UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS BIOLÓGICAS

TESIS DOCTORAL

Diversidad de patógenos en aves silvestres neotropicales: estrategias de descubrimiento e identificación del papel de las especies hospedadoras

Diversity of pathogens in Neotropical wild birds: discovery strategies and identification of host species roles

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Daniel Alejandro Truchado Martín

DIRECTORES

Laura Benítez Rico
Javier Pérez Tris

Madrid

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS BIOLÓGICAS

TESIS DOCTORAL

Diversidad de patógenos en aves silvestres neotropicales: estrategias de descubrimiento e identificación del papel de las especies hospedadoras

Diversity of pathogens in Neotropical wild birds: discovery strategies and identification of host species roles

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Daniel Alejandro Truchado Martín

DIRECTORES

Laura Benítez Rico

Javier Pérez Tris

La presente Tesis Doctoral ha sido financiada mediante un contrato predoctoral de personal investigador en formación (CT27/16-CT28/16) concedido por la Universidad Complutense de Madrid, los proyectos CGL2013-41642-P/BOS y CGL2017-82117-P de la Agencia Estatal de Investigación y por lo proyectos CEBA2015-TROCSYMB y CEBA2016-HUMMAL del programa *Investissement d'avenir* concedidos por la Agencia Nationale de la Recherche.

Nullius in verba

ACKNOWLEDGEMENTS (agradecimientos)

En primer lugar, he de dar las gracias a mis directores, Laura y Javi, por confiar en mí para llevar a cabo este bonito proyecto. A Laura, por haber sido como mi madre científica transmitiéndome la pasión por la virología y el trabajo bien hecho, por tu paciencia en enseñarme desde el principio todo lo que no sabía y por preocuparte también en el plano humano por mi bienestar (especialmente en evitar que me desmayara de hambre si las reuniones se alargaban más allá de las 14:00h). Y a Javi, por hacerme ver las cosas desde una perspectiva en la que nadie había pensado antes y por haberme dado la oportunidad de viajar a uno de los lugares que jamás voy a olvidar como es Nouragues.

Muchas gracias también a Izaskun, Jose Manuel, Belén, Pablo y todas aquellas personas de dentro y fuera de la facultad que con vuestro trabajo me habéis permitido y ayudado a avanzar de forma más eficiente en esta investigación. Y gracias a Helena y Lucía por su revisión de erratas, un trabajo muy tedioso en el que agradezco especialmente vuestra ayuda.

En el plano personal, no puede faltar mi agradecimiento a mi Equipo Abantos, quienes habéis sido como mi familia durante esta etapa. Lucía, gracias (de nuevo) por estar siempre pendiente de mí tanto en el laboratorio como fuera de él, por esa sinergia que tenemos trabajando juntos o por estar ahí tanto en los momentos malos como en los buenos (como nuestros miércoles locos para desconectar de tanta malaria). David, gracias por todos los momentos mágicos que hemos vivido (y los que nos quedan), por nuestras pausas del café o todas nuestras ideas rocambolescas en general. Estoy muy orgulloso de que seas el próximo doctorando del laboratorio 6 de micro, ¡ya verás qué rápido pasa el tiempo entre virus! Y a Elena, mi compañera de redes sociales y mi super batwoman favorita con la cual he ido creciendo como científico desde que llegué al departamento. Nunca olvidaré cómo ganamos juntos aquel concurso de la Tesis en 3 minutos en su edición de 2019, ¡qué mezcla de emociones! Espera, ¿han quitado el cronómetro?... Por supuesto, no me puedo dejar fuera a todos los doctorandos o doctores de espíritu joven de Zoología con los que he vivido un ambiente laboral inigualable: Pablo, Gabri, Héctor, Antón, Javi y Amparo, de los cuales he aprendido muchas cosas y con quienes las comidas de tupper y los momentos fuera de la facultad nos aliviaban de nuestra estresante existencia.

Igualmente agradecido estoy a mis amigos de fuera de la facultad. A Tania, por ser mi compañera científica y de mil cosas en la vida, por todos los momentos en los que nos hemos “llorado” cronológicamente (incluso estando en países diferentes) que han facilitado que no desistiera en los momentos más difíciles. A Bea y Noe, por darme apoyo y siempre estar ahí para escuchar mis explicaciones científicas y por los buenos ratos que hemos pasado juntxs, como vuestra visita a Lund mientras estaba de estancia.

No podía olvidar a mi familia, en especial a mis padres y a mi hermano, quienes han visto de cerca mis éxitos y mis momentos de estrés pero siempre me han apoyado pase lo que pase y se han indignado los primeros al ver las condiciones en las que trabajamos en la ciencia en España.

Y por último pero no menos importante, a Sergio. Aunque hayas llegado en la última etapa de esta tesis, has sido uno de mis pilares principales, presumiendo de mis logros y apoyándome en la recta final, cosa nada fácil pero que has sabido hacer a la perfección. Con gente como tú da gusto compartir cualquier momento en la vida.

INDEX

Abstract 1

Resumen 3

List of abbreviations 6

General introduction 7

Objectives 13

General methods 14

General results 25

Chapter 1: Comparative metagenomics of Palearctic and Neotropical avian cloacal viromes reveal geographic bias in virus discovery 26

Chapter 2: A novel group of avian astroviruses from Neotropical passerine birds broaden the diversity and host range of *Astroviridae* 62

Chapter 3: A novel and divergent gyrovirus with unusual genomic features detected in wild passerine birds from a remote rainforest in French Guiana 81

Chapter 4: Identifying divergent host roles for avian malaria in a community of Neotropical understory passerine birds 99

General discussion 112

Conclusions 118

Future research 120

References 121

ABSTRACT

Introduction

With this PhD thesis, we want to shed some light on the field of pathogens circulating in wildlife, more specifically, in wild birds. In a global context where pathogen transmission among wildlife, domestic animals and humans is higher than ever before, studies following a discovery-driven approach are essential for human health and biodiversity conservation. However, with some exceptions, few pathogens have been studied in wild birds. In our research, we have focused on two important groups of pathogens carried by this group of animals: viruses and haemosporidians, two models that allow different perspectives in the study of avian pathogens, from discovery to ecological function. Furthermore, we want to highlight the value of remote regions as sources of novel information about pathogen diversity and discovery-driven approaches as essential tools for their study. To this end, we sampled the understory community of wild birds in a tropical rainforest of the Nouragues Natural Reserve (French Guiana) and we analyzed its cloacal virome and the community of haemosporidians infecting them.

Objectives

The general objective of this PhD thesis is to highlight how discovery-driven research on pathogens of wildlife living in remote regions contributes substantially to expand the knowledge in the fields of virology and parasitology. With this information, we will improve the understanding of the diversity, host range, ecology and prevalence of both cloacal viruses and haemosporidians infecting birds from Nouragues. Therefore, the four main objectives of this thesis can be summarized in biodiscovery, screening of pathogens in Guianan birds, highlight remote regions and non-traditional hosts as sources of relevant new information in viral discovery and the analysis of avian malaria parasites in Nouragues.

Methodology

In order to characterize the cloacal virome and the community of avian haemosporidians, we have utilized the state-of-the-art techniques available. For virome analysis, we sampled birds using cloacal swabs that were subsequently subjected to deep sequencing. In the case of avian malaria parasites of Nouragues, we implemented the most recent and

sensitive methods of molecular screening to correctly describe their patterns of distribution and host range.

Results and conclusions

Biodiscovery

In chapter 1, we showed that by sampling non-traditional hosts living in remote regions, we expand the available knowledge about viral diversity. In chapters 2 and 3 we described in detail four novel astroviruses that constitute a divergent branch within the family *Astroviridae* and a putative new species, and a novel gyrovirus (gyrovirus 11; GyV11) which is also classified as a divergent branch with respect to the other gyroviruses. In chapter 4, we reported 11 different *Plasmodium* lineages infecting passerines of Nouragues, 4 of which are new lineages.

Screening of pathogens in Guianan birds

In chapter 2, we showed how passerines play an important role in the circulation of astroviruses within the avian community of Nouragues. In chapter 3, we showed that the prevalence of gyrovirus 11 (GyV11) is low in the same analyzed bird community and that its transmission is likely mediated by the fecal-oral route. In chapter 4, we showed that avian malaria parasites are diverse and common in undergrowth birds of the rainforest at Nouragues.

Highlight remote regions and non-traditional hosts as sources of relevant new information in viral discovery

In chapter 1, we demonstrate that, by sampling non-traditional hosts such as passerines, we obtained a diversity of viruses which was previously unknown. Moreover, we showed how viruses from Nouragues (remote region) presented nucleotide sequences more different from reference sequences in databases than viruses from La Herrería (more researched area).

Analysis of avian malaria parasites in Nouragues

In chapter 4, we demonstrate that passerine species from Nouragues were unevenly affected by avian malaria. There are divergent host roles regarding the disease, with a few species concentrating the majority of *Plasmodium* infections and some species avoiding the infection.

Sampling understudied hosts inhabiting remote regions is a prerequisite for expanding substantially the knowledge about diversity of viruses and parasites in birds. Furthermore, discovery-driven approaches constitute an efficient tool to speed up the description of novel viruses and parasite carried out by wild animals.

RESUMEN

Introducción

Con esta tesis doctoral se pretende arrojar luz en el campo de los patógenos que circulan en la fauna silvestre, más en concreto, en las aves silvestres. En un contexto global en el que la transmisión de patógenos entre fauna silvestre, animales domésticos y humanos es mayor que nunca, los estudios con aproximaciones enfocadas al descubrimiento de nuevos patógenos en la fauna silvestre son esenciales para la salud humana y la conservación de la biodiversidad. Sin embargo, con algunas excepciones, pocos son los patógenos estudiados en aves silvestres. En nuestra investigación, nos hemos centrado en dos grupos importantes de patógenos que portan este grupo de animales: los virus y los hemosporidios, dos modelos que permiten diferentes perspectivas en el estudio de los patógenos de aves. Además, pretendemos resaltar el valor de las zonas remotas como fuente de nueva información en el ámbito de la diversidad de patógenos y de las aproximaciones enfocadas al descubrimiento como una herramienta esencial para su estudio. Para ello, muestreamos la comunidad de aves silvestres del sotobosque en un bosque tropical húmedo de la Reserva Natural de Nouragues (Guayana Francesa) y analizamos su viroma cloacal y la comunidad de hemosporidios que los infectan.

Objetivos

El objetivo general de esta tesis es destacar cómo la investigación dirigida al descubrimiento de nuevos patógenos en fauna silvestre y zonas remotas contribuye sustancialmente a ampliar el conocimiento en los campos de la virología y la parasitología. Con esta información, se mejorará la comprensión de la diversidad, rango de hospedadores, ecología y prevalencia tanto de los virus cloacales como de los hemosporidios que infectan a las aves de Nouragues. Por tanto, los cuatro objetivos principales de esta tesis se resumen en biodescubrimiento, el cribado de patógenos en las aves de Guayana, resaltar los hospedadores no tradicionales y las zonas remotas como

fuente de nueva información relevante en el descubrimiento de virus y el análisis de los parásitos de la malaria aviar en Nouragues.

Metodología

Para caracterizar el viroma cloacal y la comunidad de hemosporidios utilizamos las técnicas más modernas disponibles. Para el análisis del viroma, muestreamos las aves con hisopos cloacales que posteriormente fueron sometidos a secuenciación profunda. En el caso de los parásitos de malaria aviar de Nouragues, empleamos las técnicas más actuales y sensibles de cribado molecular para describir correctamente los patrones de distribución y el rango de hospedadores.

Resultados y conclusiones

Biodescubrimiento

En el capítulo 1, demostramos que, analizando hospedadores no tradicionales y zonas remotas se aumenta el conocimiento disponible de la diversidad viral. En los capítulos 2 y 3 describimos en detalle cuatro nuevos astrovirus que constituyen una rama divergente en la familia *Astroviridae* y una posible nueva especie, y un nuevo gyrovirus (gyrovirus 11; GyV11), el cual aparece también como una rama divergente del resto de gyrovirus. En el capítulo 4 reportamos 11 linajes diferentes de *Plasmodium* que infectan a los paseriformes de Nourgaues, 4 de los cuales son nuevos linajes.

Cribado de patógenos en las aves de Guayana

En el capítulo 2, mostramos cómo los paseriformes juegan un papel importante en la circulación de astrovirus en la comunidad de aves de Nouragues. En el capítulo 3, mostramos que la prevalencia de gyrovirus 11 (GyV11) es baja en la misma comunidad de aves y que su transmisión mediante ruta fecal-oral es muy probable. En el capítulo 4 demostramos que los parásitos causantes de malaria aviar son diversos y comunes en los pájaros del sotobosque de la reserva de Nouragues.

Resaltar los hospedadores no tradicionales y las zonas remotas como fuente de nueva información relevante en el descubrimiento de virus

En el capítulo 1 mostramos cómo muestreando hospedadores no tradicionales como son los paseriformes se obtiene una diversidad de virus previamente desconocida. Además, demostramos que los virus de Nouragues (zona remota) son significativamente más

diferentes en sus secuencias nucleotídicas a las secuencias de referencia en bases de datos que los virus de La Herrería (zona más estudiada).

Análisis de los parásitos de la malaria aviar en Nouragues

En el capítulo 4, mostramos cómo la malaria aviar afecta de forma desigual a las especies de paseriformes de Nouragues. Existen papeles divergentes en los hospedadores en relación a esta enfermedad, habiendo unas pocas especies que concentran la mayoría de las infecciones por *Plasmodium* y alguna especie que evita la infección.

Muestrear hospedadores poco estudiados que habitan en zonas remotas es un requisito indispensable para ampliar de forma sustancial el conocimiento disponible acerca de la diversidad de virus y de parásitos de aves. Además, las aproximaciones enfocadas al descubrimiento suponen una herramienta eficiente para acelerar la descripción de nuevos virus y patógenos transportados por la fauna silvestre.

LIST OF ABBREVIATIONS

CAV	Chicken anemia virus
cytb	Cytochrome b
FGHEV	French Guiana hepevirus
FGPV	French Guiana picornavirus
FGRV	French Guiana reovirus
GyV	Gyrovirus
ICTV	International Comitee on Taxonomy of Viruses
LHAdV	La Herrería adenovirus
LHHEV	La Herrería hepevirus
NGS	Next generation sequencing
NTR	Non-translated region
ORF	Open reading frame
PasAstV	Passerine astrovirus
RdRp	RNA-dependent RNA-polymerase
UTR	Untranslated region
VTM	Viral transport medium

GENERAL INTRODUCTION

Pathogen diversity in wildlife

The knowledge about pathogen diversity in wildlife is in its infancy. Current anthropogenic impact on the environment is creating perturbations that increase the chance of pathogen transmission among wildlife, domestic animals and humans (Cunningham et al. 2017). In fact, emerging infectious diseases have become one of the main challenges of the present era (WHO 2019). Since the beginning of the twenty-first century, humanity has faced several pandemics (such as H1N1 influenza A virus, SARS, MERS, Ebola, Zika and, more recently, COVID-19) that have affected public health and the economy of the world (Ma et al. 2020; El Zowalaty and Järhult 2020). The majority of these pandemics had an origin in viruses of wildlife and have revealed our vulnerability to zoonotic transmission of unknown but dangerous pathogens circulating in wildlife, especially mammals and birds (Carroll et al. 2018). Apart from viruses, other zoonotic agents such as malaria parasites infecting apes and monkeys have been demonstrated to infect humans and cause severe disease (Baird 2009). Therefore, discovering the diversity of pathogens with zoonotic potential circulating in wild animals has become a central endeavor of the One Health initiatives (Lebov et al. 2017; Carroll et al. 2018). Wildlife can also act as a reservoir of pathogens that infect domestic animals. Avian metapneumovirus (Rizotto et al. 2019), foot and mouth disease virus (Morgan et al. 2006) or Nipah (Olival et al. 2020) and Hendra viruses (Plowright et al. 2011) are known examples of these pathogens. On the other hand, anthropogenic impact is also increasing the transmission of pathogens to wildlife in a process defined as "pathogen pollution" by some authors (Daszak et al. 2000; Adlard et al. 2015; Cunningham et al. 2017). Canine distemper in carnivores or chytridiomycosis in amphibians are two important diseases spread by human activity that have caused acute population declines of several species of wild animals (Cunningham et al. 2017). Also, introduction of novel pathogens in naïve ecosystems can dramatically affect wildlife, as in the case of poxvirus and malaria in birds of Galápagos and Hawaii where conservation of their unique avifauna can be seriously compromised (Land et al. 1986; van Riper III et al. 2002; Thiel et al. 2005). Thus, in order to efficiently increase our knowledge about pathogen diversity in wildlife and how they are present in wild animals from different ecosystems, it is crucial to expand the scope of research to novel hosts and remote regions using discovery-driven approaches to that end.

Importance of birds in pathogen transmission

As mentioned before, birds are one of the main animal groups carrying potential zoonotic pathogens that could occasionally infect humans. With more than 10,000 recognized species and more than 18,000 predicted (Barrowclough et al. 2016), the class Aves is the most diverse group of terrestrial vertebrates. Moreover, birds are one of the most ubiquitous groups of animals as they can be found in practically all ecosystems on Earth. Their worldwide distribution and high phylogenetic diversity make them play important roles in the ecosystems where they live, providing essential services such as pollination, pest control or seed dispersal (Şekercioğlu et al. 2016). On the other hand, this global presence increases their exposure to diverse microorganisms of the environment (Benskin et al. 2009) which, together with their different ecologies, make birds good candidates for spreading pathogens that cause disease (Şekercioğlu et al. 2016). Several behavioral and physiological features of birds favor them to be the origin of pathogen spillovers. First, wild birds usually share their roosting and feeding places with other vertebrates, thus facilitating the occurrence of inter-species pathogen transmission (Chan et al. 2015). Also, migratory birds can carry with them pathogens between their wintering sites and their breeding areas, introducing and promoting the transmission of external etiological agents in the avian populations and poultry enclosures they come in contact with (Pérez-Tris and Bensch 2005; Dhama et al. 2008; Viana et al. 2016). Bacteria such as *Borrelia burgdorferi*, *Campylobacter* or *Salmonella*; haemosporidians of the genus *Plasmodium* or viruses such as West Nile virus, avian pneumovirus or Newcastle disease virus are known to be circulated by migratory birds (Reed et al. 2002; Dhama et al. 2008; Levin et al. 2013). It has also been shown that some avian species, especially ducks, may propagate large quantities of a great diversity of viruses in the environment where they live without them being affected by disease (Hulse-Post et al. 2005; Wille et al. 2018).

The most famous cases of human epidemics and pandemics with an avian origin are the outbreaks of avian influenza viruses occurred between 1997-2005 (H5N1) and in 2013 (H7N9) (Luiz Proença-Módena et al. 2007; Lai et al. 2013). Also, the H1N1 influenza A virus causing a pandemic in 2009 had some genes derived from avian influenza viruses (Smith et al. 2009) and several other subtypes of avian influenza virus have been shown to be transmitted to humans (Peiris et al. 2007). Furthermore, reports of human infections caused by other pathogens with an avian origin such as West Nile virus and Usutu virus have increased during the last years (Vilibic-Cavlek et al. 2019). On the other hand, West

Nile virus also had devastating effects on wildlife of North America, causing important declines in wild bird populations (LaDeau et al. 2007). These detrimental consequences caused by introduced pathogens are well known in Hawaii, where several species of their endemic avifauna are now extinct due to the effect of poxvirus and malaria infections, both introduced by human activity (Warner 1968).

Poultry industry represents one of the most studied examples of pathogen transmission among domestic animals, wildlife and humans, being more frequent in the form of spillovers from wild birds to poultry in farms or in living bird markets (Kim et al. 2012; Prosser et al. 2013; Bahl et al. 2016). In some cases, such as low pathogenic avian influenza viruses in wild aquatic birds, viruses become highly pathogenic after chickens get infected causing large losses due to the high mortality (Verhagen et al. 2015). Once poultry species are infected, close contact with poultry industry workers eases the cross-species transmission and even recombination with human strains in zoonotic viruses (Gray and Kayali 2009), multiplying the chances of an emerging viral disease outbreak. Fowlpoxvirus or avian metapneumovirus are also known to infect poultry from wild birds, causing significant negative effects (Rizotto et al. 2019; Giotis and Skinner 2019). In other words, viruses carried by wild birds are of great concern not only for human health and veterinary medicine but also for the economy as they can cause important losses due to the associated mortality in poultry industry. On the contrary, domestic birds have also a negative impact in wild birds. Several cases of reverse spillover caused by viral strains highly similar to vaccine isolates have been reported in wild birds in areas where poultry production is important. Newcastle disease virus, avian paramyxoviruses, infectious bronchitis viruses or avian coronaviruses closely related to strains used for vaccination have been detected in wild birds, increasing the risk of novel outbreaks of these diseases in wildlife (Garcia et al. 2013; Devlin et al. 2016; Rohaim et al. 2017; Xiang et al. 2017). Unfortunately, spillovers are not the only threat for wild bird populations. The introduction of domestic animals and their associated pathogens, habitat loss and climate change will alter the relationships between wildlife and pathogens in the ecosystems, having a yet unknown effect on the avifauna (Daszak et al. 2000; Thompson et al. 2010; Gallana et al. 2013).

All these examples highlight the importance of both surveillance programs for known parasites and microbial biodiscovery to be able to anticipate to novel pathogen outbreaks, especially those with zoonotic potential or of conservation concern (Daszak et al. 2000;

Carroll et al. 2018). Although the virome of poultry and other domestic birds has received attention (Mihalov-Kovács et al. 2014; Kwok et al. 2020) viruses circulating in wild bird populations remain little explored. The studies available related to the virome of wild birds are mainly focused on waterfowl as they are the natural host of several avian influenza strains (Fawaz et al. 2016; Wille et al. 2018; Zhao et al. 2018; Ramírez-Martínez et al. 2018). However, little is known about the virome of important groups such as passerines, which comprise approximately the 60% of extant birds, or wild bird communities of the tropics, where we can find the highest diversity of bird species in the world (Orme et al. 2005). Therefore, discovery-driven approaches using specific designs are needed to accurately characterize the virome of wild birds and, eventually, expand the knowledge in the field of virology.

Apart from viruses, avian malaria parasites and other haemosporidians are important pathogens carried by wild birds with a particular interest in conservation biology. Malaria is a vector-borne disease caused by protozoan parasites (order Haemosporida) infecting the liver and erythrocytes in several groups of vertebrates. In humans, malaria has been one of the most deadly diseases throughout human history and, now, is still of great concern in the tropical regions, especially in Africa (Breman 2001). Children under 5 years are the most vulnerable, representing 67% of all deaths caused by this infection in 2018 (World Health Organization 2019). Interestingly enough, avian heaemosporidian parasites have been key elements in the early stages of the research of malaria. They are grouped into four genera: *Plasmodium*, *Heamoproteus*, *Leucocytozoon* and *Fallisia*; however, research has been mainly directed towards *Plasmodium*, as it is the only genus infecting humans and the only malaria parasites *sensu stricto* (Valkiūnas et al. 2005). Although the etiological agent causing human malaria was discovered by Charles Louis Alphonse Laveran in 1880, important morphological and ecological characteristics as well as its transmission by mosquitoes were described by Vassily Danilewsky and Ronald Ross when they were studying avian heamosporidians (Cox 2010; Marzal 2012). Subsequently, bird models have played an important role in the understanding of the life cycle of malaria parasites and in the development of anti-malaria treatments and vaccines (Kalra et al. 2006; Marzal 2012). Furthermore, the abundance and host diversity range of avian haemosporidians together with the relatively ease they can be sampled have also made them a perfect model for studying host-parasite relationships (Valkiunas 2004; Marzal 2012). The identification and classification of avian haemosporidian was based

on morphological features of blood stages during a long time (Rivero and Gandon 2018). However, the development of different molecular detection techniques (Bensch et al. 2000; Fallon et al. 2003; Hellgren et al. 2004; Waldenström et al. 2004; Bell et al. 2015; Ciloglu et al. 2019) started a new era in this field, allowing the flourishing of publications and, therefore, the creation of MalAvi, a global database of avian haemosporidians with detailed information regarding infections all over the world (http://130.235.244.92/Malavi/) (Bensch et al. 2009). To this day, MalAvi represents the reference database for avian malaria parasites as it catalogues all lineages described (Rivero and Gandon 2018).

Since the discovery of rodent malaria in 1948, avian hemosporidian parasites have taken a second place in the research of human malaria. However, the diversity of PCR protocols available for their detection progressively changed the aim of the studies with avian haemosporidians towards host-parasite relationships (Fecchio et al. 2020b). The astounding diversity of lineages, their presence in all avian clades and regions on Earth (except in Antarctica due to the absence of mosquitoes) or the existence of specialist and generalist lineages are the main reasons why these parasites are the perfect model for research on host-parasite interactions (Rivero and Gandon 2018; Fecchio et al. 2020b). The most generalist lineages usually belong to the genus *Plasmodium* (Valkiunas 2004) while *Haemoproteus* and *Leucocytozoon* lineages tend to be more specialist (Fecchio et al. 2020b), although several *Haemoproteus* lineages have been reported to be generalist in megadiverse regions of South America (Moens and Pérez-Tris 2016). Regarding prevalence, *Plasmodium* and *Haemoproteus* are the most world-wide distributed and abundant in bird populations, likely being this the reason why studies on avian heamosporidians are highly biased towards these two genera. *Leococytozoon* is more prevalent in temperate regions and research on *Fallisia* is scarce, with only one species described (Valkiunas 2004). All these parasite infections frequently have a negative impact on bird populations, decreasing survival rates and breeding success (Marzal et al. 2005; Martínez-De La Puente et al. 2010). These adverse effects can be even more dramatic in naïve populations where the parasite is new to the community (Atkinson and Samuel 2010). Hence, transmission and dynamics of heamosporidian parasites, as well as other pathogens, should be taken into account for avian diversity conservation in order to avoid disease impacts on fitness of threatened species.

How can we discover novel pathogens carried by birds?

If we are to gain a complete knowledge of the diversity of pathogens that threaten wildlife and human health, what should we do not to find the same pathogens once and again? A correct description of nature is essential in order to know the natural history of pathogens, the ecological relationships among the different species and their ecosystemic functions (Greene 2005). Missing this information will hamper the correct understanding of the whole system, so we need factual knowledge of particular organisms as a basis for a correct wildlife conservation and a good human health. Information about pathogens of wildlife is especially valuable when research is carried out in remote regions, where data is very limited or absent and a wide variety of yet-unknown pathogens circulate. Among all remote regions on Earth, low latitude tropical areas are known to contain a higher species richness of multicellular organisms, human pathogens and other microorganisms than any other region in the planet (Guernier et al. 2004; Sherratt and Wilkinson 2009). This increase of biodiversity towards the Equator seems to be also true for haemosporidian parasites lineages of the genera *Plasmodium* and *Haemoproteus* (Clark et al. 2014) but no study has dealt with viral diversity of wild birds in those places from an ecological point of view. Moreover, tropical regions have been predicted as hotspots of emerging infectious diseases with wildlife origin while research and surveillance for emerging pathogens have been mainly focused in higher-latitude countries (Jones et al. 2008). At the same time, more sampling effort is needed in tropical areas for avian haemosporidians to expand the knowledge about lineage diversity in these areas and clarify the global phylogeny of this group of parasites (Fecchio et al. 2020b). Then, these tropical remote regions are perfect candidates to harbor an unknown diversity of novel pathogens and host-pathogen relationships yet to be discovered.

