of effector cells, including monocytes, granulocytes and effector T cells, to the injured or infected site. They also activate the cells to mount an immune response and initiate wound healing [7-10]. Elucidative examples of inflammatory chemokines are the C-C motif chemokine ligands 3, 4 and 5 (CCL3, CCL4 and CCL5), which signal through the C-C motif chemokine receptor 5 (CCR5) [11, 12].

The study we performed on leporid *CCL3*, *CCL4* and *CCL5* genes resulted from the previously reported uniqueness of *Oryctolagus CCR5*, later extended to the genera *Pentalagus* and *Bunolagus* [13, 14]. The receptor in these genera underwent a gene conversion event with CCR2, where the CCR5 sequence motif QTLKMT in the second extracellular loop was replaced by the CCR2 motif HTIMRN. This recombinatory event was not observed in *Sylvilagus* or *Lepus* species [13]. Importantly, the second extracellular loop of the chemokine receptor CCR5 has been identified as the major determinant for ligand specificity [15]. In light of these observations, we genetically characterized the three inflammatory chemokine ligands of CCR5 in *Oryctolagus*, *Sylvilagus* and *Lepus* to understand the functional significance of this gene conversion event in *Oryctolagus* CCR5. However, we found only a few amino acid substitutions in the regions of leporid CCL3, CCL4 and CCL5 related to signaling and receptor-binding. The necessity of studying CCR5-ligand binding affinity and pattern became evident, not only for the gene conversion occurring in the second extracellular loop of the receptor, but also for the species- or genus-specific amino acid substitutions in the receptor-binding regions of the ligands. Due to the importance of this inflammatory chemokine receptor-ligand system in host innate immune defense, the relevance of these inter-species differences in receptor and ligands to MYXV susceptibility/resistance should be investigated.

The biological function of human SAMD9 is not known, but it has been suggested to play an important role in the signaling network as an interferon-stimulated gene (ISG), because *SAMD9* can be regulated by interferons (IFNs), tumor necrosis factor (TNF) and cellular stress signals [16-18]. In addition, human SAMD9 exhibits anti-viral functions by inhibiting MYXV replication, an action counteracted by the viral protein M062 [19]. Also, the close paralogue of human SAMD9, the SAMD9-like (SAMD9L) protein, is thought to play a role in cellular signaling, because it was found to be induced by type I IFNs [20].

It has been reported that, *in vitro*, European rabbit SAMD9 possibly binds to M062, but the difficulty of investigating SAMD9 in the rabbit has not allowed further

advances [19]. Nevertheless, European rabbits infected with MYXV deficient in *M062R* did not exhibit any classical clinical signs of myxomatosis and rabbit cells infected with the *M062R*-deficient virus underwent abortive infection [19].

Our studies confirmed the absence of the *SAMD9* gene in the house mouse genome, which still retains the paralogue, *SAMD9L*. In addition, we found that other mammalian genomes have apparently lost either *SAMD9* or *SAMD9L*, suggesting some overlapping functional redundancy between the two proteins. We collected both *SAMD9* and *SAMD9L* gene sequences for European rabbit from the Ensembl database, but, with the identified roles in human innate immunity and with the clear host range function in rabbit, it is important to verify the presence of the *SAMD9* and *SAMD9L* genes in other leporid species. Furthermore, the impact of possible inter-specific genetic differences between these genes should also be evaluated functionally in European rabbit and *Sylvilagus* host cells to determine whether SAMD9 is in fact also the binding partner of M062 in these species.

The ancient struggle for survival between host and virus led to the development of multiple strategies to antagonize each other. Because it is critical for the host to mount an effective innate immune response immediately after infection, the detection of intracellular viral RNA from actively replicating viruses by RIG-I-like receptors (RLRs) is an indispensable system for anti-viral immunity [21-23]. The role of RIG-I and MDA5, two members of the RLR family, in recognizing different classes of RNA viruses or DNA viruses that express unusual RNA species makes them strong candidates to be under intense selective pressures, as has been previously demonstrated for another set of host defense genes involved in recognizing pathogen molecules, the Toll-like receptors (TLRs) [24-26].