One of such remote tropical regions is French Guiana, located in the Guianan shield, in the north of South America. This Neotropical region is considered one of the main hotspots of avian diversity (Thiollay 2002) although few studies deal with pathogens of wild birds in this region (Bartlett and Bain 1987). In this thesis, we describe the cloacal virome and the avian haemosporidian community present in a wild bird community from The Nouragues Natural Reserve (French Guiana) to broaden the knowledge related to pathogens carried by wildlife in remote regions. Using a discovery-driven approach to that end, we describe and characterize new viruses and haemosporidian lineages as well as novel host-pathogen relationships unknown so far. All this information will provide us

with important data on the diversity of avian pathogens in a remote and pristine tropical area where potential candidates for causing future disease outbreaks may circulate and scientific research is scarce to date.

OBJECTIVES

The main objective of this thesis is to highlight how a discovery-driven research of novel pathogens of wildlife in remote areas can substantially contribute to the existing knowledge in the fields of virology and parasitology. We set out to contribute with data that improve the understanding of the diversity, host range and prevalence of cloacal viruses and haemosporidian parasites of Neotropical wild birds. To this end, we use state-of-the-art methods that may significantly advance the knowledge of these pathogens. A logical first step is the characterization of the cloacal virome of Neotropical wild bird populations using Next generation sequencing (NGS) methods. We also implement the most recent, highly sensitive methods of molecular screening of avian haemosporidian parasites to the study of Neotropical parasites, a strategy which may improve our capacity to correctly describe their patterns of distribution among host species and prevalence in these bird communities. Thus, the four main objectives of this PhD thesis were biodiscovery, the screening of pathogens in Guianan birds, highlight remote regions and non-traditional hosts as sources of relevant new information in viral discovery and the analysis of avian malaria parasites in Nouragues.

Biodiscovery

In all four chapters, the thesis has a strong discovery component. The main objective of chapter 1 was to describe the diversity of viruses carried by a group of non-traditional hosts such as two wild bird communities where passerines prevailed. Also, we wanted to characterize, with the highest possible level of detail, those viruses with special interest for veterinary science, human health or ecology.

Screening of pathogens in Guianan birds

In order to know which viruses and malaria parasites circulate within the unique and diverse avifauna of the Nouragues Natural Reserve, we carried out prevalence analyses of novel viruses and avian malaria parasites in chapters 2, 3 and 4. In the case of avian haemosporidians, we used the most sensitive methods available to produce more accurate estimates of their prevalence in Neotropical bird communities.

Highlight remote regions and non-traditional hosts as sources of relevant new information in viral discovery

In chapter 1, we compared the cloacal virome of wild bird populations of a remote and pristine region (the Nouragues Natural Reserve) with the cloacal virome of wild birds of a more studied area (La Herrería forest) to draw attention to the importance of sampling in remote regions and in understudied hosts as sources of relevant novel information in virus diversity.

Analysis of avian malaria parasites in Nouragues

In chapter 4, we carried out a screening analysis of avian malaria parasites in wild passerines of the Nouragues Natural Reserve with emphasis on the identification of divergent ecological roles of host species in the bird community, from key reservoir hosts for malaria to parasite avoiders.

GENERAL METHODS

Sampling

We sampled the understory population of birds of the Nouragues Natural Reserve (French Guiana) in order to analyze the pathogens they carried, focusing on two important groups: viruses and haemosporidian parasites. Moreover, we sampled the bird community of La Herrería forest (Spain), to compare the viruses they carried with those in Nouragues. The Nouragues Natural Reserve is located in a pristine tropical rainforest in northern South America while La Herrería is a broadleaved forest located at 900 m.a.s.l. in the center of the Iberian Peninsula whose bird populations have been more studied (Figure 1). For the analysis of the avian virome, we did not discriminate any group of birds as we carried out an exploratory study of viruses in a group of animals never sampled before. On the other hand, we selected only passerines for the study of avian haemosporidian parasites as they are the prevailing group in Nouragues and they also concentrate the majority of haemosporidian infections reported to date (http://130.235.244.92/Malavi/). Birds were mist-netted in January 2016 (rainy season) and October-November 2016 (dry season) in Nouragues and in April-July 2018 in La Herrería. They were taken standard morphometric measurements, ringed to avoid repeated sampling of the same individuals, and released unharmed at the site of capture. Blood samples (Nouragues) were taken by

venipuncture and the volume of blood extracted depended on the body size of the bird (always < 1% of their body mass) (Carpenter and Campbell 1988). The volume of blood was preserved in two different media: pure ethanol, for haemosporidian analysis, and RNAlater (Ambion, Life Technologies, Carlsbad, CA, USA), for virus analysis (Table 1). Also cloacal (Nouragues and La Herrería) and oral (Nouragues) samples were collected using sterile swabs (Nerbe Plus), which were preserved in 800 µl of Viral transport medium (VTM) (Becton Dickinson). We used the thinnest sterile swabs available due to the small size of some birds. However, cloacal samples were not taken when the size of the bird was too small (as in the case of hummingbirds) or when we considered it could result in any pain for the bird. Blood samples in ethanol were kept at room temperature in the field and at -20ºC until processing in the lab. Samples in VTM and in RNAlater were immediately frozen until molecular analyses. A total of 616 individuals belonging to 25 avian families were sampled in Nouragues and 215 birds of 13 families were sampled in La Herrería. Capture, sampling and transport of samples were authorized by the Service of Natural Environments, Biodiversity, Sites and Landscapes, Regional Directorate for the Environment, Planning and Housing at French Guiana (license 030418) and the General Directorate for the Environment of Madrid (license 10/209664.9/18). The experimental protocols were approved by the Committee on Animal Testing of Complutense University (CEA-UCM, authorisation number 44-2016).

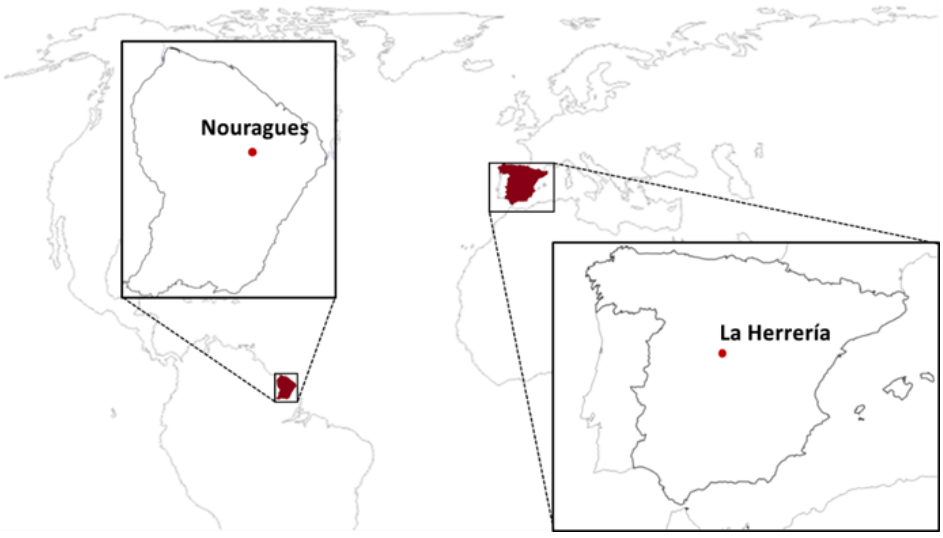


Figure 1. Location of the two sampling areas: the Nouragues Natural Reserve (French Guiana) and La Herrería forest (Spain)

Table 1. Type of samples and number of individuals analyzed in each of the chapters of the present PhD thesis

Chapter	Objective of the analysis	Sampling area	Sample type	n
1	Diversity and prevalence of astroviruses	Nouragues	Cloacal swabs	406
2	Detection and prevalence of Gyrovirus 11	Nouragues	Cloacal and oral swabs and blood in RNAlater	406
3	Comparative analysis of the cloacal virome of wild birds	Nouragues and La Herrería	Cloacal swabs	50 from Nouragues and 50 from La Herrería
4	Diversity and prevalence of avian malaria	Nouragues	Blood in ethanol	445

Viral discovery protocol

Traditionally, studies on the virome of birds have been scarce and mainly focused on species with special relevance for the economy or public health such as poultry or waterbirds (François and Pybus 2020). Thus, it is paramount to sample non-traditional hosts in remote areas in order to discover novel viruses carried by wild birds that substantially broaden the knowledge in viral diversity. In this PhD thesis, we wanted to think out of the box and analyze the virome of a community of wild birds inhabiting a remote tropical rainforest in French Guiana. This community is dominated by passerines, a group which represents 60% of extant birds but whose virome has never been analyzed. Also, French Guiana is considered one of the main hotspots of avian diversity in the world. However little is known about pathogens carried out by wild birds of this remote region where scientific research in birds has been very limited to date.

The virome of wild birds, as well as wild animals in general, has been little analyzed so we expected that the great majority of viral sequences we would obtain would correspond to novel viruses still unknown. Moreover, viruses do not have a universal genetic marker to identify them directly and they do not even share the same nucleic acids (there are DNA and RNA viruses). Therefore, we decided to follow an approach based on deep

sequencing as the best option to analyze a completely unknown virome as it has been the main tool utilized for the discovery of novel avian viral sequences in the last years (Delwart 2012; Bodewes 2018). Also, we decided to start with cloacal samples as we expected a higher diversity and abundance of viruses than in oral swabs or in blood. To maximize the number of individuals of different bird species subjected to deep sequencing we selected 50 individual cloacal samples with abundant fecal matter from as many different bird species as possible and we created five pools of 10 individuals each. Before being subjected to deep sequencing, we carried out a pretreatment in order to increase the proportion of viral nucleic acids. First, individual cloacal samples were vortexed, and the swabs were squeezed to release epithelial cells. Then, the samples were centrifuged and pellets were resuspended in PBS before being subjected to 2 freeze-thaw cycles to lyse the epithelial cells and maximize the release of viral particles. After that, a filtration through 0.45 µm pore-sized column filters enriched viral particles in the flow-through and removed cellular debris. We took 50 µl of the filtrate of each sample and mixed them in five pools of 10 samples each. Each filtrate was treated with a mixture of nucleases to digest host DNA/RNA and other unprotected nuclear acids. Finally, viral RNA/DNA was extracted with the MagMAX Viral RNA Isolation Kit (Thermo Fisher) according to the manufacturer's instructions. The protocol was adjusted taking into consideration the small amount of cells present on the swabs. Also, we ensured to keep remnant volumes of sample in the most critical steps (Figure 2).

After this pretreatment, libraries were prepared by subjecting the extracted RNA/DNA to a random PCR using QIAseq FX DNA Library Kit (Qiagen, Germany). Normalized samples were pooled and sequenced using 600-cycle (300 bp paired-end) MiSeq Reagent Kits v3 (Illumina, San Diego, CA) on a MiSeq platform. The generated raw reads were first qualitatively checked, trimmed and filtered to remove polyclonal and low quality reads (< 55 bases long) using CLC workbench (Qiagen). The remaining filtered raw reads were de-novo assembled separately using Trinity v2.6.642 (Grabherr et al. 2011) and CLC workbench and compared with a non-redundant and viral proteome database using BLASTx with an E-value cut-off of 0.001. The virus-like contigs and singlets were further compared to all protein sequences in non-redundant protein databases with a default E-value cutoff of 0.001. The viral metagenomics output has been visualized and analyzed in MEGAN (Huson et al. 2016). We followed the same protocol in the case of the deep sequencing analysis of La Herrería cloacal virome.

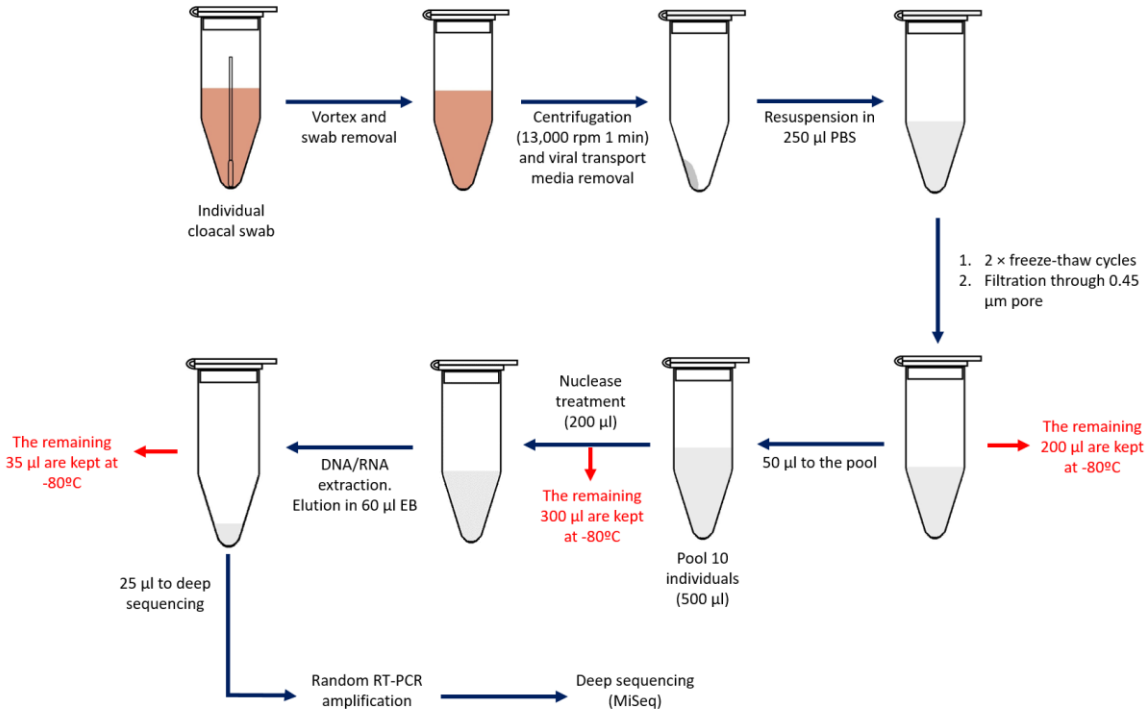


Figure 2. Protocol followed during the pretreatment of the samples for deep sequencing analysis.

Once we obtained the output of the deep sequencing analysis and we had contigs long enough to work with, we designed several primer pairs to identify by PCR/RT-PCR the birds of the pool carrying a virus of interest. We tested several primer pairs and optimized the PCR/RT-PCR protocol for each of the selected viruses using the remnant of the pooled RNA/DNA extractions as template. Thus, we did not spend the individual samples while we also identified the pool or pools containing the viral sequence we were searching. One of the main problems in this step was that the total volume of individual samples was very little and nucleic acid concentration in the extractions is usually very low, so we tried to minimize the use of individual samples. After the optimization of the PCR/RT-PCR and having found the positive pools, the individual samples which compose the positive pools were tested to identify the birds carrying the virus of interest. Finally, pools of 10 individuals were created using the cloacal samples from individuals not selected for deep sequencing to carry out the community-level prevalence analysis for the viruses in the whole set of samples. This procedure for community-level prevalence was carried out also with blood and oral samples when they were checked.

Obtaining the complete genomes of the viruses carried by a pool of 10 birds is desirable to fully characterize them, but it is not very common. Dilution with other individual

samples lower the concentration of a given virus, especially when viral load is not very high. Also, sequencing depth may not be sufficient for obtaining reads from a virus with a low intensity of infection. Only in the case of virus strains present in a relatively high prevalence in the community it is possible to fully sequence their genomes. Therefore, to avoid problems with low intensities or dilution of the original viral concentrations, individual RNA/DNA extractions were subjected to a second round of deep sequencing to complete the partial genomes of interest.

In the case we obtained the complete genome, we determined the open reading frames (ORFs) as a first step using Geneious v11 (Biomatters, New Zealand), EditSeq and SeqMan tools of the DNASTAR 5.0 software package (DNASTAR, Madison, WI), and ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/). During this step, we observed that some novel viruses did not follow the genomic architecture predicted for their group, so it was necessary to translate their ORFs to obtain the putative protein they encoded. These putative amino acid sequences were analyzed by motif searching programmes such as InterPro (http://www.ebi.ac.uk/inter_pro/), Motif Scan (https://myhits.sib.swiss/cgi-bin/motif_scan), cNLS Mapper (Kosugi et al. 2009), Phyre2 (Kelley et al. 2015) TMHMM server v2.0 (Krogh et al. 2001), FoldIndex (Prilusky et al. 2005), Multicoil scoring form (Wolf et al. 1997) and NetNES 1.1 Server (La Cour et al. 2004). Conserved motifs were searched also in the nucleotidic sequence using TFBIND (Tsunoda and Takagi 1999) and GPMiner (http://gpminer.mbc.nctu.edu.tw) for putative binding sites for transcription factors, Mfold (Zuker 2003) for stem-loop structures and Tandem repeats Finder (Benson 1999) for possible nucleotide repetitions along the untranslated region. To examine the existence of recombination events we used the software RAT (https://omictools.com/rat-tool) and RDP4 (Martin and Rybicki 2000). These analyses were carried out also with viral sequences following the genomic architecture of their group to confirm the function of their putative proteins, search for binding sites within their untranslated region and discard recombinations that may explain their genomic architecture.

In parallel, the novel viral genomes have to be placed in the phylogeny of their group in order to know which viruses are their closest relatives or how divergent they are to known viral sequences. We aligned multiple nucleotide or amino acidic sequences using MUSCLE (McWilliam et al. 2013) (https://www.ebi.ac.uk/Tools/msa/muscle/), Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) or MAFFT software (Katoh and

Standley 2013). Best substitution model was selected using jModelTest 2 (Darriba et al. 2012). Maximum likelihood phylogenetic trees were inferred using MegaX (Kumar et al. 2018) PhyML v3.1 (Guindon et al. 2010) and BEAST v1.84 (Drummond and Rambaut 2007). Distance matrices were calculated using MegaX (Kumar et al. 2018). HeatMaps were created with Morpheus software (https://software.broadinstitute.org/morpheus).

Avian haemosporidian parasites study

The approach we followed for this group of parasites was different from that for viruses. Avian haemosporidian parasites have been conventionally more studied in wild birds than viruses and more information about their nucleotide sequences is available. As a result of the abundant research on avian haemosporidian parasites, several molecular detection protocols have been developed during the last 20 years (Bensch et al. 2000; Fallon et al. 2003; Hellgren et al. 2004; Waldenström et al. 2004; Bell et al. 2015; Ciloglu et al. 2019) making the work with this type of parasites easier than the work with viruses. We selected the multiplex PCR designed by Ciloglu et al. (2019) as the best screening method in our case for several reasons. First, it is a one-step protocol that discriminates the three main genera of avian heamosporidian parasites (*Haemoproteus*, *Plasmodium* and *Leucocytozoon*) in only one PCR. In a context where we ignored the prevalence of the infection, the alternative nested PCR designed by Hellgren et al. (2004) would have been more time-consuming for the screening of samples. Secondly, the multiplex PCR has been demonstrated to be more sensitive than other alternatives, especially for multiple infections (Ciloglu et al. 2019), which is important to detect positive individuals with low parasitemias. However, this multiple PCR has also some drawbacks. One of them is that the region that the primers amplify is not the consensus *cytb* sequence used to define lineages and available in MalAvi. To circumvent this, we carried out the nested PCR designed by Hellgren et al. (2004) only in individuals which resulted positive in the multiplex PCR to confirm the infection and determine the parasite lineage. However, as the multiplex PCR is more sensitive than the nested PCR, we expected that some positive individuals in the multiplex PCR were negative in the nested PCR, especially with faint or weak bands, presumably corresponding to low intensity infections. Taking this into consideration, we decided to do analyses at the parasite genus level instead of lineage level when lineage identity could not be determined in many infections. Another downside of the multiplex is that faint non-specific bands greater than 600 bp appear rarely according to the authors (Ciloglu et al. 2019). Nevertheless, we could observe that

these unspecific products were highly frequent in our sample and sometimes resulted almost as intense as bands corresponding to positive infections (Figure 3). In the case of *Haemoproteus*, we observed non-specific bands with the size of the expected amplification product for this genus (~500 bp), especially in samples of the species *Mionectes macconnelli* (Figure 3). As no band was obtained in the nested PCR for these samples, we carried out the specific PCR only with the primer pair for *Haemoproteus*, resulting all of them negative again.

Taking all this into consideration, we decided to focus on the results related to *Plasmodium* for the fourth chapter of the thesis, as we did not observe any unspecific product of the expected size of its band (377-379 bp) unlike it happened with *Haemoproteus*. Nevertheless, we carried out a specific PCR for unclear samples using only the primer pair of interest (*Plasmodium* or *Haemoproteus*, separately) as a confirmatory PCR. Only those samples clearly positive in the multiplex PCR or those unclear samples that resulted positive in the specific PCR were considered positive.

For prevalence analysis, we followed a standard ammonium acetate protocol to extract the DNA from blood samples in ethanol (Green et al. 2012). We measured the concentration of DNA in our samples to equal final concentrations to 25 ng/µl. The quality of the extracted DNA was tested by amplifying bird sexing markers (Fridolfsson and Ellegren 1999), considering "good quality" those samples amplified by this sexing PCR.

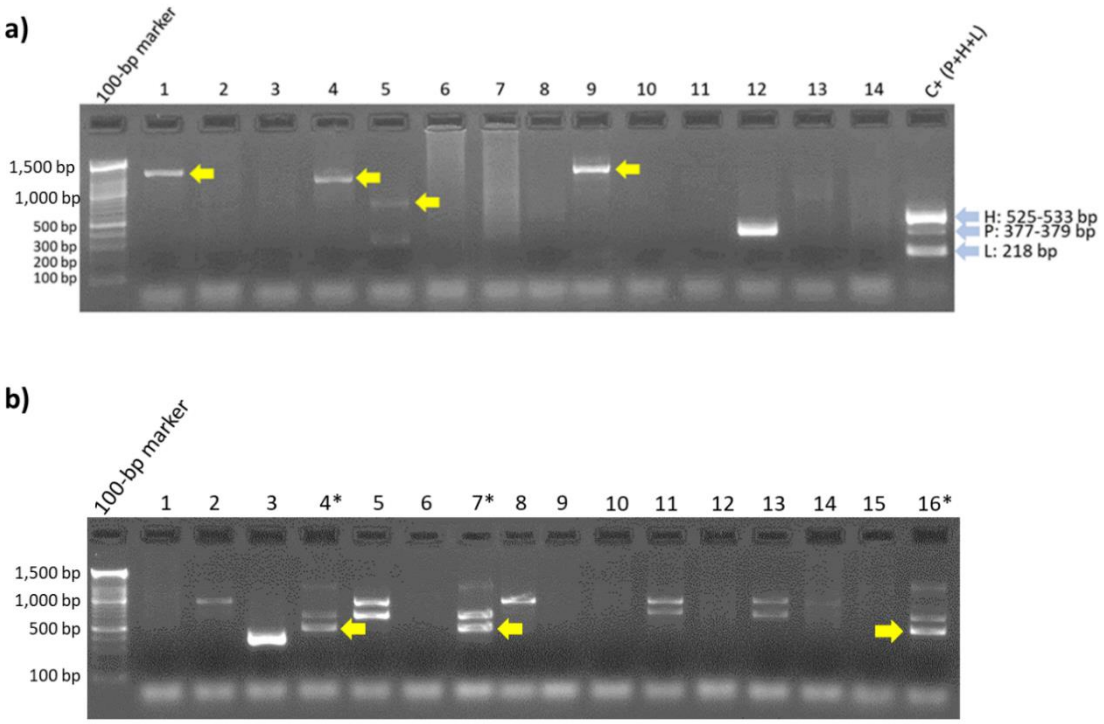


Figure 3. Agarose gel electrophoresis of multiplex PCR products of the screening of haemosporidian parasites in the samples from Nouragues. (a) Yellow arrows show non-specific products greater than 600 bp. The individual in lane 12 was positive for *Plasmodium* infection. (b) Yellow arrows show non-specific products of ~500 bp in the samples from the species *Mionectes macconnelli*, marked with asterisks. H, *Haemoproteus*; P, *Plasmodium*; L, *Leucocytozoon*; C+, positive control.

Prior to the screening of samples from Nouragues, we checked the sensitivity of the multiplex PCR using as templates all combinations of DNA extractions from individuals infected by a known parasite of the genus *Plasmodium*, *Haemoproteus* or *Leucocytozoon*. Thus, we created *ad hoc* artificial double and triple infections to test them together with the original single infections. We observed that the multiple PCR was able to detect double and triple infections. In fact, it showed that the sample believed to be a single *Plasmodium* infection was actually a double *Plasmodium* and *Leucocytozoon* infection (Figure 4).

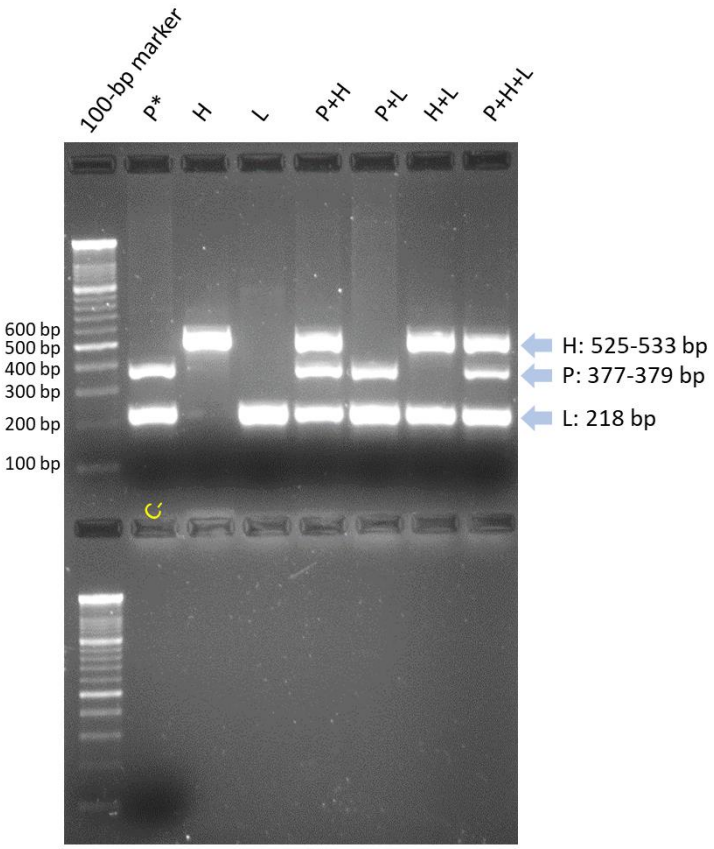


Figure 4. Agarose gel electrophoresis showing the products of the multiplex PCR obtained using different *ad hoc* combinations of single infections as templates. The asterisk shows the sample classified as single *Plasmodium* infection that resulted also positive for *Leucocytozoon* after the multiplex PCR. Due to this, all mixtures involving *Plasmodium* were also positive for *Leucocytozoon*. H, *Haemoproteus*; P, *Plasmodium*; L, *Leucocytozoon*; C-, negative control.

After equaling concentrations, checking the quality of the DNA and checking the sensitivity of the method designed by Ciloglu et al. (2019), we carried out the community-prevalence analysis using the aforementioned multiplex PCR. Faint bands not confirmed in a second test were considered as negative PCR results. Positive individuals were then subjected to the nested PCR designed by Hellgren et al. (2004).