MYXV infection of primary human macrophages is sensed largely by RIG-I, which consequently mediates the co-induction of type I IFN and TNF, demonstrating a novel innate defense mechanism to restrict viral infection in human cells [27]. Since the European rabbit is the known susceptible host of MYXV, it is thought that the virus possesses the capacity to subvert RIG-I sensing in rabbit cells, which would impose strong selective pressures on the *RIG-I* gene. The gene is thus expected to mutate to keep the pace with viral strategies to evade RIG-I detection.

Striking evidence of long-term trans-acting selective pressures was found in our studies for both the mammalian and leporid *RIG-I*, *MDA5* and *LGP2* genes. Studies have extensively described the role of RIG-I in binding to specific patterns for each virus species, but our results lead us to postulate that the multitude of viruses

potentially infecting different hosts, or even the same host, has imposed different pressures on *RIG-I* evolution. Since sensing of MYXV in human cells has been demonstrated for this specific RLR, the identification of codons under strong positive selection and amino acid differences in the RIG-I repressor domain between the susceptible European rabbit and the natural brush rabbit host demonstrate the importance of performing functional studies to evaluate the role of RIG-I in MYXV infection.

The first endogenous lentivirus identified in any species, the rabbit endogenous lentivirus type K (RELIK), is present in the genome of the European rabbit and in other leporid species from the genera *Lepus*, *Sylvilagus* and *Bunolagus* [28-30]. Because there are no exogenous retroviruses known to currently infect leporids, endogenous retroviruses such as RELIK were suggested to play a role in shaping the evolution of host genes encoding factors with anti-retroviral restriction activity. One of these anti-retroviral restriction factors is the TRIM5α protein, already identified in European rabbit and European brown hare (*Lepus europaeus*), which restricts the replication of several retroviruses in both species [31, 32].

The study of TRIM5α evolution in Primates revealed strong evidence of positive selection in a specific region of the protein, the variable region v1 in the PRYSPRY domain [33-36]. This same TRIM5α region exhibited significant differences between European rabbit and European brown hare and, therefore, we were interested in extending the study to other leporid species. The divergent species-specific pattern observed between the *Oryctolagus* and *Lepus* PRYSPRY-domains was also present in *Sylvilagus* TRIM5α, supporting the hypothesis that endogenous lentiviruses like RELIK might have been the driving force of leporid TRIM5α evolution. Recently, a study evaluated the role of RELIK in the evolution of TRIM5α by testing the ability of European rabbit and Eastern cottontail (*Sylvilagus floridanus*) TRIM5α to restrict RELIK capsid (CA)-containing viruses, and indeed the results supported the premise that the endogenous lentivirus RELIK might exert selective pressure on lagomorph TRIM5α [37].

2. Future perspectives

Several other host candidate genes could provide insight into MYXV susceptibility versus resistance differences in leporid species. Amongst them, we suggest a few that should be pursued in future studies: tumor necrosis factor (TNF), interferon-gamma (IFN-γ), double-stranded RNA-dependent protein kinase R (PKR) and 2'-5'-oligoadenylate synthetase (2'-5'OAS).

MYXV produces two European rabbit-specific viroreceptors, M002 and M007, which mimic the TNF and IFN-γ receptors, respectively. Particularly, it has been shown that M007 binds and inhibits European rabbit IFN-γ and does not bind either human or murine IFN-γ [38]. Such observations support our interest in studying genetic differences in the IFN-γ and TNF genes of European rabbit and the natural MYXV long-term host, *Sylvilagus*, as well as the binding affinity and inhibition level of M007 and M002 to the corresponding cytokines in different MYXV hosts.