Therefore, the protocol to study the prevalence and diversity of avian haemosporidian infection presented several challenges, different from those in the study of viral diversity due to the nature of the parasites themselves (Figure 4). But, together, they represent two different approaches to study the diversity of parasites carried by wild animals from remote regions.

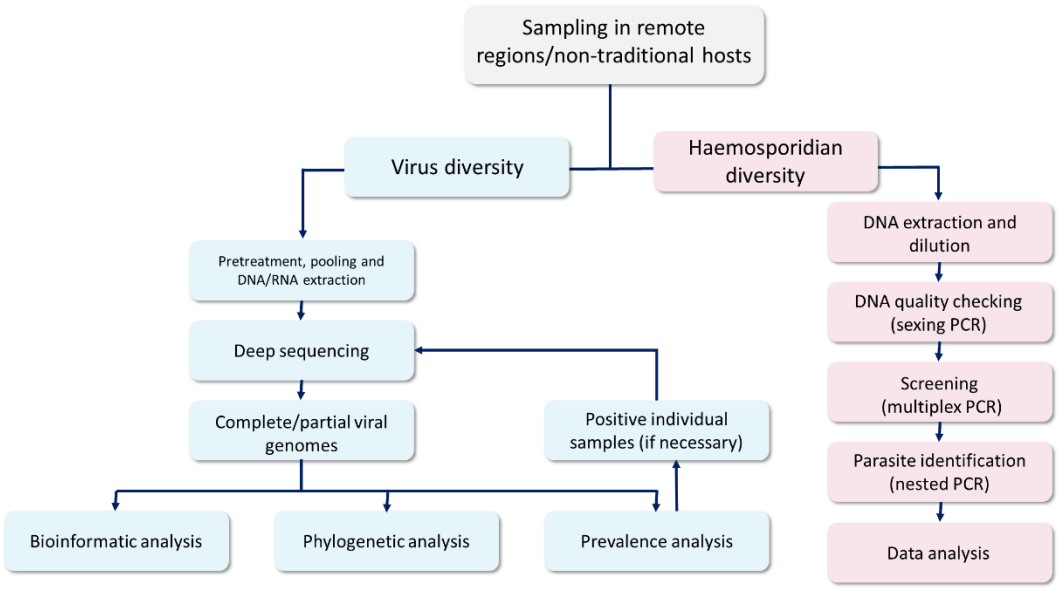


Figure 4. Protocols followed in this PhD thesis for the discovery of viruses and haemosporidians carried by wild birds from Nouragues.

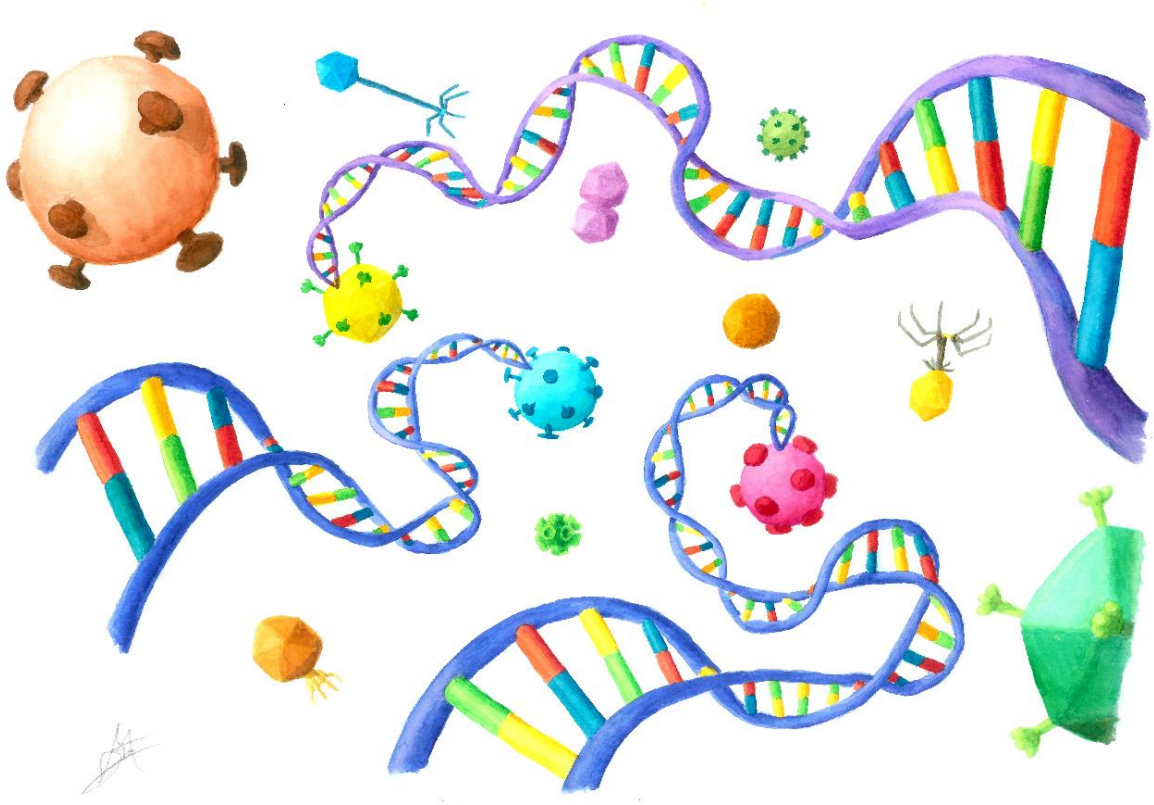
Species and lineages nomenclature

We followed the avian taxonomy proposed in the IOC World Bird List 10.2 (https://www.worldbirdnames.org/) (Gill F. et al. 2020) for bird species, the International Comitee on Taxonomy of Virus recommendations for the name of the novel viruses (https://talk.ictvonline.org/) and MalAvi criteria for the name of new *Plasmodium* lineages (http://130.235.244.92/Malavi/) (Bensch et al. 2009).

GENERAL RESULTS

With this thesis, we have contributed to better understand not only the diversity of viruses and blood parasites carried by birds, but also the difficulties and challenges associated with their discovery. We have demonstrated that if we look in underexplored carrier hosts and regions we find a new diversity of pathogens and relationships with their hosts. This results change the current paradigm about the distribution of pathogens among their hosts both in the present time (passerines can be considered as important hosts for some viruses) and in evolutionary time (the phylogenetic spectrum and the diversity of genomic structures become wider for these viruses). The use of new techniques showed that parasite prevalence may be higher than we thought, although is not without its problems.

Chapter 1: Comparative metagenomics of Palearctic and Neotropical avian cloacal viromes reveal geographic bias in virus discovery



This chapter is based on the manuscript: **Daniel A. Truchado**, Alejandro Llanos-Garrido, David A. Oropesa-Olmedo, Belén Cerrada, Pablo Cea, Michaël A. J. Moens, Esperanza Gomez-Lucia, Ana Doménech, Borja Milá, Javier Pérez-Tris, Daniel Cadar and Laura Benítez (2020). Comparative metagenomics of Palearctic and Neotropical avian cloacal viromes reveal geographic bias in virus discovery. *Microorganisms*. Accepted for publication.

Comparative metagenomics of Palearctic and Neotropical avian cloacal viromes reveal geographic bias in virus discovery

Daniel A. Truchado[1,2], Alejandro Llanos-Garrido[3], David A. Oropesa-Olmedo[1], Belén Cerrada[1], Pablo Cea[1], Michaël A. J. Moens[4], Esperanza Gomez-Lucia[5], Ana Doménech[5], Borja Milá[6], Javier Pérez-Tris[2], Daniel Cadar[7] and Laura Benítez[1]

[1] Department of Genetics, Physiology and Microbiology. Faculty of Biology. Complutense University of Madrid. José Antonio Novais, 12. 28040, Madrid, Spain

[2] Department of Biodiversity, Ecology and Evolution. Faculty of Biology. Complutense University of Madrid. José Antonio Novais, 12. 28040, Madrid, Spain

[3] Department of Organismic and Evolutionary Biology. Faculty of Arts & Sciences. Harvard University. 38 Oxford St, Cambridge, MA, USA

[4] Fundación de Conservación Jocotoco, Lizardo García E9-104 y Andrés Xaura, 170143 Quito, Ecuador

[5] Department of Animal Health. Faculty of Veterinary Medicine. Complutense University of Madrid. Avda. Puerta de Hierro s/n, 28040, Madrid. Spain

[6] National Museum of Natural Sciences, Spanish National Research Council (CSIC), Madrid 28006, Spain

[7] Bernhard Nocht Institute for Tropical Medicine, WHO Collaborating Centre for Arbovirus and Haemorrhagic Fever Reference and Research, National Reference Centre for Tropical Infectious Diseases, Hamburg, Germany

ABSTRACT

Our understanding about viruses carried by wild animals is still scarce. The viral diversity of wildlife may be best described with discovery-driven approaches to the study of viral diversity that broaden research efforts towards non-canonical hosts and remote geographic regions. Birds have been key organisms in the transmission of viruses causing important diseases, and wild birds are threatened by viral spillovers associated with human activities. However, our knowledge of the avian virome may be biased towards poultry and highly pathogenic diseases. We describe and compare the fecal virome of two passerine-dominated bird assemblages sampled in a remote Neotropical rainforest in French Guiana (Nouragues Natural Reserve) and a Mediterranean forest in central Spain (La Herrería). We used metagenomic data to quantify the degree of functional and genetic novelty of viruses recovered by examining if the similarity of the contigs we obtained to reference sequences differed between both locations. In general, contigs from Nouragues were significantly less similar to viruses in databases than contigs from La Herrería using

Blastn but not for Blastx, suggesting that pristine regions harbor a yet unknown viral diversity with genetically more singular viruses than more studied areas. Also, we describe putative novel viruses of the families *Picornaviridae, Adenoviridae, Reoviridae* and *Hepeviridae*. These results highlight the importance of wild animals and remote regions as sources of novel viruses that substantially broaden the current knowledge of the global diversity of viruses.

INTRODUCTION

The current knowledge about the virome of wild animals is still incipient. During the last 30 years, zoonotic viruses with an origin in wildlife have been the main cause of disease outbreaks, sometimes pandemic (Carroll et al. 2018). This rapid, global spread of new viruses has revealed our vulnerability to emerging diseases that have caused a great number of negative effects both in human health and the economy (El Zowalaty and Järhult 2020). In order to prevent future outbreaks, the search for novel viruses with a zoonotic potential in wildlife has become one of the main objectives of the One-Health initiative (Lebov et al. 2017; Carroll et al. 2018). Furthermore, understanding the virome of wild animals is not only important to detect potential novel zoonotic viruses before their emergence. In the last years, several studies analyzing viral diversity have challenged the view of viruses as only pathogens and have demonstrated that they constitute symbiotic microorganisms in the majority of the cases, even producing positive effects in their hosts (Roossinck and Bazán 2017). Therefore, research on viral diversity of wildlife is necessary to have a complete view of the global virome while it will help understand virus-host relationships and virus ecology. However, if we want to characterize the viral diversity of wildlife accurately, discovery-driven approaches are the optimal way of doing so. Describing novel viral strains while studying traditional hosts is no longer sufficient and specific designs that allow a greater efficiency of virus discovery are needed. Also, expanding the scope of virus discovery to non-canonical hosts and remote regions is paramount to significantly increase the current knowledge of viral diversity.

With around 10,000 described species (which may double according to phylogenetic diversity (Barrowclough et al. 2016)), the Class Aves is the most diverse tetrapod clade, inhabiting every continent across the globe. This ubiquitous presence increases their

exposure to diverse microorganisms (Benskin et al. 2009) and, together with their diverse ecologies, make birds good candidates for microorganism circulation in different ecosystems. For example, migratory birds connect ecosystems that are separated by hundreds of kilometers, carrying parasites from their breeding areas to their wintering sites and vice versa (Dhama et al. 2008; Viana et al. 2016). Moreover, they may spread large quantities of a great diversity of viruses during a long period of time (Hulse-Post et al. 2005) without any clinical signs (Hulse-Post et al. 2005; Wille et al. 2018) and their tendency to roost and feed in heterospecific groups favor the occurrence of inter-species pathogen transmission and the emergence of novel viruses (Chan et al. 2015). However, little is known about the virome of wild birds compared to their microbiome (François and Pybus 2020).

Despite the importance of increasing the knowledge of avian viruses, discovery-driven approaches to characterizing their global diversity have been rare. The majority of studies of wild bird viruses have focused on the surveillance of specific viruses that produce zoonotic infections or on those that cause massive mortalities such as highly pathogenic avian influenza virus, Newcastle disease virus, West Nile virus or Usutu virus (Li et al. 2017; Brown and Bevins 2017; Wille et al. 2018; Michel et al. 2018). Only a few studies analyze the complete virome of wild bird populations, being mainly aimed at waterfowl (Fawaz et al. 2016; Wille et al. 2018; Zhao et al. 2018; Ramírez-Martínez et al. 2018). At the same time, the virome of wild passerine populations has never been analyzed, even though they constitute approximately 60% of avian diversity (Cracraft and Barker 2009). The only report of a wild passerine's virome to date comes from the cloacal sample of one individual of the species *Sicalis flaveola* (Duarte et al. 2019). Thus, studying the virome of understudied wild bird populations will provide us with novel information about animal viruses that may be useful in the future by better preparing us for possible viral outbreaks or spillovers.

Apart from viruses that may cause disease, analyzing the virome of wild birds living in remote regions is especially interesting as it could provide useful knowledge about new virus-host relationships or the ecology of viruses circulating in ecosystems rarely disturbed by humans. The Guianan shield is one of such remote areas. Located in the Neotropics, this region is one of the main hotspots of avian diversity in the world, with more than 700 documented species (Thiollay 2002). The Guianan shield, and particularly French Guiana, is sparsely inhabited by humans and, therefore, anthropic impact is scarce.

On the other hand, human impact in French Guiana, albeit rather low, can be sufficient to introduce novel viruses that might put its unique avifauna at risk. Anthropic impact on world avian population has involved a dramatic increase in the number of endangered avian species in the last years (Bodewes 2018) and spillovers of viral strains coming from poultry vaccines have already been reported in wild bird populations (Garcia et al. 2013; Rohaim et al. 2017). By better characterizing the virome of wild birds in remote regions, not only will we expand our knowledge about the global viral diversity, but we will also be able to recognize the potential threats that endangered species might be facing and observe the influence of those viruses on the dynamics, structure and functioning of the ecosystems.

Next generation sequencing (NGS) has been the principal tool for the discovery of novel virus sequences from avian samples, being wild birds the group where they have been mainly described in the last years (Delwart 2012; Bodewes 2018). However, to our knowledge, no metagenomics or metatranscriptomics study has been performed in populations of wild birds from remote regions and, as mentioned before, the virome of some important avian groups remains virtually unknown even in more studied areas. Therefore, NGS may increase the rate of virus discovery compared to other approaches.

In this study, we used a NGS approach to analyze the cloacal virome of passerine-dominated wild bird communities from two different habitats: a remote and primary rainforest in French Guiana with limited research about avian viruses and a Mediterranean forest in Spain, a region where research about birds and their parasites is more abundant. Our general objective was to show how deep sequencing analyses of samples coming from remote areas and understudied wildlife species can efficiently increase the knowledge of virus diversity by contributing with relevant information about novel viral sequences and viral-host relationships. To this end, we analyzed whether sampling passerine birds (a non-canonical bird taxon in studies of avian viruses) in a remote area (Neotropical rainforest) contributed genetically or functionally more singular viruses than sampling passerines in a more researched area (temperate European forest). We also identified new avian and possibly non-avian viruses carried by wild birds in these two locations. All this information will provide us with important knowledge on virus diversity and virus ecology in pristine areas.

MATERIALS AND METHODS

Sample collection

A random sampling of understory bird species was carried out in two different sites: the Nouragues Natural Reserve (French Guiana) and La Herrería forest (Spain). The Nouragues Natural Reserve is located in a tropical rainforest in northern South America (4º05'N, 52º40'W), where average temperature is near 26ºC throughout the year and relative humidity is usually high. The climate is very wet in general (annual precipitation > 3,000 mm) although there is a dry season with substantially less rainfall between August and November (Grimaldi and Riera 2001). Birds were mist-netted in Pararé and Inselberg camps in January 2016 (rainy season) and October-November 2016 (dry season). La Herrería is a broadleaved forest located at 900 m.a.s.l. in the center of the Iberian Peninsula (40°34'N, 4°09'W) with a continental Mediterranean climate. The average annual precipitation is around 800 mm and there is a dry season between June and September. Although the annual mean temperature is 13ºC, the average monthly temperatures range from 5 ºC in January to 23 ºC in July (Font 2000). Sampling in La Herrería forest was carried out during the bird breeding season (April-July) of 2018.

In both locations, the same standardized protocol was followed to prevent differences due to distinct sampling methods. Cloacal samples were collected using sterile swabs (Nerbe Plus), which were preserved in 800 µl of universal viral transport medium (VTM) (Becton Dickinson) and kept frozen until molecular analyses. A total of 406 cloacal samples from 72 bird species were collected in the Nouragues Natural Reserve, and 92 cloacal samples from 20 bird species were obtained in La Herrería. Birds were marked with metal rings to avoid resampling, and released unharmed at the site of capture. The different number of species sampled in each site is representative of the different species diversity of Neotropical rainforest and Mediterranean broadleaved forests. All methods were carried out in accordance with European Union and national French and Spanish regulations. Capture, sampling and transport of samples were authorized by the Service of Natural Environments, Biodiversity, Sites and Landscapes, Regional Directorate for the Environment, Planning and Housing at French Guiana (license 030418) and the General Directorate for the Environment of Madrid (license 10/209664.9/18). The experimental protocols were approved by the Committee on Animal Testing of Complutense University (CEA-UCM, authorisation number 44-2016).

Sample selection

For each locality (Nouragues and La Herrería), 50 cloacal samples with abundant fecal matter (to make sure that those samples were properly collected) were selected, including as many different avian species as possible to maximize the number of species analyzed. We grouped them to create five pools of 10 samples in each of the two localities. When more than one individual of the same species were present, they were grouped in the same pool. In total, 32 and 18 different species were selected in Nouragues and La Herrería, respectively. Species of the order Passeriformes accounted for 92% of samples in Nouragues and 94% in La Herrería, other species were small-sized birds sharing the forest undergrowth with passerines (Tables S1.1 and S1.2). Regarding their foraging niche, the vast majority (47 birds) of the 50 individuals sampled in Nouragues Natural Reserve belonged to species that regularly feed on invertebrates (invertivorous or omnivorous). Only two individuals of the species *Pseudopipra pipra* (frugivore) and one of the species *Micrastur ruficollis* ssp. *concentricus* (carnivore) belonged to species that do not feed on invertebrates. The totality of the birds analyzed in La Herrería belonged to species that feed on invertebrates (mostly arthropods), especially during the breeding season, when the samples were taken.

Sample processing and next generation sequencing

Individual cloacal samples were vortexed, and swabs were squeezed to release epithelial cells before being discarded. The VTM was centrifuged at 13,000 rpm for 1 min to pellet out epithelial cells. Pellets were resuspended in 250 µl of PBS and subjected to 2 freeze-thaw cycles at -80 ºC to maximize the release of viral particles and filtered through 0.45 µm pore-sized column filters at 8,000 rpm for 5 min. An aliquot (50 µl) of the filtrate of each sample was combined with nine others to make five pools. Each pool was treated with a mixture of nucleases (Turbo DNase, Ambion, Carlsbad, CA, USA; Baseline-ZERO, Epicenter, Madison, WI, USA; Benzonase, Novagen, San Diego, CA, USA; RNAse One, Promega, Fitchburg, WI, USA) to digest unprotected nucleic acids including host DNA/RNA. Lastly, viral RNA/DNA was extracted with the MagMAX Viral RNA Isolation Kit (Thermo Fisher) according to the manufacturer's instructions. The extracted viral nucleic acids were subjected to library preparation, after random RT-PCR amplification, by using QIAseq FX DNA Library Kit (Qiagen, Germany). Normalized samples were pooled and sequenced using 600-cycle MiSeq Reagent Kits v3 (Illumina, San Diego, CA) on a MiSeq platform. The generated raw reads were first

qualitatively checked, trimmed and filtered to remove polyclonal and low-quality reads (< 55 bases long) using CLC workbench (Qiagen). The remaining filtered raw reads were de-novo assembled separately using Trinity v2.6.642 (Grabherr et al. 2011) and CLC workbench and compared with a non-redundant and viral proteome database using BLASTx with an E-value cut-off 0.001. The virus-like contigs and singlets were further compared to all protein sequences in non-redundant protein databases with a default E-value cutoff of 0.001. The viral metagenomics output has been visualized and analyzed in MEGAN (Huson et al. 2016).

Detection of individuals positive for the viruses

The birds carrying the viruses of interest were detected by PCR or RT-PCR using a specific set of primers we designed for each virus. First we carried out the PCR/RT-PCR using DNA/RNA extractions of the pools as a template and, once we knew the positive pool, we repeated the PCR/RT-PCR with the individual extractions. We visualized the PCR/RT-PCR product in a 2% agarose gel stained with GelRed® 100×.

Genomic analysis of the novel viruses

Genome sequence analysis and genomic organization were performed using Geneious v11 (Biomatters, New Zealand), EditSeq and SeqMan tools of the DNASTAR 5.0 software package (DNASTAR, Madison, WI), and ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/). Sequences of the putative proteins of the novel viruses were analyzed using InterPro (http://www.ebi.ac.uk/inter pro/) and Motif Scan (https://myhits.sib.swiss/cgi-bin/motif_scan) software to find conserved motifs. Similarity and possible recombination events along the amino acidic sequence of ORF1 of the different members of the *Hepeviridae* family were examined using RAT (https://omictools.com/rat-tool) and the region beyond the polymerase of unknown function was analyzed using Phyre2 in search of similarity to known protein motifs (Kelley et al. 2015).

Phylogenetic and taxonomic analysis

Multiple alignment of amino acid sequences were carried out in MUSCLE (McWilliam et al. 2013) (https://www.ebi.ac.uk/Tools/msa/muscle/). Distance matrices and maximum likelihood phylogenetic trees were inferred using MegaX (Kumar et al. 2018).

Statistical analysis

Contigs obtained with MEGAN in Nouragues and La Herrería were aligned to reference sequences in Genbank using Blastn and Blastx (https://blast.ncbi.nlm.nih.gov/Blast.cgi). We used Blastn and Blastx to hint into the genetic novelty (as scored by divergence in nucleotide sequence) and the putatively functional novelty (divergence in the encoded protein sequence) of newly discovered viruses, respectively. Before the statistical analysis, we removed all hits showing similarity to previously published viruses from our study (GyV11 and PasAst-V-1-4). We used bit scores and identity values as two variables indicative of how similar the contig is to reference sequences in databases (the higher the bit score, the lower the probability of an alignment by chance; and the higher the identity value, the more similar to reference sequence). Variation in bit scores and identity percentages of these alignments between sampling locations were analyzed using generalized linear mixed models with normally distributed errors in R v 4.0.2 (R Development Core Team 2013). In these models, site was included as a fixed factor and virus family was a random factor. Alignment length was included as a fixed covariate when identity percentage was the response variable to control for different alignment lengths. All variables were z-standardized to bring them to the same scale prior to the analyses.

RESULTS

Comparative composition of Nouragues/La Herrería cloacal viromes

In Nouragues (French Guiana), we obtained a total of 1,888,004 reads, 0.53% of them showing similarity to viruses while in La Herrería (Spain), we obtained a total of 2,919,868 reads, 0.27% of them showing similarity to viral sequences (Figure 1). Regarding viral contigs, RNA viruses were the predominant group in both localities representing 90.9% in Nouragues and 83.5% in La Herrería (Figure 2). However it is remarkable the low number of viral contigs showing similarity to ssRNA (-). DNA viruses were less abundant than RNA viruses (4.8% in Nouragues and 15.7% in La Herrería) and contigs showing similarity to unclassified viruses were a minority in both locations (Figure 2). The great majority of the putative viral contigs in both locations showed similarity to animal viruses (Figure 2). In Nouragues, the families *Polycipiviridae* and *Reoviridae* represented, by far, the main families of animal viruses (70.3% and 28.1%

respectively; Figure 3). Nevertheless, sequences from the families *Astroviridae*, *Anelloviridae* and *Picornaviridae* were also present and we were able to obtain the complete genomes of four different astroviruses, a novel gyrovirus and a novel picornavirus (Table 1). Identity values were relatively high in general for contigs of putative avian viruses in Nouragues (> 60%; Table 1) but these identities were lower when analyzing complete genomes.

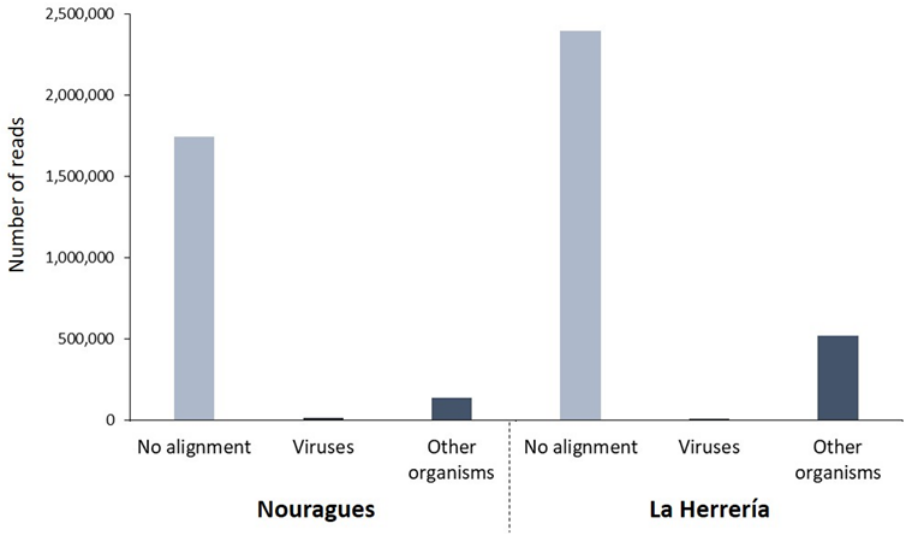


Figure 1. Comparison among the number of reads corresponding to viral sequences, the number of reads with similarity to other organisms and the reads with no alignment in Blast from the deep sequencing analyses carried out with cloacal samples of wild birds from Nouragues (French Guiana) and La Herrería (Spain)

Figure 2. Viral contig classification by genome type and by potential hosts obtained from the deep sequencing analyses of the cloacal samples from Nouragues (a) and La Herrería (b). The number of contigs of each category is shown in parenthesis. The size of the circles is proportional to the number of contigs obtained.

In La Herrería, the percentage of contigs with similarity to animal viruses was very close to Nouragues (94%). However, the second most relevant group were plant viruses (2.7%) followed closely by other eukaryotic viruses (2.6%; Figure 2). The number of different virus families of contigs showing similarity to animal viruses was slightly higher in La Herrería than in Nouragues, most of them being ssRNA (+) or dsDNA viruses (Figure 3). Also, the proportion of contigs belonging to virus families exclusively infecting vertebrates in La Herrería was higher than in Nouragues, where contigs of viruses of invertebrates prevailed. Among ssRNA (+) animal viruses, the greater part of the contigs showed similarity to viruses likely coming from invertebrate hosts. Fewer different contigs of putative avian virus were obtained in La Herrería deep sequencing and no complete genome could be assembled (Table 2).