A multitude of poxvirus immunomodulatory proteins target elements in the IFN response pathway, including PKR and 2'-5'OAS, which are host proteins activated by double-stranded RNA (dsRNA) produced during poxviral transcription [39, 40]. Extensively studied examples are the vaccinia virus (VACV) *E3L* and *K3L* genes, which encode the prototypical poxvirus inhibitors of these two host proteins. VACV E3, encoded by the *E3L* gene, is a dsRNA-binding protein that prevents PKR and 2'-5'OAS activation by sequestering viral dsRNA [41, 42]. E3 can also bind directly to PKR, preventing the phosphorylation of eukaryotic translation initiation factor 2α (eIF2α) [43]. The *K3L* gene product (protein K3) is a structural mimic of eIF2α that functions as a suicide pseudosubstrate of PKR, competitively inhibiting eIF2α phosphorylation [42, 44]. MYXV encodes the M029 protein, a member of the poxvirus E3 family of dsRNA-binding proteins, as well as the M156 protein, which exhibits a similar crystal structure as the VACV K3 protein and is an eIF2α viral mimic [45, 46]. Because MYXV produces these immunomodulatory proteins, it is important to study the genetic aspects of both PKR and 2'-5'OAS in leporid species. In Primates, the PKR-K3L system showed remarkable evidence that the host protein can overcome viral mimicry through evolutionary strategies that result in a molecular 'arms race' reflected in host protein regions that interact with the viral protein [47].

Besides focusing on the genetic and evolutionary aspects of host immune genes in leporid species with variable susceptibilities to MYXV, studying the evolution of viral resistance in European rabbit populations prior to and after the introduction of MYXV

will allow us to understand the impact of recent (~50 years) intense selective pressures on the species genome. The comparison of European rabbit samples from different time points, previous to and following MYXV outbreaks, might allow detection of genomic regions with evidence of selective sweeps, a genetic mechanism predicted to occur in the surroundings of genomic loci that are under positive selection in a population [48]. After identifying these regions, we could then screen for genes potentially responsible for the resistance to MYXV infection. This approach is likely to advance our understanding of the genetic basis of resistance in certain European rabbit populations in Australia. For such studies, it would be crucial to obtain samples from locations where resistance to myxomatosis has been previously described.

While conducting the studies included in this dissertation, we were only able to collect samples from one of the natural long-term MYXV hosts, the brush rabbit (*Sylvilagus bachmani*). It will be important in the future to obtain samples from the tapeti (*Sylvilagus brasiliensis*), the South American host of MYXV, to complete the comparative approach performed throughout this work.

Once MYXV lost pathogenicity in its *Sylvilagus* hosts, it likely exerted no further selective pressure on these hosts, and elicited no additional adaptive changes in crucial host immune genes. To confirm these assumptions, it would be useful to perform studies to date the co-evolution between *Sylvilagus* hosts and the natural circulating strains of MYXV. The 'arms race' between MYXV and its two *Sylvilagus* host species is expected to have slowed down considerably.

3. References

1. Kerr P, McFadden G: Immune responses to myxoma virus. *Viral Immunology* 2002, 15:229-246.
2. Stanford MM, Werden SJ, McFadden G: Myxoma virus in the European rabbit: interactions between the virus and its susceptible host. *Veterinary Research* 2007, 38:299-318.
3. Liu J, Wennier S, McFadden G: The immunoregulatory properties of oncolytic myxoma virus and their implications in therapeutics. *Microbes and Infection* 2010, 12:1144-1152.
4. Spiesschaert B, McFadden G, Hermans K, Nauwynck H, Van de Walle GR: The current status and future directions of myxoma virus, a master in immune evasion. *Veterinary Research* 2011, 42:76.
5. Kerr PJ: Myxomatosis in Australia and Europe: a model for emerging infectious diseases. *Antiviral Research* 2012, 93:387-415.
6. Zuniga MC: A pox on thee! Manipulation of the host immune system by myxoma virus and implications for viral-host co-adaptation. *Virus Research* 2002, 88:17-33.
7. Rossi D, Zlotnik A: The biology of chemokines and their receptors. *Annual Review of Immunology* 2000, 18:217-242.
8. Laing KJ, Secombes CJ: Chemokines. *Developmental & Comparative Immunology* 2004, 28:443-460.
9. Fernandez EJ, Lolis E: Structure, function, and inhibition of chemokines. *Annual Review of Pharmacology and Toxicology* 2002, 42:469-499.
10. Moser B, Willimann K: Chemokines: role in inflammation and immune surveillance. *Annals of the Rheumatic Diseases* 2004, 63 Suppl 2:ii84-ii89.
11. Mantovani A: The chemokine system: redundancy for robust outputs. *Immunology Today* 1999, 20:254-257.
12. Zlotnik A, Yoshie O: The chemokine superfamily revisited. *Immunity* 2012, 36:705-716.
13. Carmo CR, Esteves PJ, Ferrand N, van der Loo W: Genetic variation at chemokine receptor CCR5 in leporids: alteration at the 2nd extracellular domain by gene conversion with CCR2 in *Oryctolagus*, but not in *Sylvilagus* and *Lepus* species. *Immunogenetics* 2006, 58:494-501.