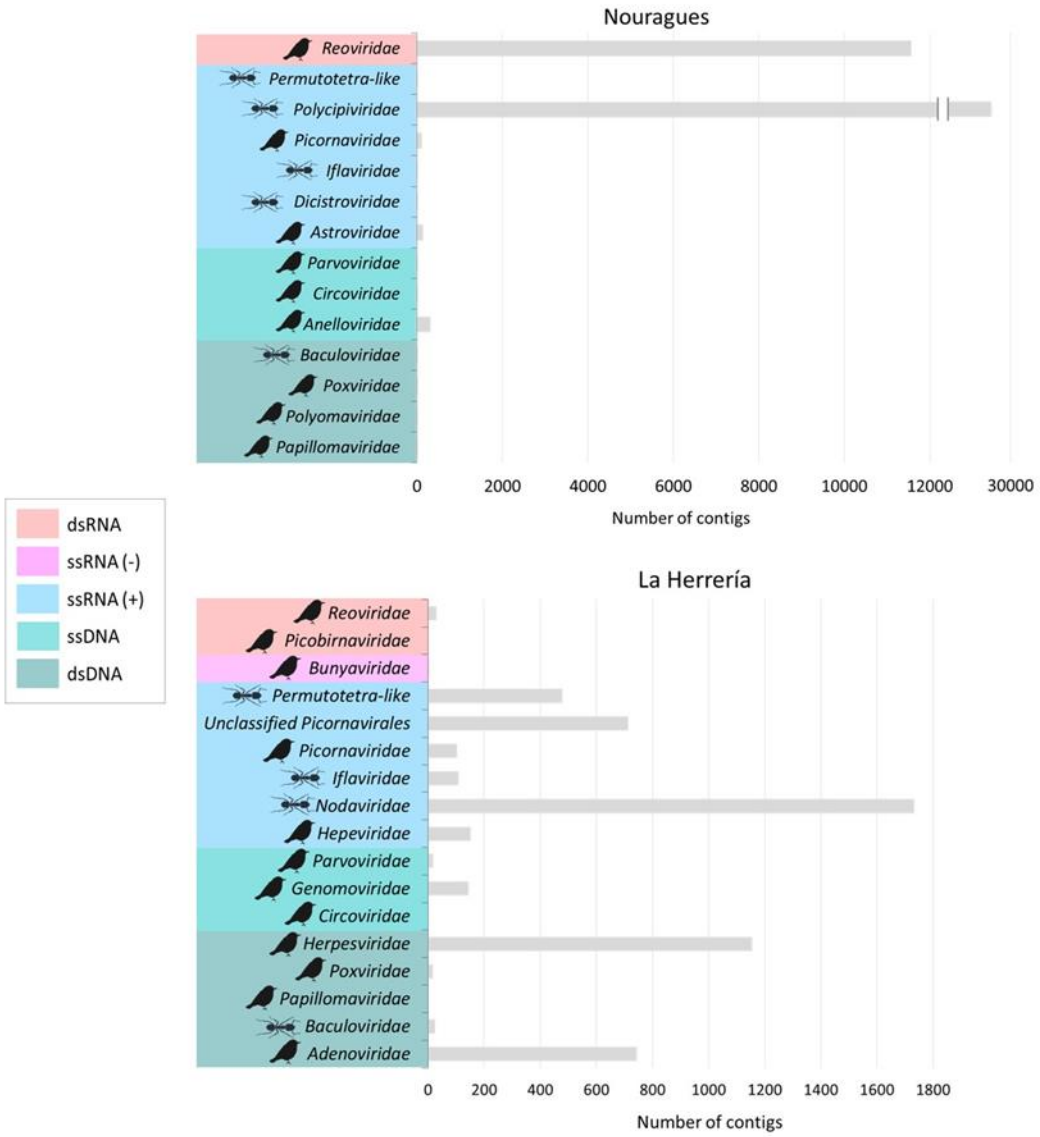


Figure 3. Contigs showing similarity to viruses of families infecting animals sequenced from cloacal samples of wild birds from Nouragues (French Guiana) and La Herrería forest (Spain). The colors indicate the type of genome of each family. The ant silhouette highlights those viral families infecting invertebrates. The bird silhouette highlights viral families with reported avian viruses.

Table 1. Contigs and complete genomes obtained by deep sequencing from cloacal samples of wild birds from Nouragues showing similarity to putative avian viruses. Closest homologs, identity and coverage values for partial genomes were obtained using Blastx.

Group	Family	Length (b)	Virus name (or closest homolog)	Accession	Genome region	Identity (%)	Coverage (%)
dsDNA	*Papillomaviridae*	407	Francolinus leucoscepus papillomavirus 1	YP_003104804	L1	79	99
	Poxviridae	176	Fowlpox virus	AAZ14082	39 kDa core protein	100	63
ssDNA	*Circoviridae*	1,955	Canary circovirus	NP_573442	Replicase	67	24
	Parvoviridae	346	Goose parvovirus	ABD76400	Nucleocapsid protein	61	76
	Anelloviridae	2,138	GyV11	MH638372	Complete genome	-	-
ssRNA (+)	*Astroviridae*	6,745	PasAstV-1	MK096773	Complete genome	-	-
		6,864	PasAstV-2	MK096774	Complete genome	-	-
		6,628	PasAstV-3	MK096775	Complete genome	-	-
		6,870	PasAstV-4	MK096776	Complete genome	-	-
	Picornaviridae	7,645	FGPV		Complete genome	-	-

When comparing the total number of viral reads in both locations to reference sequences in databases, we observed that there were no reads with long alignment lengths and low similarity values in La Herrería using Blastn or Blastx, as it happened in Nouragues (Figure 4). Bit scores and identity percentages were significantly lower in Nouragues than in La Herrería using Blastn (Table 3; Figure 5). However, bit scores and identity values were significantly lower in La Herrería using Blastx. When we compared the values obtained in Blastn and Blastx of these two variables for a given contig, we observed that unclassified viruses and a great part of RNA viruses in Nouragues showed the lowest values of bit scores both in Blastn and Blastx (Figure 6). A similar result was obtained for La Herrería. Results for identity values are less clear. However, the majority of RNA viruses showed lower values in Blastn and Blastx in comparison to DNA viruses. Unclassified viruses of Nouragues showed high identity values in Blastn but low values in Blastx (Figure 6).

Table 2. Contigs obtained by deep sequencing from cloacal samples of wild birds from La Herrería showing similarity to putative avian viruses. Closest homologs, identity and coverage values were obtained using Blastx.

Group	Family	Contig length (b)	Closest homolog(s)	Accession	Genome region	Gap (%)	Identity (%)	Coverage (%)
ds DNA	*Adenoviridae*	2,851	*Great tit siadenovirus A*	YP_009665993	DNA polymerase	4	67	99
		1,983	*Great tit siadenovirus A*	YP_009666001	Hexon	9	73	96
		930	*Raptor siadenovirus A*	YP_004414811	100 kDa Protein	5	62	100
		1,143	*Great tit siadenovirus A*	YP_009665997	pIII (base penton)	4	74	100
		2,430	Cacatua sanguinea adenovirus	QHB43551	100 kDa Protein	71	59	100
		2,154	*Penguin siadenovirus A*	YP_009252219	100 kDa Protein	30	62	100
		507	*Turkey siadenovirus A*	NP_047394	Endoprotease	-	62	100
		348	*Turkey siadenovirus A*	NP_047395	DNA binding protein	-	64	100
		486	*Penguin siadenovirus A*	YP_009252217	Protease	23	55	100
		963	Antarctic penguin adenovirus	ALB78159	IVa2	33	78	100
ssDNA	*Poxviridae*	903	Shearwaterpox virus	ARE67299	Immunoglobulin domain	-	45	27
	Parvoviridae	472	Dependoparvovirus	QHY93491	Non-structural protein	-	86	41
			Parus major densovirus	YP_009310052	ORF5	-	50	41
ds RNA	*Reoviridae*	446	Rotavirus B	ANN82201	RdRp	-	51	71
		283	Rotavirus G	AXF38053	NSP2	-	47	63
		236	Rotavirus J	APQ41756	VP4	-	63	54

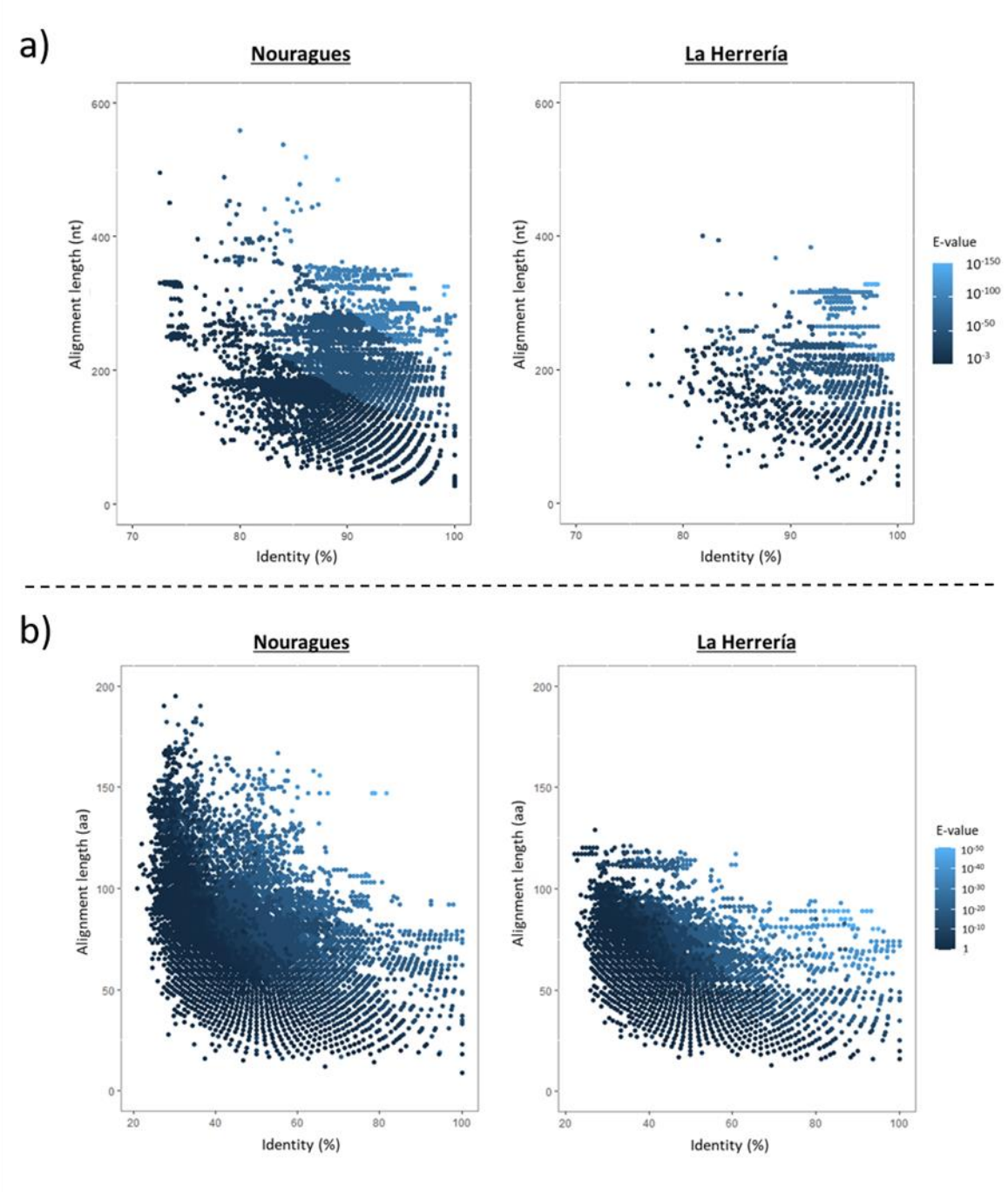


Figure 4. Alignment length (nt, nucleotides; aa, amino acids) vs. identity percentage for the contigs obtained in Nouragues and La Herrería cloacal deep sequencing analysis when compared to reference sequences in GenBank using Blastn (a) and Blastx (b). Each dot is colored according to the E-value of the alignment in order to increase the descriptive value of the graph.

Table 3. Summary of the general linear mixed models analyzing bit scores and identities of the alignments between contigs sequenced in Nouragues and La Herrería and reference sequences in Genbank database using Blastn and Blastx.

Blastn						
	Bit scores			**Identities**		
Predictors	*Incidence Rate Ratios*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	121.19	98.73 – 148.76	**<0.001**	86.09	84.19 – 88.00	**<0.001**
site [Herreria]	1.49	1.41 – 1.58	**<0.001**	4.37	3.75 – 4.99	**<0.001**
Random Effects						
σ^2	0.13			8.60		
τ_{00}	0.30 $_{\text{family}}$			17.29 $_{\text{alignment length}}$		
				22.97 $_{\text{family}}$		
ICC	0.69			0.82		
N	28 $_{\text{family}}$			28 $_{\text{family}}$		
				367 $_{\text{alignment length}}$		
Observations	36197			36197		
Marginal R^2 / Conditional R^2	0.038 / 0.701			0.040 / 0.831		
Blastx						
	Bit scores			**Identities**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.65	0.41 – 0.90	**<0.001**	0.03	-0.24 – 0.31	0.819
site [Herreria]	-0.65	-0.67 – -0.62	**<0.001**	-0.27	-0.30 – -0.25	**<0.001**
Random Effects						
σ^2	0.76			0.48		
τ_{00}	0.58 $_{\text{family}}$			0.65 $_{\text{alignment length}}$		
				0.57 $_{\text{family}}$		
ICC	0.43			0.72		
N	38 $_{\text{family}}$			38 $_{\text{family}}$		
				170 $_{\text{alignment length}}$		
Observations	112176			112176		
Marginal R^2 / Conditional R^2	0.072 / 0.475			0.011 / 0.720		

Figure 5. Comparison between values obtained in Blastn and Blastx for bit scores and identity percentages for a given contig in Nouragues and in La Herrería.

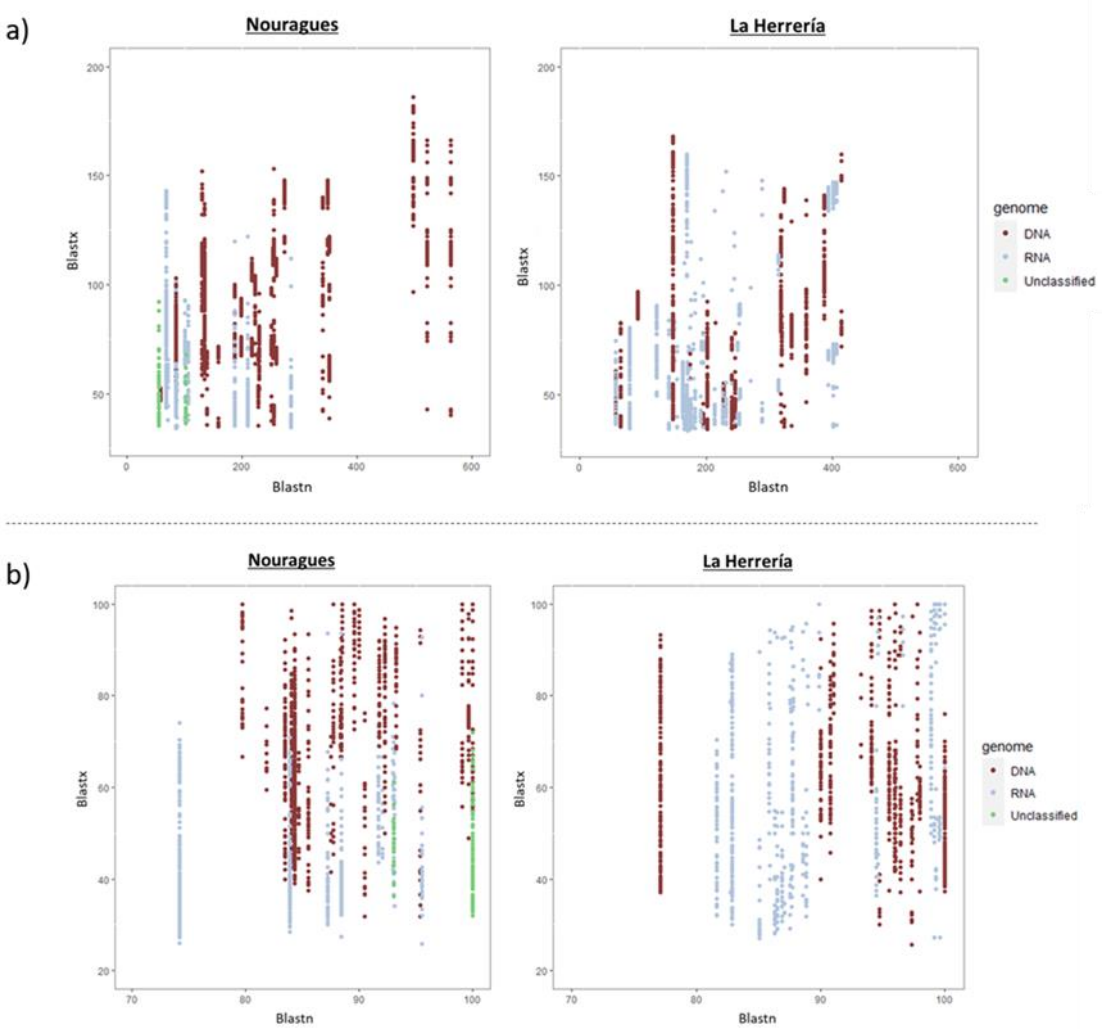


Figure 6. Comparison between values of bit scores (a) and identities (b) of contigs sequenced in Nouragues and La Herrería using Blastn and Blastx considering the type of genome of the novel viruses. The type of genome (DNA, RNA or unclassified) was obtained from the viral group each contig was assigned to by the software MEGAN.

Novel viruses found in Nouragues/La Herrería

Two novel viruses drew special attention among all possible new avian pathogens we obtained. In Nouragues, a novel picornavirus provisionally named French Guiana Picornavirus (FGPV; Genbank accession number MT792642) was detected in the cloacal sample of a Rufous-throated Antbird (*Gymnopithys rufigula*). Phylogenetic analysis place FGPV in a divergent branch sister to the genera *Avihepatovirus* and *Avisivirus* (Figure A1). However, it could not be unambiguously classified as part of either of these genera according to ICTV criteria. On the other hand, we identified a partial genome of a novel siadenovirus in La Herrería, provisionally named La Herrería Adenovirus (LHAdV), in two pools containing only passerine species. LHAdV is phylogenetically related to *Raptor adenovirus A* and *Great tit siadenovirus A* (Figure A2) and could be considered a novel species according to ICTV criteria for the genus *Siadenovirus*. Added to these two viruses are the four novel astroviruses and a novel gyrovirus detected in these deep sequencing analyses that were published elsewhere (Fernández-Correa et al. 2019; Truchado et al. 2019).

At the same time, three other viruses stood out because they were highly divergent to their closest relatives, although they cannot be clearly classified as avian pathogens. A novel reovirus, provisionally named French Guiana Reovirus (FGRV; Genbank accession numbers MT792643-48), was detected in the cloacal sample of a Wedge-billed Woodcreeper (*Glyphorynchus spirurus*) in Nouragues. The phylogenetic tree using the amino acidic sequences of the putative RNA-dependent RNA polymerase showed that this novel reovirusis grouped with Cimodo virus in a divergent clade within the subfamily *Spinareovirinae* (Figure A3). Finally, two hepe-like viruses were detected in the cloaca of two invertivore birds: a Plain Xenops (*Xenops minutus*) in Nouragues (provisionally named French Guiana Hepevirus; FGHEV) and a European Robin (*Erithacus rubecula*) in La Herrería (provisionally named La Herrería Hepevirus; LHHEV). Although only FGHEV is fully sequenced, both show unusual arrangements in their genome architectures. Phylogenetic analyses placed LHHEV as a clade within hepe-like viruses, but the phylogenetic location of FGHEV was less clear (Figure S8). The complete genome of FGHEV was deposited in Genbank under accession number MT792641. A more detailed characterization of the viruses can be found in the Appendix A.

DISCUSSION

In a context where viral discovery and surveillance in wildlife have become one of the main goals to prevent disease outbreaks and global pandemics (Carroll et al. 2018; Sun et al. 2020), this study adds information about novel viruses harbored by birds in wild populations and potential gaps of knowledge of the global virus diversity associated with the paucity of research in understudied species and geographic regions. Using a discovery-driven approach to uncovering virus diversity, we have examined the cloacal virome of passerine-dominated bird communities sampled in a tropical rainforest and in a Mediterranean habitat, in order to reveal the importance of remote areas and wildlife as sources of relevant new information in the field of virology. Also, we shed some light on the cloacal virome of wild passerines, addressing the issue for the first time despite the relevance and ubiquity of this group of birds. We discovered several novel and divergent viruses of the families *Adenoviridae*, *Anelloviridae*, *Astroviridae*, *Hepeviridae*, *Picornaviridae* and *Reoviridae*. The novel viruses of the families *Anelloviridae* and *Astroviridae* were described in detail elsewhere (Fernández-Correa et al. 2019; Truchado et al. 2019).

Comparative composition of Nouragues/La Herrería cloacal viromes

The presence of contigs showing little or no similarity to reference sequences highlights the need for further research in the virome of birds and other wildlife. It is especially noteworthy the high number of reads showing no similarity with sequences in the GenBank database. As this proportion of reads with unknown origin is very similar in Nouragues and La Herrería, this result could be explained mainly because wild birds, and especially wild passerines, have remained understudied in this type of analyses compared to other avian groups such as poultry or waterfowl. It is considered that up to 90% of viral reads in deep sequencing analyses can be considered as "viral dark matter" (Krishnamurthy and Wang 2017) as they do not align to any available viral sequence, especially when dealing with highly divergent viruses and short fragments. Also, research in virology has been highly biased towards human and other mammalian viruses (Huson et al. 2007; Krishnamurthy and Wang 2017), so analyzing the virome of non-canonical hosts represents a challenge as closely related sequences are scarce or absent in public databases. Thus, the majority of the contigs sequenced in our study could not be assigned to any known taxon, showing the importance of continuing studying unexplored species and regions to expand the available reference sequences in the future.

Regarding viral reads, RNA viruses were the predominant group in both sampling sites. RNA viruses have been shown to be the most abundant group in other deep sequencing analysis from avian fecal samples (Fawaz et al. 2016; Vibin et al. 2018; Wang et al. 2019b). However, not all RNA viruses were equally represented. It is remarkable the low number of reads related to ssRNA (-) viruses we obtained in both Nouragues and La Herrería. In some studies of fecal avian virome, ssRNA (-) are not very abundant but are present to some extent (Vibin et al. 2018) while in others they are completely absent or not highlighted by the authors (Fawaz et al. 2016; Lima et al. 2017; Wang et al. 2019b; Duarte et al. 2019). Avian influenza virus, the main ssRNA (-) viruses found in birds so far, has been shown to be present in fecal samples of waterfowl, turkeys and chickens, being wild aquatic birds their main reservoir (Webster et al. 1978; Swayne 2008; Wille et al. 2019). However, avian influenza virus is not present in all bird groups. Wild passerines do not seem to play a role in the transmission of avian influenza virus (Slusher et al. 2014). As passerines represent the great majority of the birds sampled in this study and we did not collect samples from aquatic birds, this could be an explanation of the low number of reads related to ssRNA (-) viruses we obtained. On the other hand, Rosseel et al. (2015) showed that different pre-treatments of samples before deep sequencing had different effects on the detection of ssRNA(-) viral reads. More specifically, a random PCR amplification before deep sequencing had a negative impact on the number of ssRNA (-) detection. As our samples were pre-treated this way, it is possible that the number of reads related to ssRNA (-) viruses was underrepresented in our study. In relation to DNA viruses, they were a minority in both locations, representing 15.7% of viral reads in La Herrería and only 4.1% in Nouragues. These results differ from other studies of fecal virome in birds, where this percentage is much higher (Fawaz et al. 2016; Ramírez-Martínez et al. 2018; Wang et al. 2019b). Only Zhao et al. (2018) obtained a similar proportion of DNA virus reads in their study with the fecal virome of Jinding ducks although they followed a different pretreatment.

Viral classification by host showed that animal viruses represented the greatest proportion in our study. In Nouragues, prokaryotic viruses were the second most abundant group, in contrast with the results in La Herrería, where plant and other eukaryotic viruses were more abundant. The majority of animal viruses we found had similarity with viruses of invertebrates, as expected given the type of sample and the diet of the birds analyzed. Sequences of insect viruses were also abundant in previous deep sequencing analyses of

insectivorous animals such as birds (Phan et al. 2013; Fawaz et al. 2016; Wang et al. 2019b) or bats (Li et al. 2010; Wu et al. 2012), also reflecting their dietary preferences. The remarkably high percentage (60,8 %) of contigs showing similarity to the *Polycipiviridae* family (arthropod viruses) in Nouragues cloacal virome could be explained because of the great number of invertivore species among the birds sampled, being an example of how the virome can be influenced by the structure and ecology of host community, as it was shown in other wild bird populations (Wille et al. 2018). However, a divergent member of this family has recently been reported in the stool of a bat, being the first time that these viruses appear in a vertebrate (Temmam et al. 2019). So, although the presence of viruses of the *Polycipiviridae* family in cloacal or rectal samples seems likely due to diet, the possibility that this newly described viral family has members infecting vertebrate hosts cannot be ruled out.

Our discovery-driven approach to documenting virus diversity showed that the endeavor of uncovering new avian viruses may not only benefit from studies of non-canonical host species, but also from sampling remote areas seldom explored in virus research. Viral contigs found in Nouragues showed, in general, significantly lower bit scores and identity percentages using Blastn. This would imply that nucleotide sequences of viruses present in the cloaca of wild birds in remote regions are less similar to known viruses than the viruses carried by wild birds living in more studied ecosystems. Furthermore, although we found higher diversity of virus families in La Herrería, the singularity of these viruses was lower compared to Nouragues, where all complete or almost complete viral genomes were very divergent to their closest relatives in the phylogenetic analyses. On the contrary, bit scores and identities were, in general, significantly lower in Nouragues than in La Herrería, suggesting that amino acidic sequences of viruses from Nouragues are more similar to those in databases. This apparently contradiction could reflect the difference between genetic singularity (nucleotide sequences) and functional singularity (amino acidic sequences). Nucleotide sequences in Nouragues are less similar to reference sequences probably due to the singularity of the sampling area and carriers, very different from the traditional ones. Moreover, the isolation of Nouragues ecosystem implies more viruses with unknown nucleotide sequences of which we detect only their functionality through their amino acidic sequences using Blastx. Also, contigs showing similarity to unclassified viruses in Nouragues showed high identity values in Blastn but, however, low bit scores which could reflect alignments by chance of short viral reads to

reference sequences. Taking all this into consideration, we can suggest that wild birds of Nouragues carry genetically more singular viruses belonging to fewer virus families than wild birds of La Herrería. However, this result does not seem to be true for functional singularity of the viruses as it is captured from protein sequence divergence, so more comparative studies in remote regions are needed considering both types of singularity in order to clarify this trend.

Novel viruses of interest found in Nouragues/La Herrería

Focusing on viruses of vertebrates, the main families we found in our study have been frequently found in other fecal viromes of birds. For example, picornaviruses are usually found in this type of samples, irrespective of whether birds are domestic or free-living or if they are healthy or not (Lina et al. 1973; Day et al. 2010; Phan et al. 2015; Fawaz et al. 2016; Wang et al. 2019b; Wille et al. 2019). Astrovirus sequences also have been frequently reported in previous metagenomics analysis of avian fecal samples (Day et al. 2010; Mihalov-Kovács et al. 2014; Phan et al. 2015; Wang et al. 2019b). However, the four novel astrovirus genomes we obtained from wild birds from Nouragues constitute a putative new species, providing important new information in relation to the family *Astroviridae* (Fernández-Correa et al. 2019). The same happens with GyV11, the divergent novel gyrovirus we found in the same set of samples (Truchado et al. 2019). The four astroviruses detected in Nouragues and GyV11 are examples of divergent viruses related to important avian pathogens involved in intestinal disorders and so far unknown circulating in a remote region. Nevertheless, they were not the only putative avian viruses whose complete genome was obtained in this population of Neotropical birds. FGPV was present in the cloacal sample of a Rufous-throated Antbird, being the first time that a picornavirus is detected in a bird of the family Thamnophilidae. The family *Picornaviridae* is the most diverse among ssRNA (+), with more than 75 accepted species infecting mainly mammals and birds (Zell et al. 2017). However, the discovery of a novel, divergent picornavirus likely corresponding to a new species infecting birds shows that there is an unknown diversity of this group of viruses yet to be discovered. In fact, novel and divergent picornaviruses have been recently detected in hosts rarely sampled before (de Souza et al. 2019; Wang et al. 2019a; Wille et al. 2019) and, prior to these studies, only five out of 18 species of avian picornaviruses had been described in wild birds (Woo et al. 2010; Boros et al. 2013; Phan et al. 2013; Boros et al. 2017; Pankovics et al. 2018). It is interesting that the Rufous-throated Antbird positive to FGPV

was also positive for astrovirus, showing a possible coinfection. Unfortunately, as picornavirus infections are usually asymptomatic, it is difficult to determine the effect of FGPV, as in the case of other novel members of the family.

Another novel putative avian virus we detected, in this case in La Herrería forest, is LHAdV. Adenoviruses are pathogens of different species of vertebrates, mainly causing infections of the respiratory and digestive systems. They are transmitted through respiratory droplets and via fecal-oral route, so adenoviruses have been commonly found in other studies of fecal viromes in birds (Phan et al. 2013; Lima et al. 2017; Zhao et al. 2018). In this group of animals, adenoviruses are widespread among domestic and wild birds, where they usually produce opportunistic infections when the host's health is compromised (Fitzgerald et al. 2020). One exception to this is *Turkey siadenovirus A*, currently classified in the genus *Siadenovirus*, which causes severe infections and mortality in poultry (Domermuth et al. 1979; Fitzgerald et al. 2020). The novel adenovirus we found as part of the fecal virome of passerines in La Herrería forest, LHAdV, is proposed to be a novel species of the genus *Siadenovirus*. The closest relatives of LHAdV are *Great tit siadenovirus A* and *Raptor siadenovirus A*, two of the siadenoviruses detected in birds. *Great tit siadenovirus A* was also found in a passerine bird showing acute enteritis, being likely (but not proven to be) the cause of this clinical sign (Kovács et al. 2010). On the other hand, *Raptor siadenovirus A* was detected in three different dead individuals of hawks and eagle owls showing diverse pathologies such as hepatomegaly, splenomegaly or proventricular and ventricular dilation (Zsivanovits et al. 2006). Thus, LHAdV could be an important pathogen for wild passerines circulating in La Herrería forest that may cause similar clinical conditions as its closer relatives. However, given the nature of our sampling method, we cannot prove that the individuals carrying LHAdV showed any clinical sign related to an adenovirus infection.

Regarding FGRV and the two novel hepe-like viruses we have detected in our samples, we cannot clearly classify them as avian pathogens or as viruses of invertebrates that were in the diet of the analyzed birds. Reoviruses are dsRNA viruses infecting a wide variety of hosts and causing gastroenteritis and respiratory diseases in vertebrates. The closest reovirus to FGRV is Cimodo virus, detected in Africa and which likely infects mosquitoes (Hermanns et al. 2014). Both reoviruses appear to forrm a new genus within the subfamily *Spinareovirinae*. FGRV was present in the cloacal sample of a Wedge-billed Woodcreeper (*Glyphorynchus spirurus*), an insectivorous bird of Nouragues, so it is

possible that this novel reovirus was infecting insects that this bird had fed on. However, we cannot discard that FGRV and Cimodo virus might be arboviruses transmitted by mosquitoes (or other arthropods) to birds, as neither of them have been tested to infect bird cell cultures (Hermanns et al. 2014).

A similar situation occurs in the case of the two novel hepe-like viruses we detected: FGHEV and LHHEV. Hepeviruses are important zoonotic viruses causing hepatitis E and splenomegaly with high mortality rates among vertebrates. Hepe-like viruses are a group phylogenetically related to hepeviruses recently described whose effect on their host is still unknown (Wu et al. 2018). Our results suggest that FGHEV and LHHEV would belong to this latter group for several reasons. Firstly, their ORF arrangement and genome length are more similar to those of the hepe-like viruses than to those of hepeviruses (Purdy et al. 2017). Phylogenetic analyses supports this hypothesis in the case of LHHEV, placing it clearly within the hepe-like group, whose members have been mainly detected in invertebrates (Shi et al. 2016; Wu et al. 2018; Dong et al. 2020) or in fecal samples of animals feeding on them (Williams et al. 2018; Reuter et al. 2018, 2020). Nonetheless, the phylogenetic position of FGHEV is more ambiguous, as it could be related to hepe-like viruses of invertebrates or to hepeviruses of vertebrates depending on the fragment selected to infer the tree. Given that the species of birds carrying both novel hepe-like viruses feed on invertebrates, both FGHEV and LHHEV could be actually viruses of invertebrates that were part of the diet of the birds. However, the fact that we were able to sequence the whole genome of FGHEV and almost the complete genome of LHEEV, argues in favor that both viruses could have maintained their integrity until the end of the avian digestive tract. In that case, birds would act as dispersers of these viruses and their putative invertebrate hosts could become infected if they got in close contact with bird droppings. This transmission route through the feces of predators has been shown to happen for a virus infecting the gypsy moth (Reilly and Hajek 2012) and could be used also by the aforementioned FGRV and *Polycipiviridae* virus in Nouragues, as both groups were the most abundant in those cloacal samples. Unfortunately, little is known yet about the pathogenicity and ecology of the hepe-like group. Therefore, we cannot rule out that FGHEV and LHHEV are exclusively novel avian pathogens or that they could infect and circulate among both invertebrate and vertebrate species, especially FGHEV. In any case, these unclear results for FGHEV are the evidence that it belongs to a yet unknown

diversity of viruses circulating in this remote area so further research is needed to clarify its ecology, epidemiology and its closer relatives in the phylogeny.

CONCLUSIONS

In general, our research reveals how extending the focus to non-canonical hosts and regions is crucial for viral discovery. Most of the potential avian viruses we obtained using a discovery-driven approach were different enough to be considered novel species or even genera. This seems to be mainly the effect of the limited existing research on the virome of wild birds and, especially, of wild passerines. Therefore, carrying out deep sequencing analyses in bird species other than poultry and waterfowl contributes substantially to gain better insight into avian virology. In fact, none of the novel viruses described in birds during recent years were detected in poultry (Bodewes 2018). Moreover, our results also show how remote regions harbor an unknown diversity of viruses that is yet to be described, and how preserving and studying pristine forests is highly relevant in the research of emerging infectious diseases to avoid future spillovers that affect humanity and biodiversity. Thus, a good approach to widen the knowledge about animal viruses in general would be to combine sampling in understudied animal hosts with analyzing the virome of wild animals in remote regions.

Appendix A- Description of the novel viruses of interest found in Nouragues and La Herrería

Possible viral avian pathogens identified in Nouragues and La Herrería

A partially complete fragment showing similarity to Duck hepatitis A virus (DHAV) was sequenced in the metagenomic analysis of cloacal samples from Nouragues. A Rufous-throated Antbird (*Gymnopithys rufigula*) was identified by RT-PCR as the carrier of the virus. After a second deep sequencing analysis of this individual's cloacal sample, we obtained a complete genome of 7,645 b (provisionally named French Guiana picornavirus; FGPV). The highest identity values in the amino acid sequence of the putative capsid protein (P1) of FGPV were 32.23% with *Avihepatovirus* and 31.04% with *Avisivirus*. However, the identity values for the 2C+3CD amino acid sequence were higher, reaching a 47.75% with *Avihepatovirus* and 38.30% with *Avisivirus*. A phylogenetic tree based on the amino acid sequence of the P1 region placed FGPV in a divergent branch within *Picornaviridae* family, sister to the clade grouping the genera *Avihepatovirus*, *Avisivirus*, *Aalivirus* and *Orivirus*, all of them detected only in avian species (Figure A1). Taking all these results together, the novel FGPV could not be unambiguously considered as part of the existing *Avihepatovirus* or *Avisivirus* genera according to ICTV criteria as identity for P1 is < 33%, but 2C+3D is > 36%.

Regarding La Herrería, we assembled ten non-overlapping contigs with similarity to adenoviruses from two pools containing only passerine species. As the contigs belong to different parts of the genome, we cannot discard that they come from the same virus. The phylogenetic analysis of this novel adenovirus was carried out using the longest fragment (2,851 b), corresponding to the putative DNA polymerase. This novel adenovirus (provisionally named La Herrería adenovirus; LHAdV) forms a sister clade to *Raptor adenovirus A* and *Great tit siadenovirus A* within the *Siadenovirus* genus (Figure A2). Phylogenetic distance based on DNA polymerase amino acid sequence is greater than 0.15 with the closest known adenovirus (*Great tit siadenovirus A* , p-dist = 0.36) so, according to ICTV criteria for the genus *Siadenovirus*, this novel adenovirus could be considered a putative new species.

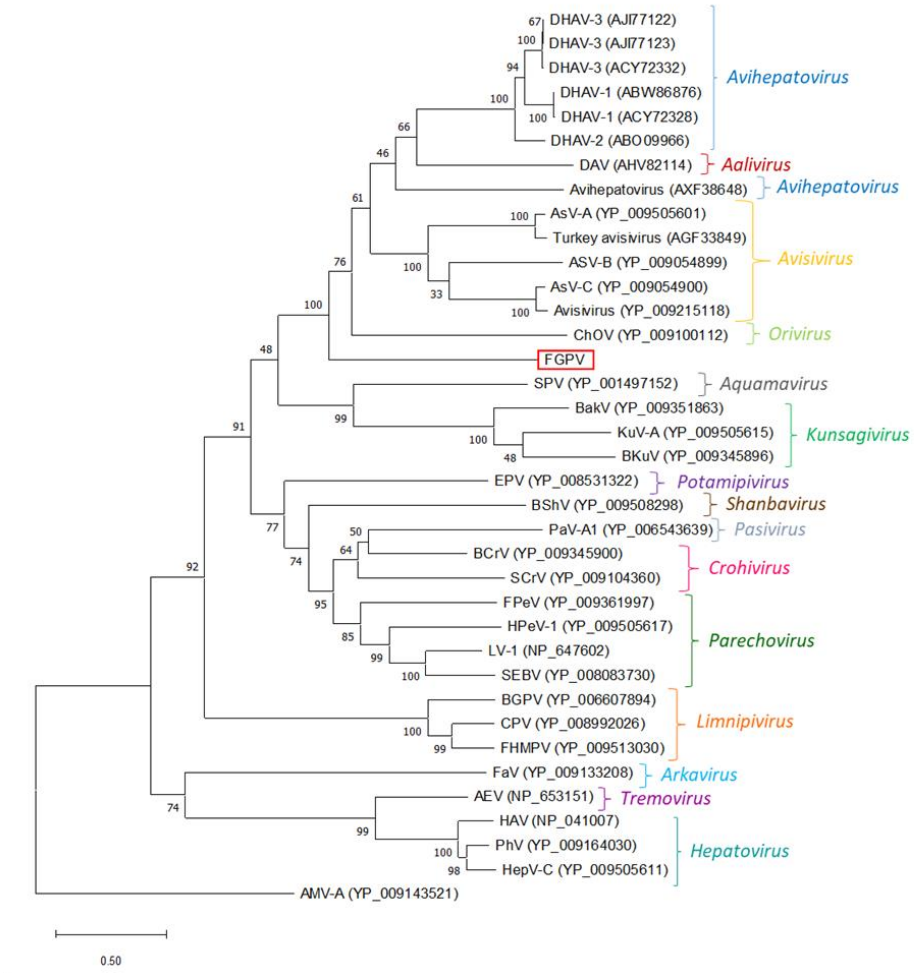


Figure A1. Maximum likelihood tree based on amino acid sequences of P1 region of the family *Picornaviridae* using LG + G + I + F method (bootstrap value = 1,000 replicates). The different genera within the family are indicated to the right of the tree. Taxon information includes species names and GenBank accession numbers. Branch lengths measure the number of substitutions per site. *Ampivirus A* (AMV A) was used as outgroup. AEV, avian encephalomyelitis virus; AsV, avisivirus; BakV, bakunsavirus; BCrV, bat crohivirus; BGPV, bluegill picornavirus; BKuV, bat kunsagivirus; BShV, bat picornavirus; ChOV, chicken orivirus; CPV, carp picornavirus; DAV, duck aalivirus; DHAV, duck hepatitis A virus; EPV, eel picornavirus; FaV, falcovirus; FGPV, French Guiana picornavirus; FHMPV, fathead minnow picornavirus; FPeV, ferret parechovirus; HAV, hepatitis A virus; HepV-C, hepatovirus C; HPeV, human parechovirus; KuV, kunsagivirus; LV, ljunganvirus; PhV, phopivirus; PaV, pasivirus; SCrV, shrew crohivirus; SEBV, Sebokele virus; SPV, seal picornavirus.

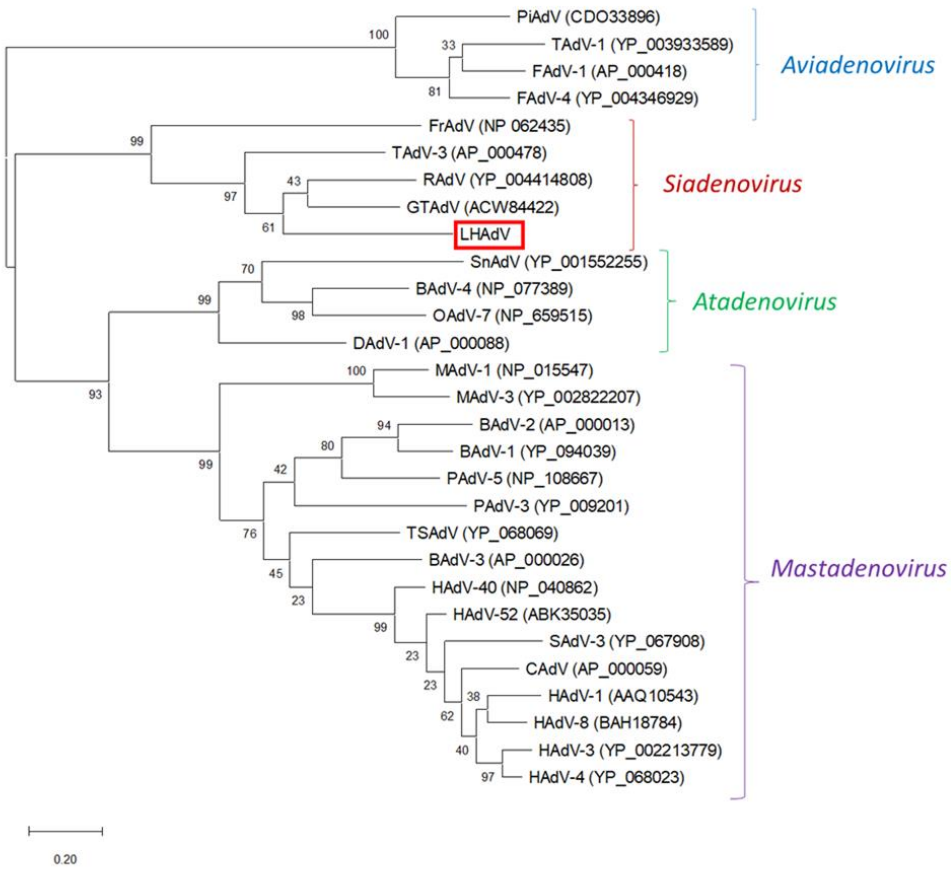


Figure A2. Maximum likelihood tree based on amino acid sequences of DNA polymerase of the family *Adenoviridae* using LG + G method (bootstrap value = 1,000 replicates). The different genera within the family are indicated to the right of the tree. Taxon information includes species names and GenBank accession numbers. Branch lengths measure the number of substitutions per site. BAdV, Bovine adenovirus; CAdV, *Canine mastadenovirus A*; DAdV, Duck adenovirus; FAdV, Fowl aviadenovirus; FrAdV, *Frog siadenovirus*; GTAdV, *Great tit siadenovirus A*; HAdV, Human adenovirus; LHAdV, La Herreria adenovirus; MAdV, Murine adenovirus; OAdV, Ovine adenovirus; PAdV, Porcine adenovirus; PiAdV, *Pigeon aviadenovirus*; RAdV, *Raptor siadenovirus A*; SAdV, Simian adenovirus; SnAdV, *Snake atadenovirus A*; TAdV, Turkey adenovirus; TSAdV, *Tree shrew mastadenovirus A*.

Highly divergent viruses identified in the cloacal viromes

In the deep sequencing analysis from both regions, we identified three complete or almost complete viral genomes that, although they showed some similarity to families of animal virus, they were highly divergent from their closest relatives: a reo-like virus in Nouragues and two hepe-like viruses, one at each location.

As for the reo-like virus, we first obtained a partially complete genome showing similarity to this group of viruses. A primer set was designed to identify the carrier of the virus, a Wedge-billed Woodcreeper (*Glyphorynchus spirurus*) whose individual cloacal sample was subjected to a second deep sequencing analysis to obtain the complete genome of the novel reovirus. We finally obtained six complete fragments, for a total of 18,567 bp (approximately 75% of the complete genome of other reoviruses). These complete fragments showed similarity to Cimodo virus (Genbank accession number KF880748), an unclassified reovirus. The phylogenetic tree using the amino acid sequences of the putative RNA-dependent RNA polymerase (RdRp) showed that this novel reovirus, provisionally named French Guiana Reovirus (FGRV), is grouped with Cimodo virus in a divergent clade within the subfamily *Spinareovirinae* (Figure A3).

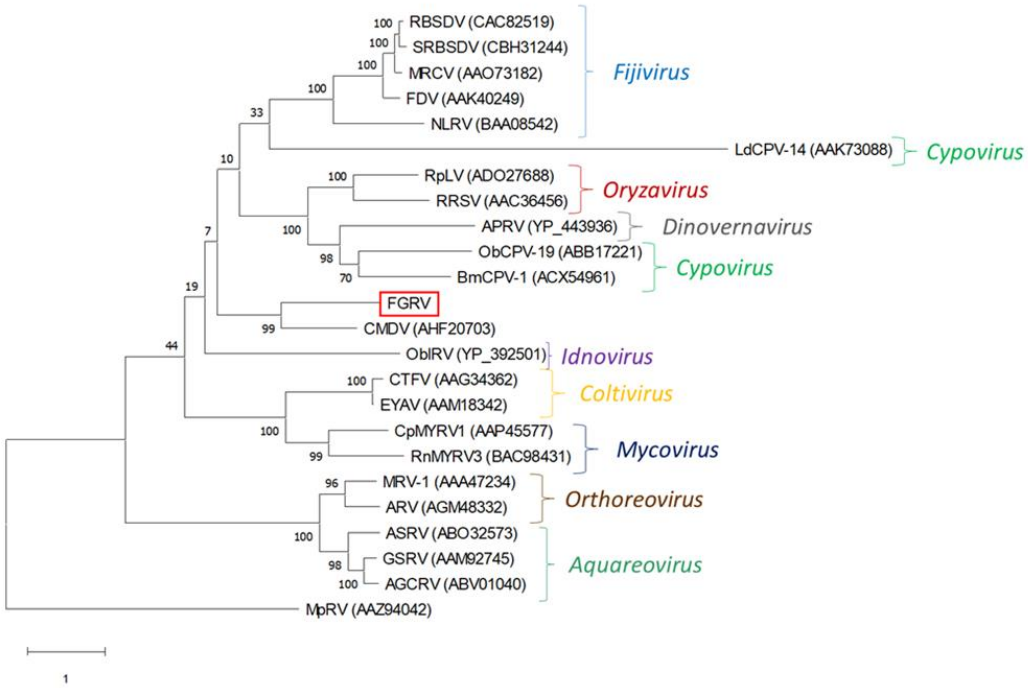


Figure A3. Maximum likelihood tree based on amino acid sequences of RNA-dependent RNA polymerase using LG + G + F method (bootstrap value = 1,000 replicates) showing the position of French Guiana reovirus (FGRV) within the subfamily *Spinareovirinae*. The different genera within the family are indicated to the right of the tree. Taxon information includes species names and GenBank accession numbers. Branch lengths measure the number of substitutions per site. Micromonas pusilla reovirus (MpRV), a representative of the subfamily *Sedoreovirinae*, was used as outgroup. AGCR, American grass carp reovirus; APRV, Aedes pseudoscutellaris reovirus; ARV, avian orthoreovirus; ASRV, Atlantic salmon reovirus; BmCPV, Bombyx mori cypovirus; CpMYRV, Cryphonectria parasitica mycoreovirus; CTFV, Colorado tick fever

virus; CMDV, Cimodo virus; EYAV, Eyach virus; FDV, Fiji disease virus; GSRV, golden shiner reovirus; LdCPV, Lymantria dispar cypovirus; MRCV, mal de Rio Cuarto virus; MRV, mammalian orthoreovirus; NLRV, Nilaparvata lugens reovirus; ObCPV, Operophtera brumata cypovirus; ObIRV, Operophtera brumata idnoreovirus; RBSDV, rice black streaked dwarf virus; RnMYRV, Rosellinia necatrix mycoreovirus; RpLV, raspberry latent virus; RRSV, rice ragged stunt virus; SRBSDV, southern rice black streaked dwarf virus.

On the other hand, a complete and a partial genome, both showing similarity to hepevirus, were sequenced in the metagenomic analyses from Nouragues and La Herrería respectively. We carried out RT-PCRs using specific primers for each of these viruses to identify the positive individuals: a Plain Xenops (*Xenops minutus*) for French Guiana Hepevirus (FGHEV); and a European Robin (*Erithacus rubecula*) for La Herrería Hepevirus (LHHEV). The complete genome of FGHEV was 7,595 b, unusually longer than other members of the family *Hepeviridae* (Purdy et al. 2017). Also unusually long is ORF1 (6,455 b), encoding the putative non-structural polyprotein, which shows the conserved methyl-transferase, helicase and RdRp domains but has an extensive coding region of unknown function between RdRp and the stop codon (Figure A4). The analysis using Phyre2 showed that this region located at the end of ORF1 has 43% similarity (48.7% confidence) to PTPA-like superfamily, which groups protein tyrosine phosphatases. The arrangement of ORF2 in FGHEV, completely embedded in ORF1, is different from the layout of these two ORFs in hepeviruses, where they are completely separated (Figure A5). These characteristics, however, are common in the recently described hepe-like virus group, a sister clade to *Hepeviridae*. LHHEV, though incomplete, shares some of these features with FGHEV. ORF1 is unusually long and ORF2 appears to overlap with it, in a different reading frame (Figure A5). Moreover, the incomplete LHHEV genome is 5,451 b long, which would suggest that the complete genome is also unusually long for a hepevirus. Similarity analyses carried out with RAT software along the ORF1 sequences for both viruses show that similarity with other hepe and hepe-like viruses, although low in general, peaks at the conserved motifs (Figure A6 and A7).

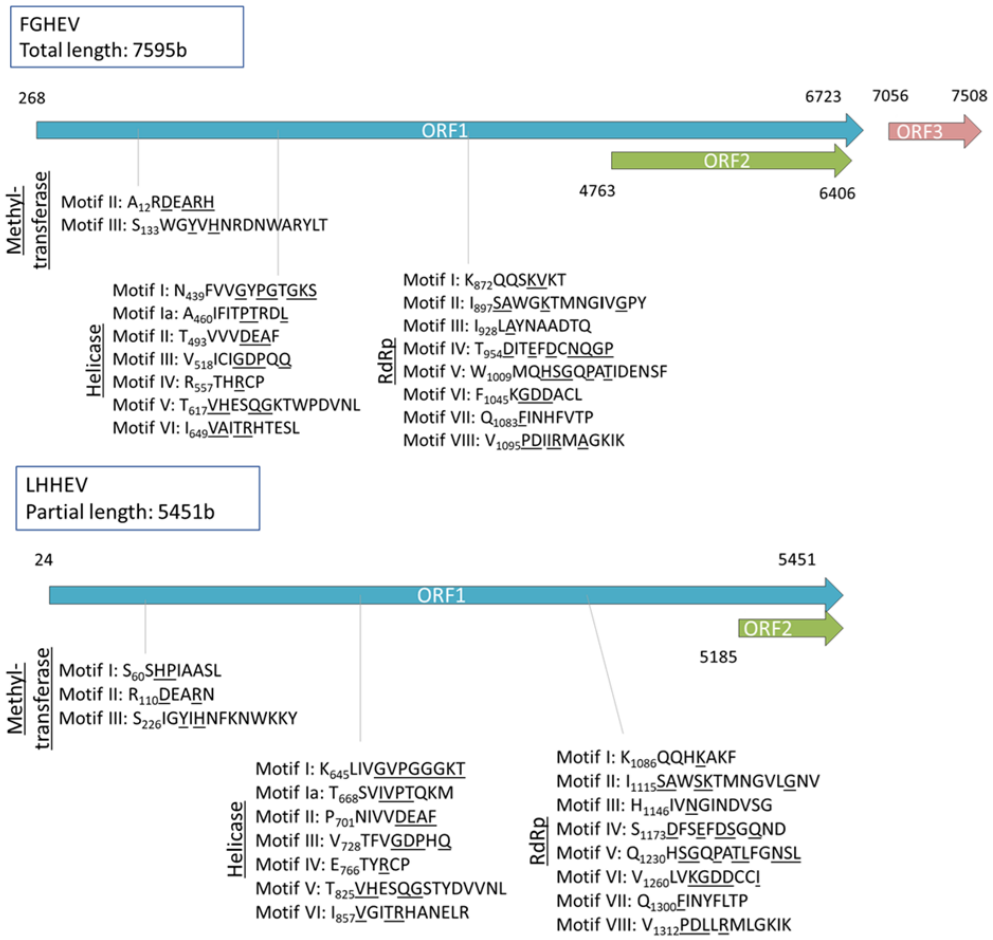


Figure A4. Putative methyl-transferase, helicase and RNA-dependant-RNA polymerase (RdRp) motifs within the ORF1 of French Guiana hepevirus (FGHEV) and La Herrería hepevirus (LHEHV). Conserved amino acids coincident with those describe by (Koonin et al. 1992) are underlined.

Phylogenetic analyses place LHHEV as a sister taxon to murine feces associated and Hubei hepe-like viruses, either using RdRp amino acid sequence or the whole ORF1 (Figure A8). On the other hand, the phylogenetic location of FGHEV is less clear. When using RdRp, it is grouped in a basal group with Elicom virus and Barns Ness breadcrumb sponge hepe-like virus 1. However, when the whole ORF1 amino acid sequence is selected to infer the phylogeny, FGHEV is located halfway between hepe-like viruses and actual hepeviruses (Figure A8).