14. Abrantes J, Carmo C, Matthee C, Yamada F, Loo W, Esteves P: A shared unusual genetic change at the chemokine receptor type 5 between *Oryctolagus*, *Bunolagus* and *Pentalagus*. *Conservation Genetics* 2011, 12:325-330.
15. Samson M, LaRosa G, Libert F, Paindavoine P, Detheux M, Vassart G, Parmentier M: The second extracellular loop of CCR5 is the major determinant of ligand specificity. *The Journal of Biological Chemistry* 1997, 272:24934-24941.
16. Chefetz I, Ben Amitai D, Browning S, Skorecki K, Adir N, Thomas MG, Kogleck L, Topaz O, Indelman M, Uitto J, Richard G, Bradman N, Sprecher E: Normophosphatemic familial tumoral calcinosis is caused by deleterious mutations in SAMD9, encoding a TNF-alpha responsive protein. *Journal of Investigative Dermatology* 2008, 128:1423-1429.
17. Tanaka M, Shimbo T, Kikuchi Y, Matsuda M, Kaneda Y: Sterile alpha motif containing domain 9 is involved in death signaling of malignant glioma treated with inactivated Sendai virus particle (HVJ-E) or type I interferon. *International Journal of Cancer* 2010, 126:1982-1991.
18. Hershkovitz D, Gross Y, Nahum S, Yehezkel S, Sarig O, Uitto J, Sprecher E: Functional characterization of SAMD9, a protein deficient in normophosphatemic familial tumoral calcinosis. *Journal of Investigative Dermatology* 2011, 131:662-669.
19. Liu J, Wennier S, Zhang L, McFadden G: M062 is a host range factor essential for myxoma virus pathogenesis and functions as an antagonist of host SAMD9 in human cells. *Journal of Virology* 2011, 85:3270-3282.
20. Pappas DJ, Coppola G, Gabatto PA, Gao F, Geschwind DH, Oksenberg JR, Baranzini SE: Longitudinal system-based analysis of transcriptional responses to type I interferons. *Physiological Genomics* 2009, 38:362-371.
21. Kawai T, Akira S: Innate immune recognition of viral infection. *Nature Immunology* 2006, 7:131-137.
22. Loo YM, Gale M, Jr.: Immune signaling by RIG-I-like receptors. *Immunity* 2011, 34:680-692.
23. Dixit E, Kagan JC: Intracellular pathogen detection by RIG-I-like receptors. *Advances in Immunology* 2013, 117:99-125.
24. Wlasiuk G, Nachman MW: Adaptation and constraint at Toll-like receptors in primates. *Molecular Biology and Evolution* 2010, 27:2172-2186.
25. Alcaide M, Edwards SV: Molecular evolution of the toll-like receptor multigene family in birds. *Molecular Biology and Evolution* 2011, 28:1703-1715.
26. Areal H, Abrantes J, Esteves PJ: Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evolutionary Biology* 2011, 11:368.