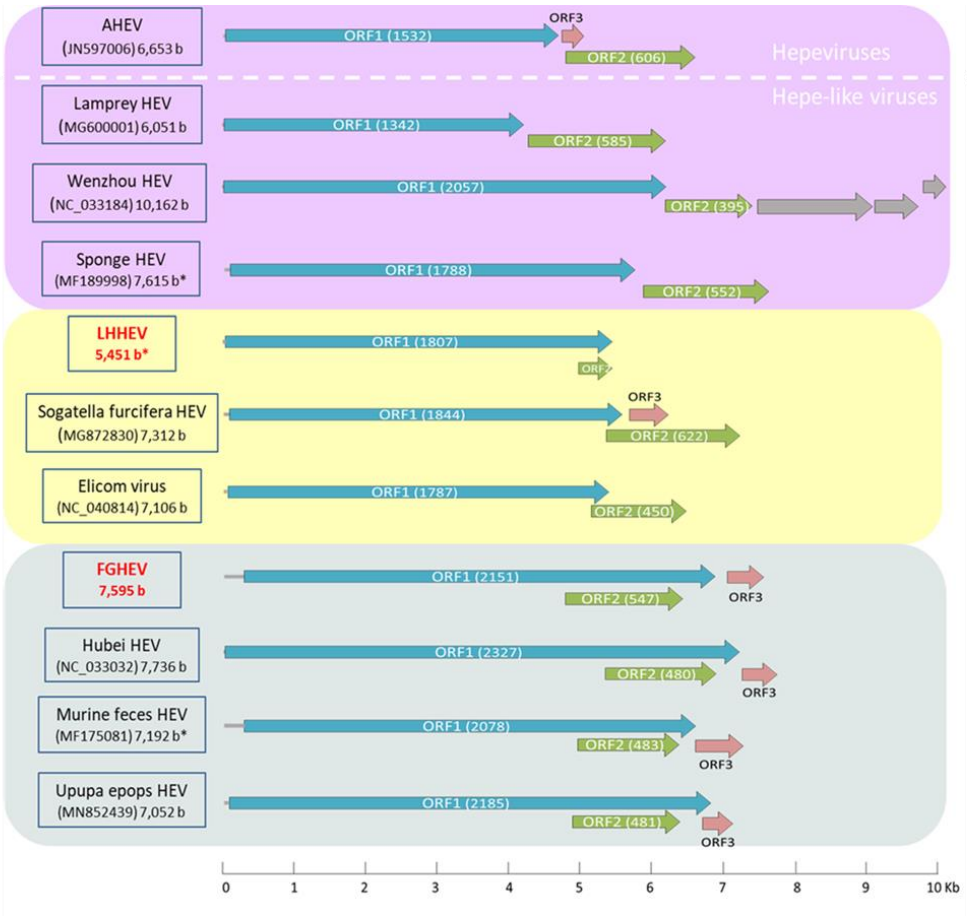


Figure A5. Schematic representation of the genomic organization of French Guiana hepevirus (FGHEV), La Herrería hepevirus (LHHEV) and other hepe-like viruses. Avian hepatitis E virus (AHEV) has been selected as a model of typical ORF arrangement in the family *Hepeviridae*. The genomes have been grouped depending on the position of ORF2: separated from ORF1 as in hepeviruses (violet), partially overlapping with ORF1 (yellow) and embedded within ORF1 (blue). The number in parenthesis indicates the lenght of the ORF and the 5´UTR is represented as a grey segment at the beginning of the genome. The grey arrows of Wenzhou HEV represent proteins of unknown function. Asterisks mark incomplete genomes. The novel hepe-like viruses from this study are highlighted in red.

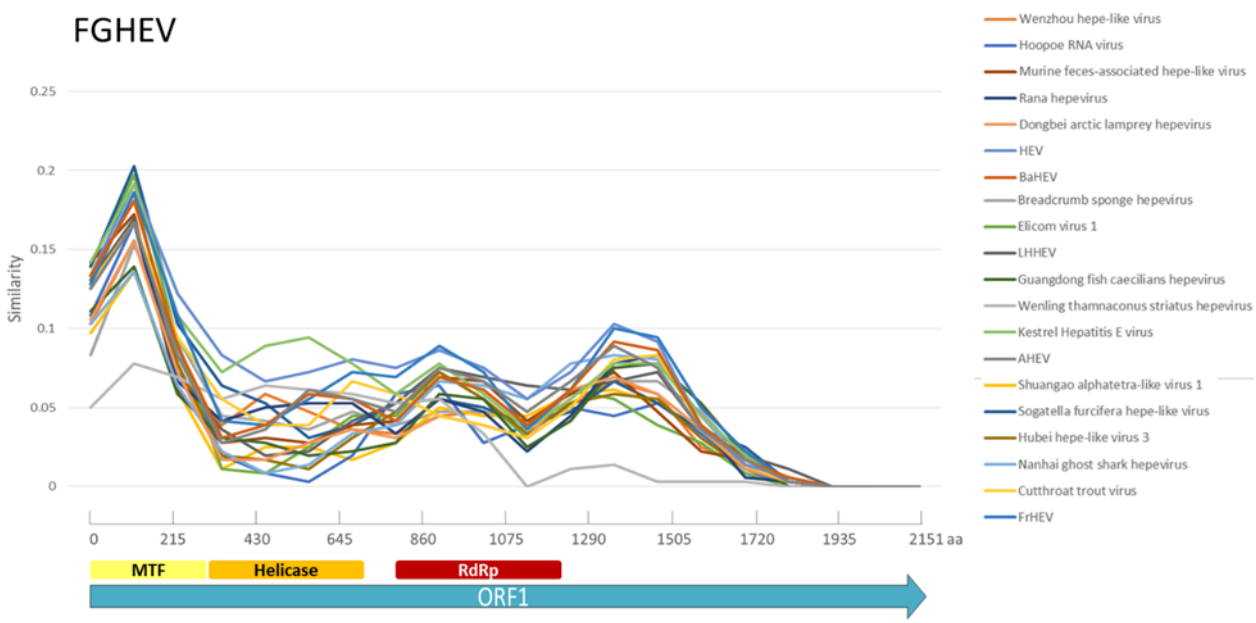


Figure A6. Similarity analysis along the ORF1 amino acidic sequence of FGHEV and other members of the family *Hepeviridae* using Recombination Analysis Tool (RAT). Color boxes indicate the conserved motifs found in the FGHEV sequence: methyltransferase (MTF), helicase and RNA-dependent RNA polymerase (RdRp).

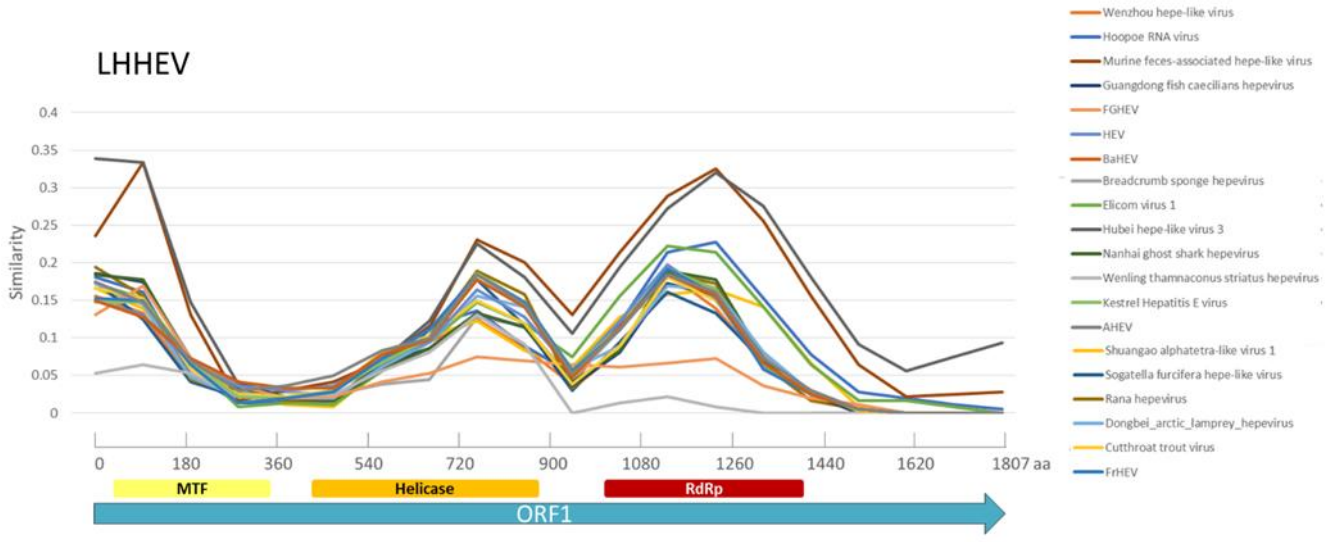


Figure A7. Similarity analysis along the ORF1 amino acidic sequence of LHHEV and other members of the family *Hepeviridae* using Recombination Analysis Tool (RAT). Color boxes indicate the conserved motifs found in the LHHEV sequence: methyltransferase (MTF), helicase and RNA-dependent RNA polymerase (RdRp).

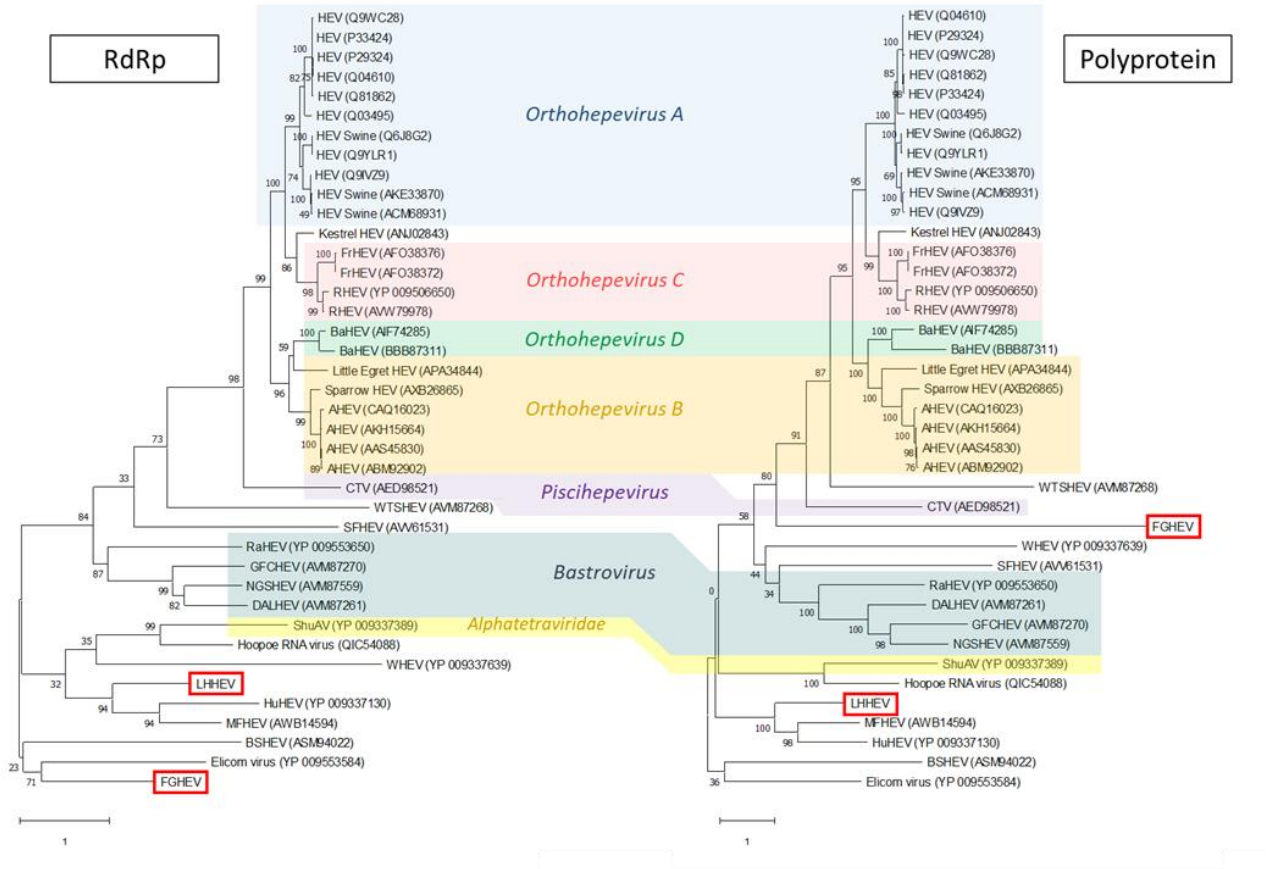


Figure A8. Maximum likelihood trees based on amino acid sequences of RNA-dependent-RNA polymerase (RdRp) and the whole ORF1 (polyprotein) of the family *Hepeviridae* using LG + G +I and LG + G methods respectively (bootstrap value = 1,000 replicates). The different genera within the family are highlighted in different colored boxes. Taxon information includes species names and GenBank accession numbers. Branch lengths measure the number of substitutions per site. AHEV, avian hepatitis E virus; BaHEV, bat hepatitis E virus; BHEV, Barns Ness breadcrumb sponge hepe-like virus 1; CTV, Cutthroat trout virus; DALHEV, Dongbei arctic lamprey hepevirus; FGHEV, French Guiana hepevirus; FrHEV, ferret hepatitis E virus; GFCHEV, Guangdong fish caecilians hepevirus; HEV, human hepatitis E virus; HuHEV, Hubei hepe-like virus 3; LHHEV, La Herrería hepevirus; MFHEV, Murine feces-associated hepe-like virus; NGSHEV, Nanhai ghost shark hepevirus; RaHEV, Rana hepevirus; RHEV, rat hepatitis E virus; SFHEV, Sogatella furcifera hepe-like virus; ShuAV, Shuangao alphatetra-like virus 1; WHEV, Wenzhou hepe-like virus 1; WTSHEV, Wenling thamnaconus striatus hepevirus

SUPPLEMENTARY MATERIAL

Chapter 1: Comparative metagenomics of Palearctic and Neotropical avian cloacal viromes reveal geographic bias in virus discovery

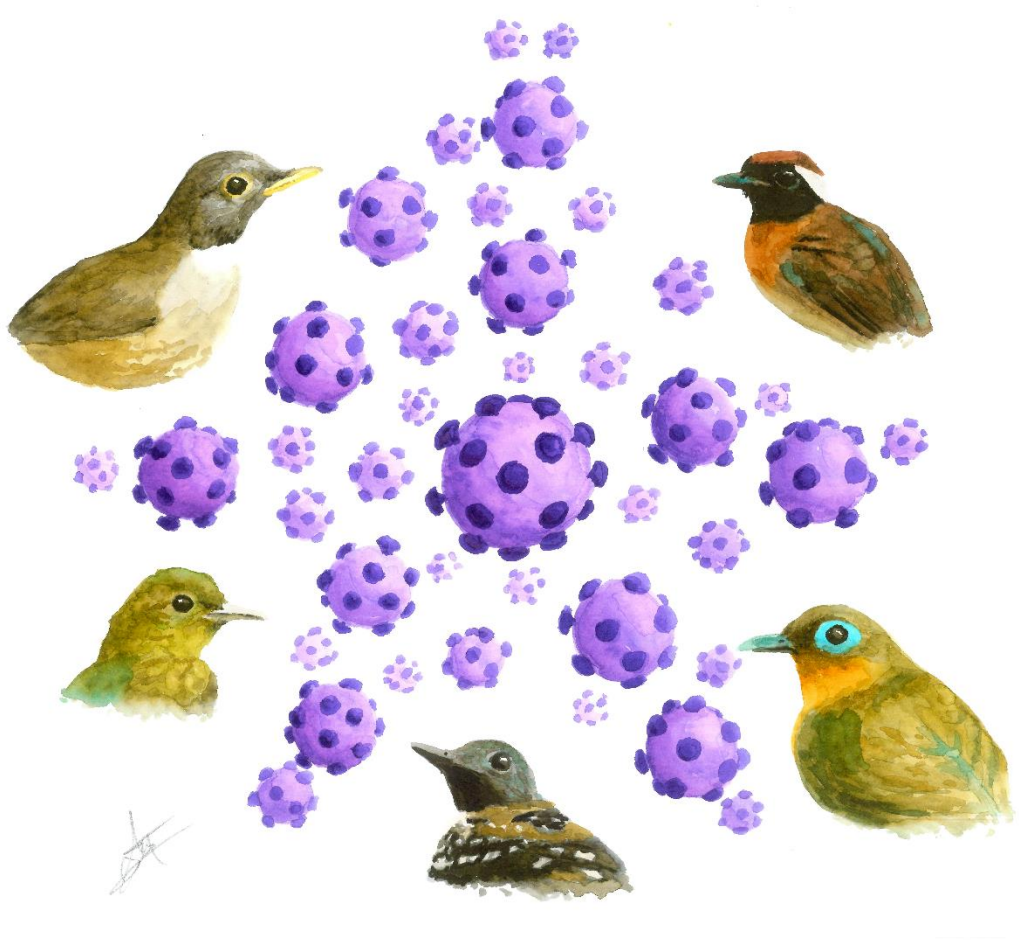
Table S1.1. List of avian species including the 50 individuals selected for deep sequencing from Nouragues Natural Reserve.

Species	Order	Number of individuals
Glyphorynchus spirurus	Passeriformes	8
Gymnopithys rufigula	Passeriformes	2
Hylophylax naevius	Passeriformes	2
Lepidothrix serena	Passeriformes	2
Leptotila rufaxilla	Columbiformes	1
Mionectes macconnelli	Passeriformes	2
Philydor erythrocercum	Passeriformes	3
Bucco capensis	Piciformes	1
Chloroceryle inda	Coraciiformes	1
Corapipo gutturalis	Passeriformes	1
Corythopis torquatus	Passeriformes	1
Pseudopipra pipra	Passeriformes	2
Myrmotherula longipennis	Passeriformes	1
Pithys albifrons	Passeriformes	2
Ramphocelus carbo	Passeriformes	2
Cyanoloxia cyanoides	Passeriformes	1
Formicarius analis	Passeriformes	1
Willisornis poecilinotus	Passeriformes	1
Micrastur ruficollis ssp. *concentricus*	Falconiformes	1
Myrmoderus ferrugineus	Passeriformes	2
Isleria guttata	Passeriformes	1
Ceratopipra erithrocephala	Passeriformes	1
Platyrinchus coronatus	Passeriformes	1
Platyrinchus saturatus	Passeriformes	1
Thamnomanes ardesiacus	Passeriformes	2
Tachyphonus surinamus	Passeriformes	1
Terenotriccus erythrurus	Passeriformes	1
Thamnomanes caesius	Passeriformes	1
Epinecrophylla gutturalis	Passeriformes	1
Turdus albicollis	Passeriformes	1
Xenops minutus	Passeriformes	1
Xiphorhynchus pardalotus	Passeriformes	1

Table S1.2. List of avian species including the 50 individuals selected for deep sequencing from La Herrería forest.

Species	Order	Number of individuals
Aegithalos caudatus	Passeriformes	1
Coccothraustes coccothraustes	Passeriformes	2
Dryobates minor	Piciformes	2
Dendrocopos major	Piciformes	1
Erithacus rubecula	Passeriformes	4
Ficedula hypoleuca	Passeriformes	4
Fringilla coelebs	Passeriformes	4
Luscinia megarhynchos	Passeriformes	1
Garrulus glandarius	Passeriformes	1
Cyanistes caeruleus	Passeriformes	4
Parus major	Passeriformes	2
Troglodytes troglodytes	Passeriformes	1
Sturnus unicolor	Passeriformes	1
Phylloscopus bonelli	Passeriformes	1
Sitta europaea	Passeriformes	1
Sylvia atricapilla	Passeriformes	10
Turdus merula	Passeriformes	8
Turdus philomelos	Passeriformes	2

Chapter 2: A novel group of avian astroviruses from Neotropical passerine birds broaden the diversity and host range of *Astroviridae*



This chapter is based on the manuscript: Fernández-Correa, I., **Truchado, D. A**., Gomez-Lucia, E., Doménech, A., Pérez-Tris, J., Schmidt-Chanasit, J., Cadar, D. & Benítez, L. (2019). A novel group of avian astroviruses from Neotropical passerine birds broaden the diversity and host range of *Astroviridae*. *Scientific reports*, *9*(1), 1-9. https://doi.org/10.1038/s41598-019-45889-3

A novel group of avian astroviruses from Neotropical passerine birds broaden the diversity and host range of *Astroviridae*

Izaskun Fernández-Correa[1], Daniel A. Truchado[1,2], Esperanza Gomez-Lucia[3], Ana Doménech[3], Javier Pérez-Tris[2], Jonas Schmidt-Chanasit[4,5,6], Daniel Cadar[4] and Laura Benítez[1]

[1] Department of Genetics, Physiology and Microbiology. Faculty of Biology. Complutense University of Madrid. José Antonio Novais, 12. 28040, Madrid, Spain

[2] Department of Biodiversity, Ecology and Evolution. Faculty of Biology. Complutense University of Madrid. José Antonio Novais, 12. 28040, Madrid, Spain

[3] Department of Animal Health. Faculty of Veterinary Medicine. Complutense University of Madrid. Avda. Puerta de Hierro s/n, 28040, Madrid. Spain

[4] Bernhard-Nocht-Institut fur Tropenmedizin, WHO Collaborating Centre for Arbovirus and Haemorrhagic Fever Reference and Research, National Reference Centre for Tropical Infectious Diseases, Bernhard-Nocht-Strasse 74, 20359, Hamburg, Germany

[5] University of Hamburg, Faculty of Mathematics, Informatics and Natural Sciences

[6] German Centre for Infection Research (DZIF), Partner Site Hamburg-Luebeck-Borstel, Greifswald-Insel Riems, Germany

These authors contributed equally: Izaskun Fernández-Correa, Daniel A. Truchado

ABSTRACT

Metagenomics is helping to expand the known diversity of viruses, especially of those with poorly studied hosts in remote areas. The Neotropical region harbors a considerable diversity of avian species that may play a role as both host and short-distance vectors of unknown viruses. Viral metagenomics of cloacal swabs from 50 Neotropical birds collected in French Guiana revealed the presence of four complete astrovirus genomes. They constitute an early diverging novel monophyletic clade within the *Avastrovirus* phylogeny, representing a putative new astrovirus species (provisionally designated as *Avastrovirus 5*) according to the International Committee on Taxonomy of Viruses (ICTV) classification criteria. Their genomic organization shares some characteristics with *Avastrovirus* but also with *Mamastrovirus*. The pan-astrovirus RT-PCR analysis of the cloacal samples of 406 wild Neotropical birds showed a community-level prevalence of 4.9% (5.1% in passerines, the highest described so far in this order of birds). By screening birds of a remote region, we expanded the known host range of astroviruses to the avian families Cardinalidae, Conopophagidae, Furnariidae, Thamnophilidae,

Turdidae and Tyrannidae. Our results provide important first insights into the unexplored viral communities, the ecology, epidemiology and features of host-pathogen interactions that shape the evolution of avastroviruses in a remote Neotropical rainforest.

INTRODUCTION

The last decades have witnessed a broadening of the global diversity of viruses thanks to rapidly evolving random amplification sequencing technology. Next Generation Sequencing (NGS) has revolutionized our knowledge by not only uncovering the diversity of viruses at an unprecedented rate, but also by allowing to gain insight into their ecology and distribution, helping to disentangle the network that connects viruses with their hosts and the geographical scenario in which such interactions take place (Bascompte 2010). However, deciphering the global virome and its structure becomes challenging not just because of the technical difficulties associated with the screening of diverse viruses, but because their many hosts are often understudied, particularly those from remote areas. The Neotropical region is one such example; being among the most biodiverse realms on Earth, Neotropical rainforests are candidates to host and conserve relevant parts of the virosphere diversity (Carroll et al. 2018). However, the knowledge of Neotropical viral diversity and their hosts is in its infancy. Given that zoonotic viruses are considered the most probable causative agents of emerging diseases (Woolhouse and Gowtage-Sequeria 2005) and that mammal and avian species are their main hosts (Taylor et al. 2001; Morse et al. 2012), the study of the wildlife viral diversity and their ecological context in these remote areas will provide important information for public and animal health to prevent potential new emerging viral epidemics (Moens and Pérez-Tris 2016; Carroll et al. 2018). For example, the great diversity of the Neotropical fauna harbors the risk of emergence of new infections in domestic animals and humans that might come into contact with these birds or ecosystem as a result of human impact on nature. Conversely, the valuable biodiversity of these regions may be threatened by invasive viruses imported into the Neotropics (Daszak et al. 2000; Keesing et al. 2010). There is growing evidence that the risk of pathogen transfer is increasing in both directions (Komar et al. 2003; Pereda et al. 2008; Ferreira-Junior et al. 2018), and a comprehensive characterization of the virome of the local fauna may be instrumental in preventing any harmful impact. However, this important knowledge would hardly become available

without exploratory approaches which address the discovery and describe the diversity of viruses that affect non-model host species.

In the last decade, astroviruses have been one of the virus families in which the number of their potential hosts has quadrupled thanks to the use of metagenomic studies (Cortez et al. 2017; Donato and Vijaykrishna 2017). New astroviruses have been discovered as part of the intestinal virome of wildlife, including bats (Hu et al. 2017) and especially birds (Honkavuori et al. 2014), although information about the presence and diversity of astroviruses in wild birds is far from being complete. Novel astrovirus genotypes have been discovered in cloacal swabs or fecal samples in a variety of bird species (Kofstad and Jonassen 2011; Chu et al. 2012b; Honkavuori et al. 2014), although observed prevalence was usually very low.

The family *Astroviridae* comprises two genera: *Mamastrovirus,* which infect mammals, and *Avastrovirus,* which infect birds. Astroviruses are non-enveloped icosahedral viruses, with a diameter of 28-30 nm. The genome of astroviruses consists of a positive sense linear single-stranded RNA molecule between 6.4 and 7.9 kb with three open reading frames (ORFs): ORF1a, ORF1b and ORF2. ORF1a and ORF1b encode two non-structural polyproteins (nsp1a and nsp1ab) which include a protease and a RNA-dependent RNA-polymerase (RdRp). In the overlapping region between ORF1a and ORF1b there is a distinctive frameshifting mechanism consisting of a heptanucleotide and a stem-loop structure (Jiang et al. 1993). ORF2 encodes a polyprotein which is the precursor of viral capsid proteins. This polyprotein is translated from a viral subgenomic RNA (sgRNA) and has two main domains: a highly conserved N terminal and a hypervariable C terminal (Pantin-Jackwood et al. 2013). A viral genome-linked protein (VPg) with a key role in genome replication or protein synthesis is also encoded in the genome (Goodfellow 2011; Fuentes et al. 2012). This protein is linked to the 5´-end of astrovirus genome and it possesses a TEEEY-like residue that has been demonstrated to be related to viral infectivity in human astroviruses (Fuentes et al. 2012).

In domestic birds, enteritis is the most frequent clinical sign associated with astrovirus infections, although they have also been correlated with other pathological conditions such as nephritis or hepatitis (Imada et al. 2000; Todd et al. 2009). However, astrovirus infections usually occur asymptomatically in this group of animals, only causing mild disease. The most frequent transmission route is fecal-oral (Guix et al. 2012; Pantin-Jackwood et al. 2013), a circumstance that increases the risk of cross-infection among

species that share habitat or get in close contact, for instance, in farms. Furthermore, astroviruses are also important for public and environmental health as a causative agent of encephalitis in cattle (Boujon et al. 2017).

The classification of astroviruses has been redefined several times since they were first discovered in 1975 (Donato and Vijaykrishna 2017). According to ICTV, the current classification does not correspond to the phylogeny of this group of viruses, and it is instead based on the host and the genetic distances (p-dist) among complete amino acid sequences of the capsid region (ORF2). The average genetic distance between groups of *Avastroviruses* is 0.704, and within the same group it ranges between 0.576 and 0.741 (Guix et al. 2012; Donato and Vijaykrishna 2017). The *Astroviridae* study group of ICTV also establishes that viruses with p-dist > 75% identity in the complete protein sequence of ORF2 should be considered as members of the same species (Donato and Vijaykrishna 2017). Currently, there are three recognized species of the genus *Avastrovirus*. *Avastrovirus 1* is apparently specific of the Galloanserae, with isolates from turkey (Turkey astrovirus-1; TAstV-1), chicken (Chicken astrovirus; CAstV), duck (Duck astrovirus; DAstV), guineafowl (Guineafowl astrovirus; GFAstV) and goose (Goose astrovirus; GAstV). *Avastrovirus 2* includes avian nephritis virus 1 and 2 (ANV-1 and ANV-2), and has been found mainly in chickens, but also in turkeys, ducks and pigeons sporadically. *Avastrovirus 3* includes isolates type 2 and 3 from turkey (TAstV-2 and TAstV-3). Therefore, most of the known data on the diversity of *Avastrovirus* comes from poultry studies, with an evident lack of information from wild species, a circumstance which is usual for most avian diseases (Fuller et al. 2012).

Using a metagenomic approach aiming to characterize the cloacal virome of birds from the primary rainforest of the Guianan shield (one of the most remote Neotropical regions) we observed a high number of viral reads related to astroviruses. Astrovirus-like contigs were assembled and four complete astrovirus genomes were successfully recovered, which we provisionally named Passerine astrovirus (PasAstV)-1-4. These four newly discovered astroviruses exhibit little similarity both among one another and with other representatives of the *Avastrovirus* genus. Therefore, we set out to describe the genomic features and hosts of these viruses, thereby contributing to broaden the knowledge of the diversity and host range of this virus family. We also placed them in the phylogeny of *Astroviridae*, where they represent a putative novel viral species.

METHODS

Sample collection

A random sampling of understory bird species was carried out in Pararé and Inselberg camps, in the Nouragues Natural Reserve, French Guiana (4º05'N, 52º40'W). The sampling area was located in a tropical rainforest, where average temperature is around 26 ºC throughout the year and relative humidity is usually high. The climate is very wet in general, with annual precipitation exceeding 3,000 mm, although there is a dry season with considerably less rainfall between August and November. Birds were mist-netted in January 2016 (rainy season) and October-November 2016 (dry season). They were taken standard morphometric measurements, ringed to avoid repeated sampling of the same individuals, and released unharmed at the site of capture. Cloacal samples were collected using sterile swabs (Nerbe Plus), which were preserved in 800 µl of universal viral transport medium (VTM) (Becton Dickinson) and kept frozen until molecular analyses. A total of 406 cloacal samples from 72 bird species were collected

Sample processing and next generation sequencing

We selected 50 samples with abundant fecal matter in cloacal swabs, including as many different bird species as possible. We grouped them to create five pools of 10 samples each (hereafter pools 1-5). The samples were vortexed, and the swabs were squeezed to release epithelial cells and discarded. The VTM was centrifuged at 13,000 rpm for 1 min to pellet out epithelial cells. VTM was removed and pellets were resuspended in 250 µl of PBS. Then, samples were subjected to 2 freeze-thaw cycles at -80 ºC to maximize the release of viral particles, and filtered through 0.45 µm pore-sized column filters at 8,000 rpm for 5 min to enrich viral particles in the flow-through. We took 50 µl of the filtrate of each sample and mixed them in five pools of 10 samples each. Each filtrate was treated with a mixture of nucleases (Turbo DNase, Ambion, Carlsbad, CA, USA; Baseline-ZERO, Epicenter, Madison, WI, USA; Benzonase, Novagen, San Diego, CA, USA; RNAse One, Promega, Fitchburg, WI, USA) to digest unprotected nucleic acids including host DNA/RNA. Finally, viral RNA/DNA was extracted with the MagMAX Viral RNA Isolation Kit (Thermo Fisher) according to the manufacturer's instructions. Standard biosafety level-2 biocontainment measures were followed during the whole process. The extracted viral RNA and DNA were subjected to random RT-PCR amplification for library preparation by using QIAseq FX DNA Library Kit (Qiagen,

Germany). Normalized samples were pooled and sequenced using 600-cycle (2 × 300 bp paired-end) MiSeq Reagent Kits v3 (Illumina, San Diego, CA) on a MiSeq platform. The generated raw reads were first qualitatively checked, trimmed and filtered to remove polyclonal and low quality reads (< 55 bases long) using CLC workbench (Qiagen). The remaining filtered raw reads were de-novo assembled separately using Trinity v2.6.642 (Grabherr et al. 2011) and CLC workbench and compared with a non-redundant and viral proteome database using BLASTx with an E-value cut-off 0.001. The virus-like contigs and singlets were further compared to all protein sequences in non-redundant protein databases with a default E-value cutoff of 0.001. The viral metagenomics output has been visualized and analyzed in MEGAN (Huson et al. 2016).

Prevalence of astroviruses in the cloacal swabs

A total of 356 cloacal samples, which were not subjected to deep-sequencing, were processed differently. Only half of the VTM volume (400 µl) was taken after vortexing. Swabs were squeezed and discarded, samples were centrifuged and VTM was removed in the same way as for samples analyzed by deep sequencing. However, the pellets were resuspended in 40 µl of PBS, avoiding freeze-thaw cycles, filtering or nuclease treatment. These samples were grouped in 36 pools, 34 of which contained 10 individuals, one of 11 individuals and one of five individuals. Viral RNA/DNA was extracted using the same protocol as for samples subjected to deep sequencing.

Degenerate primers described by Todd et al. (2009) for pan-astrovirus RT-PCR, which amplify a fragment of the RNA-dependent RNA polymerase (RdRp) region of the ORF1b, were used to detect the presence of astroviruses. Due to the presence of PCR inhibitors in feces, 5 µl of a 10^{-1} dilution from the individual RNA/DNA extraction were added as a template. PCR assays were performed with a final volume of 25 µl reaction mixture consisting of 5 µl extracted RNA/DNA, 12.5 µl *Verso 1-Step RT-PCR Hot-Start Kit* (Thermo Fisher), 0.5 µl of each primer (5 pmol), and ddH_2O up to 25 µl using a cycling condition as follows: synthesis of cDNA at 50 ºC for 15 min followed by denaturation at 95 ˚C for 15 min; 45 amplification cycles were performed at 95 ˚C for 20 s, 45 ˚C for 30 s, 72 ˚C for 1 min. Final extension was at 72 ˚C for 5 min.

Genomic analysis of the four novel PasAstV

The ORFs were predicted using ORFfinder (NCBI: https://www.ncbi.nlm.nih.gov/orffinder/). TMHMM server v2.0 (Krogh et al. 2001) was

employed for prediction of transmembrane helices in proteins; cNLS Mapper (Kosugi et al. 2009) for nuclear localization signals and NCBI Conserved domains (NCBI: https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) for the location of serine protease domains, RdRp and the conserved region of the capsid. Secondary structures were predicted by FoldIndex (Prilusky et al. 2005) for locating the viral protein associated with the genome (VPg), and by Multicoil scoring form (Wolf et al. 1997) for locating the coiled-coil regions. Hairpin loops were analysed using Mfold (Zuker 2003).

Phylogenetic and taxonomic analysis

Genome sequence analysis, genomic organization and multiple alignments were performed using Geneious v11 (Biomatters, New Zealand), EditSeq and SeqMan tools of the DNASTAR 5.0 software package (DNASTAR, Madison, WI), and BioEdit Sequence Alignment Editor (Hall 1999).

Evolutionary relationships of the four novel PasAstV with representative avastroviruses were determined by the construction of the phylogenetic tree based on amino acid sequences of the ORF2 gene. The initial complete ORF2 nucleotide data set was pruned from the potential recombinants using the various methods for recombination detection implemented in RDP4 (Martin and Rybicki 2000). Recombinations were considered only when they were detected by more than three out of seven methods having significant p-values ($p < 0.05$). The phylogenetic tree based on ORF2 amino acid sequences was inferred using the Bayesian Markov chain Monte Carlo (MCMC) approach available in BEAST v1.84 (Drummond and Rambaut 2007). The analysis was performed under the best fit amino acid substitution model identified as LG+Γ+I+F using Bayesian Information Criterion implemented as the model selection framework in jModelTest (Darriba et al. 2012) and MEGA X (Kumar et al. 2018). Monte Carlo Markov Chains (MCMC) were run for 10^7 generations, sampling every 1,000 trees. Traces were inspected for convergence with Tracer 1.5 (Drummond and Rambaut 2007). The 10,000 resulting trees were summarized with TreeAnnotator v2.1.2 (Drummond and Rambaut 2007) and the phylogeny with branch posterior probabilities was displayed in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). We also computed pairwise amino acid genetic distances (p-dist) of the capsid region (ORF2) between isolates, between astrovirus groups and between members of the same group. To do so, we used MEGA X program using the bootstrap method with 1,000 replicates and partial deletion (95%). The identity matrices were generated with the MUSCLE tool (McWilliam et al. 2013)

(https://www.ebi.ac.uk/Tools/msa/muscle/), and transformed to HeatMaps with Morpheus software (https://software.broadinstitute.org/morpheus).

RESULTS

Diversity and host range of astroviruses

The screening by pan-astrovirus RT-PCR of all 406 cloacal samples retrieved 20 individuals positive for astrovirus, representing a prevalence of 4.9%. Although our sample included birds of six different orders (Apodiformes, Columbiformes, Coraciiformes, Falconiformes, Passeriformes and Piciformes), all positive individuals were passerines. Most of the infected birds belonged to the family Thamnophilidae, but other families such as Cardinalidae, Conopophagidae, Furnariidae, Tyrannidae and Turdidae were also positive for viral infection (Table 1).

Identity matrices (Figure 1) showed high divergences of the four novel astroviruses from French Guianan passerine birds, compared to previously known astroviruses. Amino acid sequence identity of capsid protein (ORF2) varied from 35% to 93% within the genus *Avastrovirus*, but we found lower identities when we compared these four novel genomes to one another (24-29%), or when we compared them to other avastroviruses (15-36%). An ORF1b and ORF2 amino acid distance matrix analysis generated with HeatMaps confirmed the clear demarcation between passerine avastroviruses and relatives of other *Avastrovirus* genus (Figure 1). Genetic distance analysis also indicated that the passerine avastroviruses could be further subgrouped into four supported clades: *Passerine astrovirus 1* (PasAstV-1), *Passerine astrovirus 2* (PasAstV-2), *Passerine astrovirus 3* (PasAstV-3) and *Passerine astrovirus 4* (PasAstV-4), since distance values among these four astroviruses double the values found within the other *Avastrovirus* groups (Table 2). The sequences of the complete genomes of the four PasAstV were deposited in GenBank under accession numbers MK096773-76. The closest relative to all four PasAstV in databases was the only one passerine astrovirus described, found in a black-naped monarch. Its genome is partially sequenced and no data related to ORF2 is available, but it shows a 59-63% amino acid identity in ORF1b (RdRp) with the four PasAstV.

Table 1. Percentage of birds positive for astrovirus within each of the avian families that provided positive samples in French Guiana.

Family	Positive birds for astrovirus	Captured birds	% infected/sampled birds
Furnariidae	3	70	4
Thamnophilidae	13	170	8
Turdidae	1	8	13
Tyrannidae	1	40	3
Others	2	118	2

Table 2. Within genetic distances (p-dist) of the ORF2 amino acid sequence for the phylogenetic groups *Avastrovirus 1-5*

Species	p-dist	Error
Avastrovirus 1	0.449	± 0.013
Avastrovirus 2	0.394	± 0.014
Avastrovirus 3	0.146	± 0.010
Avastrovirus 4	0.307	± 0.015
Avastrovirus 5	0.671	± 0.015

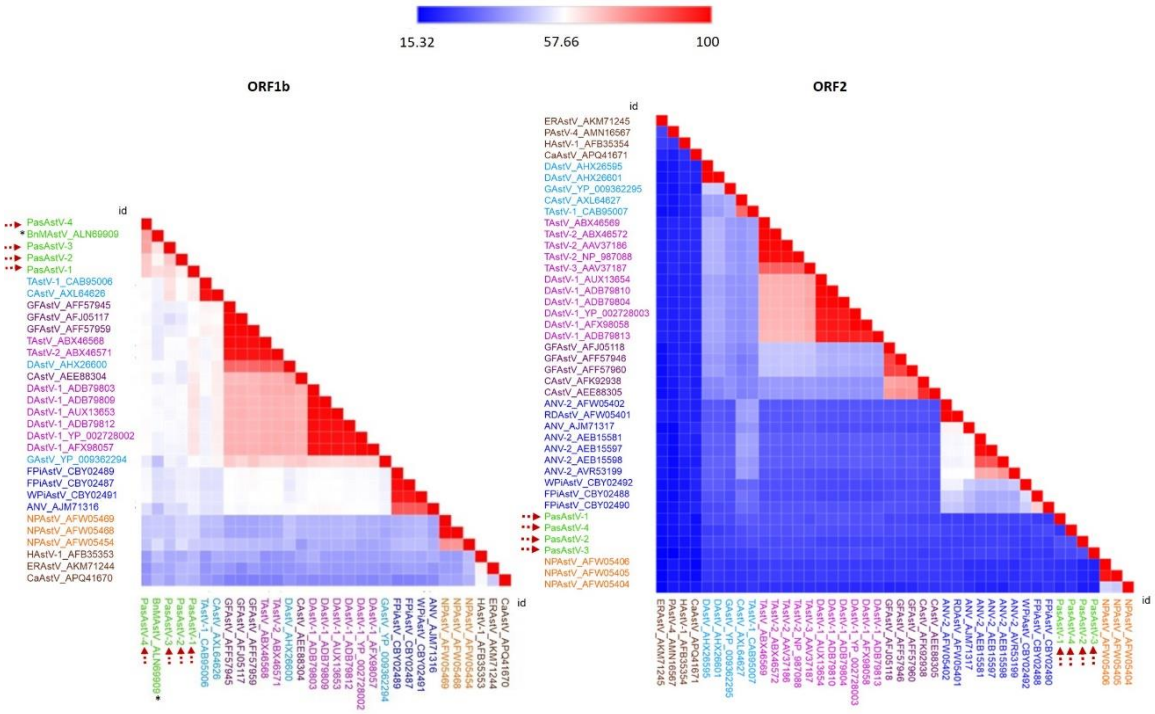


Figure 1. Colour-coded pairwise identity matrix generated from complete ORF1b and ORF2 amino acid sequences of the four novel PasAstV and representative members of the *Astroviridae* family. The PasAstV-

1-4 are marked with arrows and the black-naped monarch astrovirus isolate (BnMAstV) is marked with an asterisk. Each coloured cell represents the percentage identity score between two sequences. A colour key indicates the correspondence between pairwise identities and the colours displayed in the matrix. The *Avastrovirus* strains are colored according to the ICTV current species and our proposed new classification (see Figure 4). Difference in size between the two heatmaps is due to the different availability of sequences for both ORFs in Genbank.

Genomic characterization of the four PasAstV

The genomic organization of the four novel PasAstV complete genomes showed that each ORF is in a different reading frame for PasAstV-1 and PasAstV-3. ORFs 1b and 2 of PasAstV-2 are in +1 frame, while PasAstV-4 has ORF1a and ORF2 in the same frame and the ORF1b on +1 frame (Figure 2A). None of them have exhibited the conserved genomic RNA promoter sequence at 5' UTR (Pantin-Jackwood et al. 2013). PasAstV-2, PasAstV-3 and PasAstV-4 preserve the stem-loop structure followed by the ribosomal frame-shift heptameric signal (AAAAAAC) at the 3' end of ORF1a like the rest of *Avastrovirus*. However, the setting of the ORF1a stop codon in PasAstV-1, between the stem-loop structure and the heptameric signal, shows a typical *Mamastrovirus* structure (Pantin-Jackwood et al. 2013) (Figure 2B). Distinctive sgRNA promoter sequences (5´-AUUUGGAGNGGNGGACCNAAN-3´) were found at the 3' end of the four ORF1b (Pantin-Jackwood et al. 2013). A structure similar to the stem-loop 2-like (s2m) described for astroviruses (Pantin-Jackwood et al. 2013) was also observed at the 3' end of PasAstV-1, PasAstV-2 and PasAstV-3, but not in PasAstV-4. Several conserved protein domains were found in each PasAstV ORF (Table S2.1) except for a hypervariable region in ORF1. Conserved VPg sequences, with some amino acid variations, were conserved in all four novel astroviruses (Figure 3).

Phylogeny and taxonomy

The analysis carried out to remove potential ORF2 recombinants from our database showed no recombination events in the sequences of the four PasAstV. The Bayesian maximum clade credibility (MCC) tree of astrovirus ORF2 amino acid sequences revealed six well supported lineages of avian astroviruses (Figure 4), roughly corresponding to *Avastrovirus* species based on their genetic divergences. Also, we performed another phylogenetic analysis removing those sequences with possible recombination events (some members of GFAstV; Wood pigeon Astrovirus, WpiAstV;

DAstV; CAstV; TAstV; and ANV), obtaining the same well supported clades, ruling out the possible influence that recombination could have had in our results.

The species *Avastrovirus 2* and *Avastrovirus 3* were recovered with maximum branch support, but *Avastrovirus 1* formed a paraphyletic group with respect to *Avastrovirus 3*, which was included in its clade with the same uniqueness and statistical support as the two major clades of *Avastrovirus 1*. Therefore, if *Avastrovirus 1* and *Avastrovirus 3* are to be kept as different species, we propose distinguishing between two evolutionarily distinct groups within the species *Avastrovirus 1* (*Avastrovirus 1a* and *1b* in Fig. 4). The three sequences of Northern pintail named ”Avastroviruses Group 3” by Chu et al.(2012b) formed a distinct, early diverging clade, which we named *Avastrovirus 4*. The four newly discovered passerine astrovirus from French Guiana formed another distinct clade, which we provisionally named *Avastrovirus 5*. Between-group mean genetic distances for *Avastrovirus 4* (0.727-0.745) and *Avastrovirus 5* (0.672-0.745) are greater than genetic distances among the three previously described *Avastrovirus* species (0.550-0.680). The *Avastrovirus 5* clade had very long internal branches compared to other groups of the same phylogenetic level, and their members scored the highest intraspecific genetic distance in average (0.671).

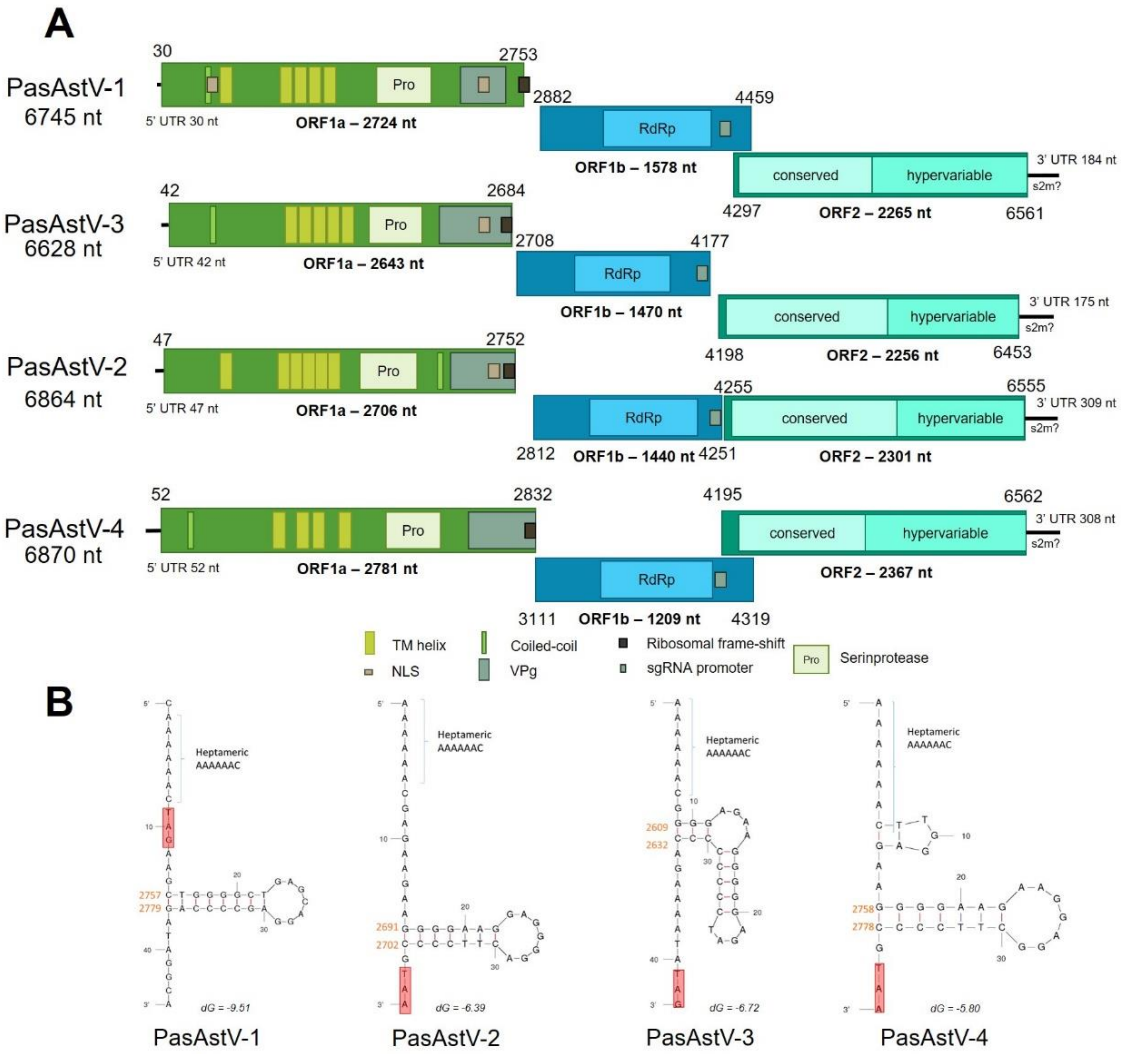


Figure 2. A) Schematic representation of the genomic organization of the four novel PasAstV. NLS, nuclear localization sequence; VPg, viral protein associated with the genome; RdRp, RNA-dependent RNA polymerase; sgRNA, subgenomic RNA; s2m, stem-loop 2-like. B) Prediction of secondary RNA structure of the ribosomal frame-shift at the 3' end of ORF1a. The termination codon of the ORF is shown in a red rectangle.

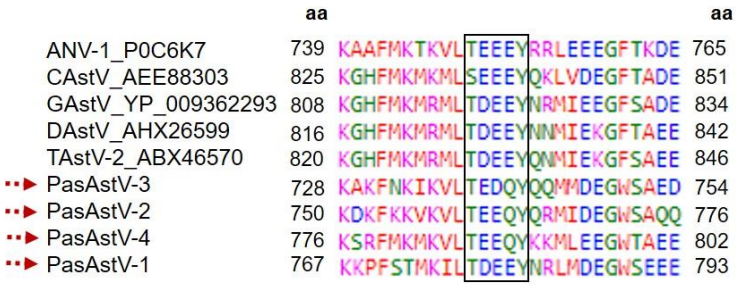


Figure 3. Alignment of the partial protein sequences of the ORF1a of passerine astroviruses (PasAstV-1-4) and some *Avastrovirus* strains at the end of ORF1a (VPg region). The sequence variation is given with respect to the conserved VPg TEEEY motif (black rectangle).

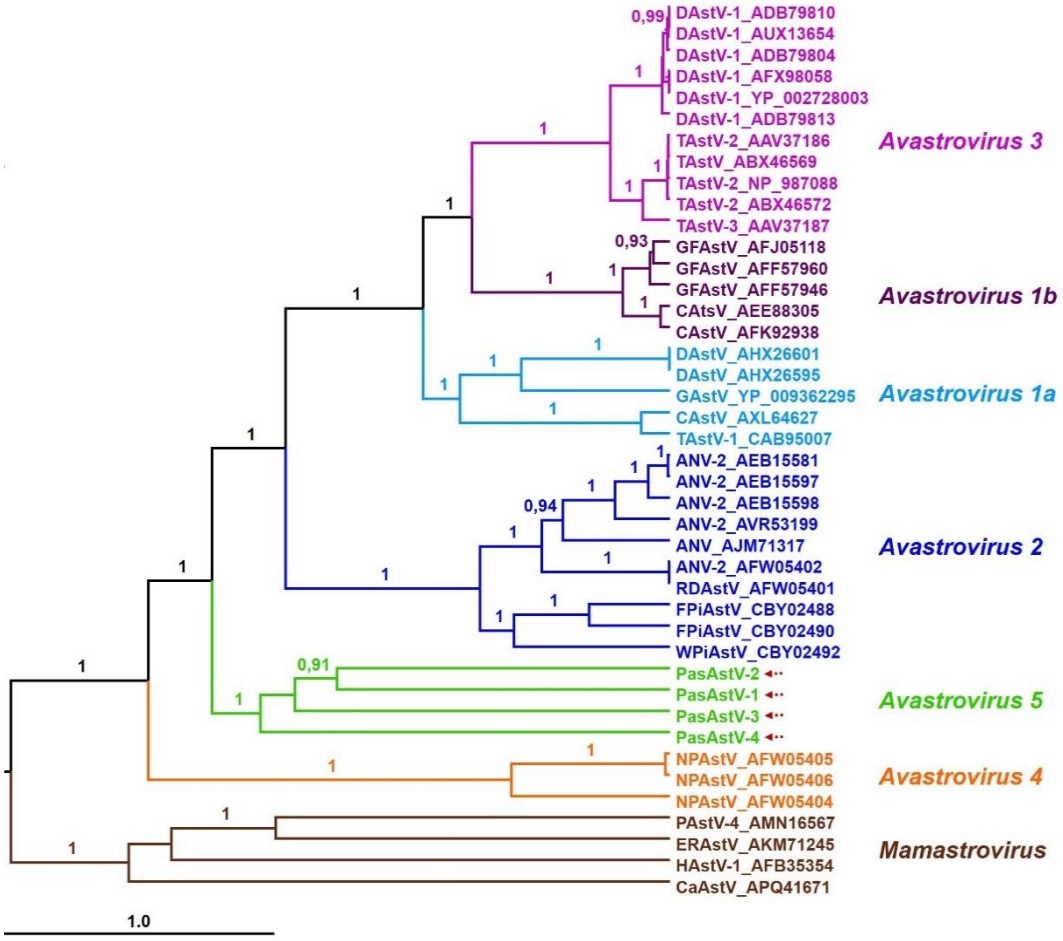


Figure 4. Bayesian maximum clade credibility (MCC) tree based on complete amino acid sequences of ORF1 showing the phylogenetic placement of the four novel PasAstV from this study compared with representative members of the *Avastrovirus* genus. Bayesian posterior probabilities (≥ 90%) are indicated at the nodes. The main lineages are indicated to the right of the tree. Taxon information includes strain names and GenBank accession numbers. The *Avastrovirus* strains are colored according to the ICTV current species and our proposed new classification. Scale bar indicates mean number of amino acid substitutions per site. Four representative members of *Mamastrovirus* are included as outgroup. ANV, Avian nephritis virus; CAstV, Chicken astrovirus; CaAstV, Canine astrovirus; DAstV, Duck astrovirus; ErAstV, European roller astrovirus; FPiAstV, Feral pigeon astrovirus; GAstV, Goose astrovirus; GFAstV, Guinefowl astrovirus; HAstV; Human astrovirus; NPAstV, Northern pintail astrovirus; PAstV, Porcine astrovirus; PasAstV, Passerine astrovirus; RDAstV, Rock dove astrovirus; TAstV, Turkey astrovirus; WPiAstV, Wood pigeon astrovirus

DISCUSSION

Metagenomics has become a powerful tool, capable of characterizing the diversity of viral communities in different ecosystems. A large amount of novel astroviruses has been discovered during the last decade, which lead to the redefinition of the *Astroviridae* classification (Cortez et al. 2017; Donato and Vijaykrishna 2017). In this study, four novel and genetically distinct astroviruses, provisionally designated as PavAstV-1-4 have been described in cloacal samples from apparently healthy populations of Neotropical passerines from French Guiana.