27. Wang F, Gao X, Barrett JW, Shao Q, Bartee E, Mohamed MR, Rahman M, Werden S, Irvine T, Cao J, Dekaban GA, McFadden G: RIG-I mediates the co-induction of tumor necrosis factor and type I interferon elicited by myxoma virus in primary human macrophages. *PLoS Pathogens* 2008, 4:e1000099.
28. Katzourakis A, Tristem M, Pybus OG, Gifford RJ: Discovery and analysis of the first endogenous lentivirus. *Proceedings of the National Academy of Sciences of the United States of America* 2007, 104:6261-6265.
29. Keckesova Z, Ylinen LM, Towers GJ, Gifford RJ, Katzourakis A: Identification of a RELIK orthologue in the European hare (*Lepus europaeus*) reveals a minimum age of 12 million years for the lagomorph lentiviruses. *Virology* 2009, 384:7-11.
30. van der Loo W, Abrantes J, Esteves PJ: Sharing of endogenous lentiviral gene fragments among leporid lineages separated for more than 12 million years. *Journal of Virology* 2009, 83:2386-2388.
31. Schaller T, Hue S, Towers GJ: An active TRIM5 protein in rabbits indicates a common antiviral ancestor for mammalian TRIM5 proteins. *Journal of Virology* 2007, 81:11713-11721.
32. Fletcher AJ, Hue S, Schaller T, Pillay D, Towers GJ: Hare TRIM5alpha restricts divergent retroviruses and exhibits significant sequence variation from closely related lagomorpha TRIM5 genes. *Journal of Virology* 2010, 84:12463-12468.
33. Sawyer SL, Wu LI, Emerman M, Malik HS: Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102:2832-2837.
34. Song B, Gold B, O'Huigin C, Javanbakht H, Li X, Stremlau M, Winkler C, Dean M, Sodroski J: The B30.2(SPRY) domain of the retroviral restriction factor TRIM5alpha exhibits lineage-specific length and sequence variation in primates. *Journal of Virology* 2005, 79:6111-6121.
35. Johnson WE, Sawyer SL: Molecular evolution of the antiretroviral TRIM5 gene. *Immunogenetics* 2009, 61:163-176.
36. Soares EA, Menezes AN, Schrago CG, Moreira MA, Bonvicino CR, Soares MA, Seuanez HN: Evolution of TRIM5alpha B30.2 (SPRY) domain in New World primates. *Infection, Genetics and Evolution* 2010, 10:246-253.
37. Yap MW, Stoye JP: Apparent effect of rabbit endogenous lentivirus type K acquisition on retrovirus restriction by lagomorph Trim5alphas. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2013, 368:20120498.

38. Mossman K, Upton C, McFadden G: The myxoma virus-soluble interferon-gamma receptor homolog, M-T7, inhibits interferon-gamma in a species-specific manner. *The Journal of Biological Chemistry* 1995, 270:3031-3038.
39. Sen GC: Viruses and interferons. *Annual Review of Microbiology* 2001, 55:255-281.
40. Johnston JB, McFadden G: Poxvirus immunomodulatory strategies: current perspectives. *Journal of Virology* 2003, 77:6093-6100.
41. Chang HW, Watson JC, Jacobs BL: The E3L gene of vaccinia virus encodes an inhibitor of the interferon-induced, double-stranded RNA-dependent protein kinase. *Proceedings of the National Academy of Sciences of the United States of America* 1992, 89:4825-4829.
42. Davies MV, Chang HW, Jacobs BL, Kaufman RJ: The E3L and K3L vaccinia virus gene products stimulate translation through inhibition of the double-stranded RNA-dependent protein kinase by different mechanisms. *Journal of Virology* 1993, 67:1688-1692.
43. Sharp TV, Moonan F, Romashko A, Joshi B, Barber GN, Jagus R: The vaccinia virus E3L gene product interacts with both the regulatory and the substrate binding regions of PKR: implications for PKR autoregulation. *Virology* 1998, 250:302-315.
44. Carroll K, Elroy-Stein O, Moss B, Jagus R: Recombinant vaccinia virus K3L gene product prevents activation of double-stranded RNA-dependent, initiation factor 2 alpha-specific protein kinase. *The Journal of Biological Chemistry* 1993, 268:12837-12842.
45. Ramelot TA, Cort JR, Yee AA, Liu F, Goshe MB, Edwards AM, Smith RD, Arrowsmith CH, Dever TE, Kennedy MA: Myxoma virus immunomodulatory protein M156R is a structural mimic of eukaryotic translation initiation factor eIF2alpha. *Journal of Molecular Biology* 2002, 322:943-954.
46. Rahman MM, Liu J, Chan WM, Rothenburg S, McFadden G: Myxoma virus protein M029 is a dual function immunomodulator that inhibits PKR and also conscripts RHA/DHX9 to promote expanded host tropism and viral replication. *PLoS Pathogens* 2013, 9:e1003465.
47. Elde NC, Child SJ, Geballe AP, Malik HS: Protein kinase R reveals an evolutionary model for defeating viral mimicry. *Nature* 2009, 457:485-489.
48. Nielsen R: Molecular signatures of natural selection. *Annual Review of Genetics* 2005, 39:197-218.