Our results significantly expanded the known range of hosts of avastroviruses and they provide further insight into genetic diversity, evolution and population structure of these viruses. Earlier reports on astroviruses from passerines were scarce or absent, with only one representative member detected in a black-napped monarch (*Hypothymis azurea*) from Cambodia (Mendenhall et al. 2015). However, our findings put forward passerines as putative cornerstone hosts of these viruses in the avifauna of the interior primary forest of the Guianan shield. We documented the presence of astroviruses in 14 species of passerine birds of the families Thamnophilidae, Cardinalidae, Conopophagidae, Furnariidae, Tyraniidae and Turdidae. The avifauna of this region is largely isolated from bird migratory routes, with most species being local residents (Thiollay 2002; Brûlé and Touroult 2014). In fact, we have never captured Nearctic migrants in various years of fieldwork at the Nouragues reserve. Therefore, further research is needed to clarify whether passerines are important astrovirus hosts worldwide (which have remained undersampled in previous research), or the pattern we found is a local occurrence singularly evolved in a remote and isolated avifauna. In relation to this, it is important to note that we found higher community-level prevalence (4.9%) than the observed in other previous studies based on cloacal swabs (1.7% in a study conducted in Cambodia and Hong Kong) (Chu et al. 2012b). The difference becomes still higher if it is restricted to passerine birds, for which only one individual tested positive out of 199 screened in Cambodia by Mendenhall et al. (2015). This represents 0.8% prevalence, to be compared with 5.1% of passerines infected in French Guiana (20 positive out of 395 tested).

The four novel astroviruses found in French Guiana contribute to substantially broaden the genetic and phylogenetic diversity of the genus *Avastrovirus*. The phylogenetic analysis of ORF2 amino acid sequences of *Avastrovirus* in our study revealed wider phylogenetic diversity than previously thought for the respective genus. Our analysis

added two new, early diverging groups to the known diversity of *Avastrovirus* genus, which before this study was composed of three recognized and accepted avastrovirus species (*Avastrovirus 1*, *2* and *3*) (Simmonds et al. 2017). According to our phylogenetic analysis, *Avastrovirus 1* is a paraphyletic group. Based on phylogenetic evidence, *Avastrovirus 3* should be a member of the same species, although the internal structure of *Avastrovirus 1* recommends dividing the whole group into three clades, with *Avastrovirus 1b* being sister to *Avastrovirus 3* and *Avastrovirus 1a* as a distinct, earliest diverging clade in that group. This topology, where *Avastrovirus 1* is not a monophyletic group, can be observed in previous phylogenetic analyses based on full-length ORF2 amino acid sequences (Liu et al. 2014, 2018; Zhang et al. 2017), although this matter is not addressed in their research. The four sequences recovered from French Guianan passerines form a distinct novel clade (*Avastrovirus 5*) that is sister to the group formed by *Avastrovirus 1,2* and *3*, although its members are more divergent from one another than the members of other clades. The fact that passerine astroviruses from French Guiana cluster together but form a clade with long internal branches suggests the possibility that greater diversity of viruses exists in this region, hidden by incomplete sampling in our very local study of birds from the Guianan shield. The diversity of *Avastrovirus* was further broadened by the recognition of *Avastrovirus 4*, a group of early diverging viruses found in a relatively well-sampled group of birds (ducks). Astroviruses included in *Avastrovirus 4* already appeared forming a different clade when they were first described in both RdRp and ORF2 phylogenetic trees (Chu et al. 2012b). Our proposal of two new *Avastrovirus* species is supported by the large between-group mean genetic distances in the amino acid sequences of ORF2 introduced by these viruses, with higher values for *Avastrovirus 4* (astroviruses from Northern pintail) and *Avastrovirus* 5 (astroviruses from French Guianan passerines) than for the previously described species *Avastrovirus 1*, *Avastrovirus 2* and *Avastrovirus 3*. Leaving aside the taxonomic debate, our results demonstrate that the real diversity of avian astroviruses is yet to be discovered, and sampling in remote areas can be critical to gain full insight into its distribution across regions and hosts.

The genomic organization of the novel passerine avastroviruses from French Guiana involved some differences on ORF layout. The genomes PasAstV-1 and PasAstV-3 have each ORF in a different reading frame, a special characteristic that only occurs in the DAstV isolate described by Fu et al. (2009), since astroviruses have usually two ORFs

on the same reading frame. PasAstV-4 has ORF1b in +1 frame, the same as ANV-1 (Guix et al. 2012). Furthermore, ORF1b and ORF2 from PasAstV-2 are on +1 reading frame, which is common in some *Mamastrovirus* and *Avastrovirus*. PasAstV-1-4 show the conserved tyrosine residue within the TEEEY-like domain in the VPg putative protein. The ORF sizes of astroviruses from French Guiana are more similar to those from *Mamastrovirus* than those from *Avastrovirus*. The 5' ends of the four passerine astroviruses are within *Astroviridae* range (11-85 nt) (Méndez et al. 2012), although they are longer than those described for other *Avastrovirus*. As in the European roller astrovirus isolate, phylogenetically related to *Mamastrovirus*, passerine astroviruses lack the gRNA promoter sequence on their 5' end (Pankovics et al. 2015). It is also interesting the *Mamastrovirus*-like structure present on the ribosomal frame-shift signal located between the ORF1a and 1b of PasAstV-1. Stem-loop location within AAAAAAC heptamer and the stop-codon in the ORF1a suggests that this astrovirus likely originated from an ancestor resulted from a recombination event between *Mamastrovirus* and *Avastrovirus*, even though recombination between the two genera have not been described so far (Donato and Vijaykrishna 2017). Likewise, only one case of recombination in a human astrovirus between ORF1a and ORF1b has been described (Wolfaardt et al. 2011). These *Mamastrovirus*-like features in avian astroviruses together with the wide genetic variety of *Astroviridae* have been proposed as a sign of possible cross-species transmission (Pankovics et al. 2015).

The four astroviruses discovered in passerines of French Guiana, provisionally named as *Avastrovirus 5*, imply an increase in the diversity of *Astroviridae*, because, although they have a typical *Avastrovirus* genomic structure, they also share some characteristics of the members of *Mamastrovirus*. Interestingly, only representatives of *Avastrovirus 5* have been found in birds from French Guiana (including a wide representation of avian lineages). One explanation for this could be that previously known diversity of astroviruses is mainly restricted to particular hosts (such as chickens, ducks or turkeys) which were not present in our study. Alternatively, French Guianan astroviruses likely represent an endemic diversity, which would persist in this remote and greatly isolated area as no migratory avian species or other major sources of interchange such as poultry trading occurs. Sampling typical hosts of *Astrovirus 1*, *2* and *3* in French Guiana would help answer this question. Astroviruses are present in the virome of some animals (Pantin-Jackwood et al. 2013; Ng et al. 2014), but there are no reports about astroviruses as a

causative agent of disease in wild birds, so they could be only the hosts of these viruses. Therefore, our results pave the road for further research about the evolutionary origin, geographic distribution and ecological relationships of these viruses, an interest that can be extended to other virus families.

SUPPLEMENTARY MATERIAL

Chapter 2: A novel group of avian astroviruses from Neotropical passerine birds broaden the diversity and host range of *Astroviridae*

Table S2.1. Comparison of the conserved domains in the protein sequences of the four novel PasAstV. The number of conserved domains in each genome is underlined. The size in amino acids of each sequence is shown (in parentheses). NLS: nuclear localization signal; VPg: viral genome-binding protein; RdRp: RNA-dependent RNA polymerase.

ORF1a (aa)	**PasAstV-1 (908)**	**PasAstV-2 (902)**	**PasAstV-3 (881)**	**PasAstV-4 (926)**
Transmembrane helix	5: 161-183, 328-350, 362-384, 399-416, 423-445	6: 168-190, 308-330, 340-362, 369-388, 398-420, 427-449	5: 298-320, 330-352, 359-378, 388-410, 417-439	4:183-202, 342-364, 374-396, 443-465
NLS	2: 154 and 840	1: 822	1: 843	0
Coiled-coils	1: 126-162	1: 703-721	1: 127-155	1:138-158
Serinprotease	1: 507-630	1: 517-627	1: 491-616	1:518-644
VPg	1: 760-820	1: 720-301	1: 706-880	1:741-926
ORF1b (aa)	**PasAstV-1 (526)**	**PasAstV-2 (480)**	**PasAstV-3 (490)**	**PasAstV-4 (402)**
RdRp	127-398	131-410	142-419	57-334
ORF2 (aa)	**PasAstV-1 (755)**	**PasAstV-2 (767)**	**PasAstV-3 (752)**	**PasAstV-4 (788)**
Conserved	36-396	83-406	68-408	84-459
Hypervariable	397-755	407-767	408-752	459-788

Chapter 3: A novel and divergent gyrovirus with unusual genomic features detected in wild passerine birds from a remote rainforest in French Guiana

This chapter is based on the manuscript: **Truchado, D. A.**, Diaz-Piqueras, J. M., Gomez-Lucia, E., Doménech, A., Milá, B., Pérez-Tris, J., Schmidt-Chanasit, J., Cadar D. & Benítez, L. (2019). A Novel and Divergent Gyrovirus with Unusual Genomic Features Detected in Wild Passerine Birds from a Remote Rainforest in French Guiana. *Viruses*, *11*(12), 1148. https://doi.org/10.3390/v11121148

A novel and divergent gyrovirus with unusual genomic features detected in wild passerine birds from a remote rainforest in French Guiana

Daniel A. Truchado[1,2], José Manuel Diaz-Piqueras[1], Esperanza Gomez-Lucia[3], Ana Doménech[3], Borja Milá[4], Javier Pérez-Tris[2], Jonas Schmidt-Chanasit[5,6] , Daniel Cadar[5] and Laura Benítez[1]

[1] Department of Genetics, Physiology and Microbiology. Faculty of Biology. Complutense University of Madrid. José Antonio Novais, 12. 28040, Madrid, Spain

[2] Department of Biodiversity, Ecology and Evolution. Faculty of Biology. Complutense University of Madrid. José Antonio Novais, 12. 28040, Madrid, Spain

[3] Department of Animal Health, Faculty of Veterinary Medicine, Complutense University of Madrid. Avda. Puerta de Hierro s/n, Madrid 28040, Spain

[4] National Museum of Natural Sciences, Spanish National Research Council (CSIC), Calle José Gutiérrez Abascal 2, Madrid 28006, Spain.

[5] Bernhard Nocht Institut for Tropical Medicine, WHO Collaborating Centre for Arbovirus and Haemorrhagic Fever Reference and Research, National Reference Centre for Tropical Infectious Diseases, Bernhard-Nocht-Strasse 74, 20359, Hamburg, Germany

[6] University of Hamburg, Faculty of Mathematics, Informatics and Natural Sciences, Mittelweg 177, 20148 Hamburg, Germany.

ABSTRACT

Sequence-independent amplification techniques have become important tools for virus discovery, metagenomics, and exploration of viral diversity at the global scale, especially in remote areas. Here, we describe the detection and genetic characterization of a novel gyrovirus, named GyV11, present in cloacal, oral, and blood samples from Neotropical wild birds in French Guiana. The molecular epidemiology revealed the presence of GyV11 only in passerine birds from three different species at a low prevalence (0.73%). This is the first characterization and prevalence study of a gyrovirus carried out in resident wild bird populations in a remote region, and provides evidence of the fecal–oral route transmission and local circulation of the virus. The molecular phylogeny of gyroviruses reveals the existence of two distinct gyrovirus lineages in which GyV11 is phylogenetically distinct from previously reported gyroviruses. Furthermore, GyV11 is placed basal in the gyrovirus phylogeny, likely owing to its ancestral origin and marked divergence. This study also provides important insights into the ecology, epidemiology, and genomic features of gyroviruses in a remote Neotropical rainforest. The pathogenesis

of this virus in avian species or whether GyV11 can infect humans and/or chickens needs to be further investigated.

INTRODUCTION

In recent years, numerous novel viruses and virus variants have been discovered by virus metatranscriptomics and metagenomics in different bird species, mainly in wild birds (Chan et al. 2015; Bodewes 2018; Duarte et al. 2019). Nevertheless, there is still very little known about viruses that circulate in wild birds, especially in remote areas, and their possible interaction with domestic fowl or even their zoonotic potential. Next generation sequencing (NGS) technology allows the identification of, in principle, all the viruses present in a given sample, which improves our knowledge on the viral diversity and evolution in that area and helps estimate the potential risk to domestic fowl or humans (Kapgate et al. 2015).

Gyroviruses (GyVs) are non-enveloped icosahedral viruses whose genome is a circular, single-stranded DNA molecule of approximately 2 Kb. They are grouped in the genus *Gyrovirus*, which has been recently reclassified in the *Anelloviridae* family (Rosario et al. 2017). Chicken anemia virus (CAV) is so far known to be the only pathogenic member of the genus and has long been the prototype and sole member of the genus *Gyrovirus* (Rosario et al. 2017). CAV was first reported in 1979 affecting domestic chickens (*Gallus gallus domesticus*) (Yuasa et al. 1983), and its infection is an economically important clinical and subclinical disease in young chickens, with a worldwide distribution (Eltahir et al. 2011). The other members of the genus *Gyrovirus* are GyV2 through GyV10 and the human gyrovirus (HGyV1). HGyV1 was identified in healthy French blood donors, as well as in blood samples from solid organ transplant recipients and in an HIV-infected person from Italy (Maggi et al. 2012; Biagini et al. 2013). This genus includes very divergent members, mainly isolated from chicken faeces, meat, or brain tissue (Rijsewijk et al. 2011; dos Santos et al. 2012; Zhang et al. 2014), as well as from human skin, faeces, and blood (Chu et al. 2012a; Gia Phan et al. 2013; Phan et al. 2015) and faeces of cats (Zhang et al. 2014) and ferrets (Fehér et al. 2014). CAV, GyV2, and GyV3 species have also been detected in fecal samples from wild birds in Brazil (Duarte et al. 2019). Therefore, only three gyroviruses have been found exclusively in bird species different from chicken: GyV8, GyV10, and ASPaGyV. GyV8 was isolated from the spleen and uropygial gland of a Northern Fulmar (*Fulmarus glacialis*) with neurological clinical

signs (Li et al. 2015). GyV10 was present in several Crested Screamers (*Chauna torquata*), some with neurologic disease and clinical and pathological features resembling CAV infection, as well as from apparently healthy birds (Goldberg et al. 2018). The latest gyrovirus discovered, ASPaGyV, was detected in a metagenomic analysis from cloacal samples in Ashy Strom Petrels (*Oceanodroma homochroa*) (Waits et al. 2018).

The pathogenic significance of these genetically novel avian and human gyroviruses is currently uncertain (Rosario et al. 2017; Yao et al. 2017), but similarities between avian and human sequences suggest that some of them may have zoonotic potential (Yao et al. 2017). This is highlighted by the observation of possible recombinants between strains from different origins (Zhang et al. 2014). In this study, we describe using unbiased deep sequencing the detection and genetic characterization of a novel gyrovirus species, designated as GyV11, from Neotropical birds in French Guiana.

MATERIAL AND METHODS

Sample collection

The study was carried out at the Pararé and Inselberg field stations in the Nouragues Ecological Research Station, managed by France's Centre National de la Recherche Scientifique (CNRS) and located in the Nouragues Natural Reserve in French Guiana. Birds of the understory were mist-netted in January 2016 (rainy season) and in October–November 2016 (dry season). From each bird, a small amount of blood, which was subsequently used for several preparations, was collected. After making a blood smear, the remaining volume was divided into two different tubes, one containing ethanol and the other RNAlater (Life Technologies, Vilnius, Lithuania). In addition, oral and cloacal swabs were taken from each individual. Cloacal swabs were not taken when the size of the bird was too small, mainly in hummingbirds. Both oral and cloacal samples were placed in 800 µL of universal viral transport medium (VTM) (Becton Dickinson, Sparks Glencoe, MD, USA). Blood samples in ethanol and RNAlater were stored at −20 °C, while oral and cloacal swabs were kept at −80 °C. A total of 406 birds from 72 species and 24 different families were sampled.

Sample processing and Next Generation Sequencing

Cloacal samples with abundant fecal matter from 50 individual birds, including the highest bird species diversity possible (31 species of 12 families), were grouped in five pools of 10 individuals. The samples were subjected and processed for NGS, as described elsewhere (Fernández-Correa et al. 2019). Shortly after, the fecal samples were vortexed and the swabs squeezed to release epithelial cells. The pellets obtained were resuspended in 250 µL PBS and subjected to two freeze–thaw cycles at −80 °C to maximize the release of viral particles. Afterwards, in order to reduce the volume of bacteria and other contaminants, the samples were filtered through 0.45 µm pore-sized columns. Filtrate of each sample (50 µL) was mixed in five pools of 10 samples each. Each filtrate was treated with a mixture of nucleases (Turbo DNase, Ambion, Carlsbad, CA, USA; Baseline-ZERO, Epicenter, Madison, WI, USA; Benzonase, Novagen, San Diego, CA, USA; RNAse One, Promega, Fitchburg, WI, USA) to digest unprotected nucleic acids including host DNA/RNA. Finally, viral RNA/DNA was extracted with the MagMAX Viral RNA Isolation Kit (Thermo Fisher, Vilnius, Lithuania) according to the manufacturer's instructions. After random RT-PCR amplification, the extracted viral nucleic acids were subjected to library preparation using a QIAseq FX DNA Library Kit (Qiagen, Germany) and sequenced using (2 × 300 bp paired-end) MiSeq Reagent Kits v3 (Illumina, San Diego, CA, USA) on a MiSeq platform. Raw reads were first trimmed and filtered to remove polyclonal and low-quality reads (< 50 bases long) using CLC workbench (Qiagen). The remaining filtered raw reads were de novo assembled separately using Trinity v2.6. (Grabherr et al. 2011) and CLC workbench and compared with a non-redundant and viral proteome database using BLASTx with a cut-off E-value of 0.001. The virus-like contigs and singlets were further compared to all protein sequences in non-redundant protein databases with a default E-value cutoff of 0.001. The viral metagenomic and metatranscriptomic output was visualized and analyzed in MEGAN (Huson et al. 2016).

Prevalence of the novel GyV11 in cloacal, oral, and blood samples

We tested for the presence of GyV11 in 356 additional cloacal samples that were not subjected to deep sequencing, and also in oral and blood samples from the 50 individuals whose cloacal samples were subjected to deep sequencing. Oral and blood samples of those individuals that tested positive in the cloacal screening were also analyzed. Cloacal

samples were analyzed in pools, while oral and blood samples were analyzed individually.

The 50 selected individuals pooled for deep sequencing were screened by PCR for the presence of different viruses in their cloacal, oral, and blood samples. In order to detect the presence of GyV11 in our samples, we designed two sets of specific primer pairs: Gyr1168F (5′–GCATCCTGGCTTCACTCCTCACA–3′)/Gyr1168R (5′–CGGCGTCCCCTGCTGCTA–3′) and Gyr266F (5′–CCGCCGCGCTGGGAGGTA–3′)/Gyr266R (5′–TTGGCGTCTGAAGCGTTGAT–3′), which amplify genome fragments of 396 bp and 361 bp, respectively. In total, 406 birds were tested for the presence of the novel gyrovirus (Table S3.1).

All DNA amplifications were performed in 25 µL of reaction mixture, containing 0.625 U of AmpliTaq DNA polymerase (Applied Biosystems, Vilnius Lithuania), PCR Buffer II (Applied Biosystems) supplemented to a final 2.5 mM $MgCl_2$, 0.8 mM of each desoxynucleotide diphosphate (Fermentas, Vilnius, Lithuania), 10 pmol/µL of each primer, and 5 µL of extracted RNA/DNA. In the case of cloacal samples, we added 5 µL of 10^{-1} dilution from the RNA/DNA extraction owing to the presence of PCR inhibitors in faeces. The reactions comprised a first denaturation step at 94 °C for 5 min, 45 cycles (denaturation at 94 °C for 30 s, annealing at 57 °C for 30 s for both primer pairs, and polymerization at 72 °C for 30 s), and a final extension step at 72 °C for 5 min. RT-PCR was carried out using Verso 1-Step RT-PCR Kit (Thermo Fisher) following the manufacturer's protocol.

Genomic Characterization and Phylogeny

The putative ORFs in the genome of the novel gyrovirus were predicted using ORFfinder and later identified by BLASTX. Pairwise similarities for the complete genome of the gyrovirus, putative ORFs, and protein sequences were calculated with Clustal Omega. To identify the main protein motifs, we applied Motif Scan (MyHits, SIB, Switzerland), cNLS Mapper (Kosugi et al. 2009), and NetNES 1.1 Server (La Cour et al. 2004). Heatmaps for protein pairwise similarities were created with Morpheus (https://software.broadinstitute.org/morpheus). Putative binding sites of the non-translated region (NTR) were identified using TFBIND (Tsunoda and Takagi 1999) and GPMiner (http://gpminer.mbc.nctu.edu.tw), and stem-loop structures were analyzed using Mfold (Zuker 2003). Tandem repeats were searched using Tandem repeats Finder

(Benson 1999). We plotted the genome organization of the novel gyrovirus employing the DNAPlotter software (Carver et al. 2009).

The genome obtained for GyV11 was compared with all complete gyrovirus VP1, VP2, and VP3 protein sequences publicly available. Sequences were aligned using the MAFFT algorithm and then visually inspected in Geneious v9.1.4. To investigate the phylogenetic relationship of the GyV11 with other gyroviruses available in databases, phylogenetic reconstructions were performed using Bayesian inference and a Monte Carlo Markov chain (MCMC) sampling method as implemented in BEAST v.1.8.3 (Drummond and Rambaut 2007), and in parallel using maximum likelihood inference in PhyML v3.1 (Guindon et al. 2010) based on VP1, VP2, and VP3 protein sequences. Analyses were performed under the best fit amino acid substitution model identified as the WAG+Γ+I for VP1 and VP2, and JTT+Γ for putative VP3 protein data set using jModelTest 2 (Darriba et al. 2012). Monte Carlo Markov chains (MCMC) were run for 107 generations, sampling every 1000 trees. Traces were inspected for convergence with Tracer 1.5 (https://github.com/beast-dev/tracer/releases). Maximum clade credibility (MCC) trees for each ORF were generated and visualized using FigTree v1.4.1 (http://tree.bio.ed.ac.uk/software/figtree/).

RESULTS

Molecular screening of GyV11

An incomplete genome of a novel gyrovirus, named GyV11 (1855-b fragment) was assembled from one of the five cloacal pools subjected to deep sequencing. In that pool, only one individual, a Ferruginous-Backed Antbird (*Myrmoderus ferrugineus*), tested positive, so this individual's cloacal sample was subjected to a second metagenomic analysis in order to fully sequence the complete genome. The same Ferrugineous-Backed Antbird also tested positive for GyV11 in the blood sample, but not in the oral sample. In the 356 additional cloacal samples that were not subjected to deep sequencing, the identical GyV11 sequence was also found in the oral sample from a rufous-rumped foliage-gleaner (*Philydor erythrocercum*) and in the cloacal sample of a white-plumed antbird (*Pithys albifrons*). All of the positive results were confirmed by Sanger sequencing. Hence, 3 out of 406 birds (0.73%) were positive for GyV11, all three of them

sampled during the dry season. Attempts to isolate GyV11 were not possible owing to very limited amount of each sample available.

Genomic characterization

The genomic sequence of this virus differs considerably from all other gyroviruses previously described (shared amino acid similarity of 25% to 29% for VP1, 13% to 21% for VP2, and 10% to 17% for VP3; Figure 1).

The genome of GyV11 is a circular DNA molecule, 2138 bases long, which exhibits the three characteristic ORFs of gyroviruses, VP1, VP2, and a putative VP3 partially overlapping each other (Figure 2). VP1, VP2, and VP3 amino acid sequence comparisons revealed that, while VP1 and VP2 are somewhat similar to those previously described, VP3 shows little similarity to other gyrovirus VP3. GyV11 VP1 encodes a putative viral capsid protein with 397 aa. This VP1 shows a typical N-terminal region rich in basic amino acids with high arginine content typical of capsid proteins. The VP2 protein is 206 aa long and contains the $WX_7HX_3CXCX_5H$ and CX_5R motifs (at positions 30 and 42, respectively) characteristic of protein tyrosine phosphatases in gyroviruses and other anelloviruses. The VP3 ORF encodes a protein of 134 aa, being the longest VP3 among gyroviruses described so far. The putative VP3 in GyV11 did not contain apoptin conserved protein domains. However, we found three motifs that are common in other gyrovirus VP3: a proline-rich region in the N-terminal region, a bipartite nuclear localization sequence (RKRPKPGTEAWLLQRKKE; position 90), and a putative nuclear export sequence (LECNEIL; position 117), with the latter two located in the C-terminal region.

The non-translated region (NTR) of the GyV11 is 435 bp in length (Figure 2) and shared an unusual constellation. Up to 11 putative transcription factor binding sites were identified: four for MZF1 (myeloid zinc finger 1) (positions 82, 136, 174, and 348), three for Sp1 (zinc finger transcription factor) (positions 58, 286, and 345), three for AP2 (positions 6, 51, and 71), and one for GATA2 (position 287). However, no canonical TATA boxes were identified. Two 4-b motifs (CGGG and GGGC) previously described in torque teno viruses were found 11 times along the NTR (positions 55, 81, 148, 283, 353, 364, 373, 376, 383, 391, and 415). Three of these motifs are located in the three GC boxes found (positions 352, 373, and 397) (Figure 2). Four putative stem-loop structures were observed between positions 250 and 370. Furthermore, a polyadenylation signal

(AATAAA; position 24) and two direct repeats (5'-AACCCTAAC-3´) connected by 4 b are located at the 5´ region of the NTR. A GC-rich region is also located in this 5´ region of the NTR, between positions 68 and 88. The promoter nucleotide motif containing a putative estrogen-response element (ACGTCA) detected in several gyroviruses (Gia Phan et al. 2013) that can upregulate transcription was found in the GyV11 NTR. The findings suggest that this virus represents a novel member in the genus *Gyrovirus*, and thus we propose it as a new gyrovirus species or genotype, designated as GyV11 (GenBank accession number MH638372).

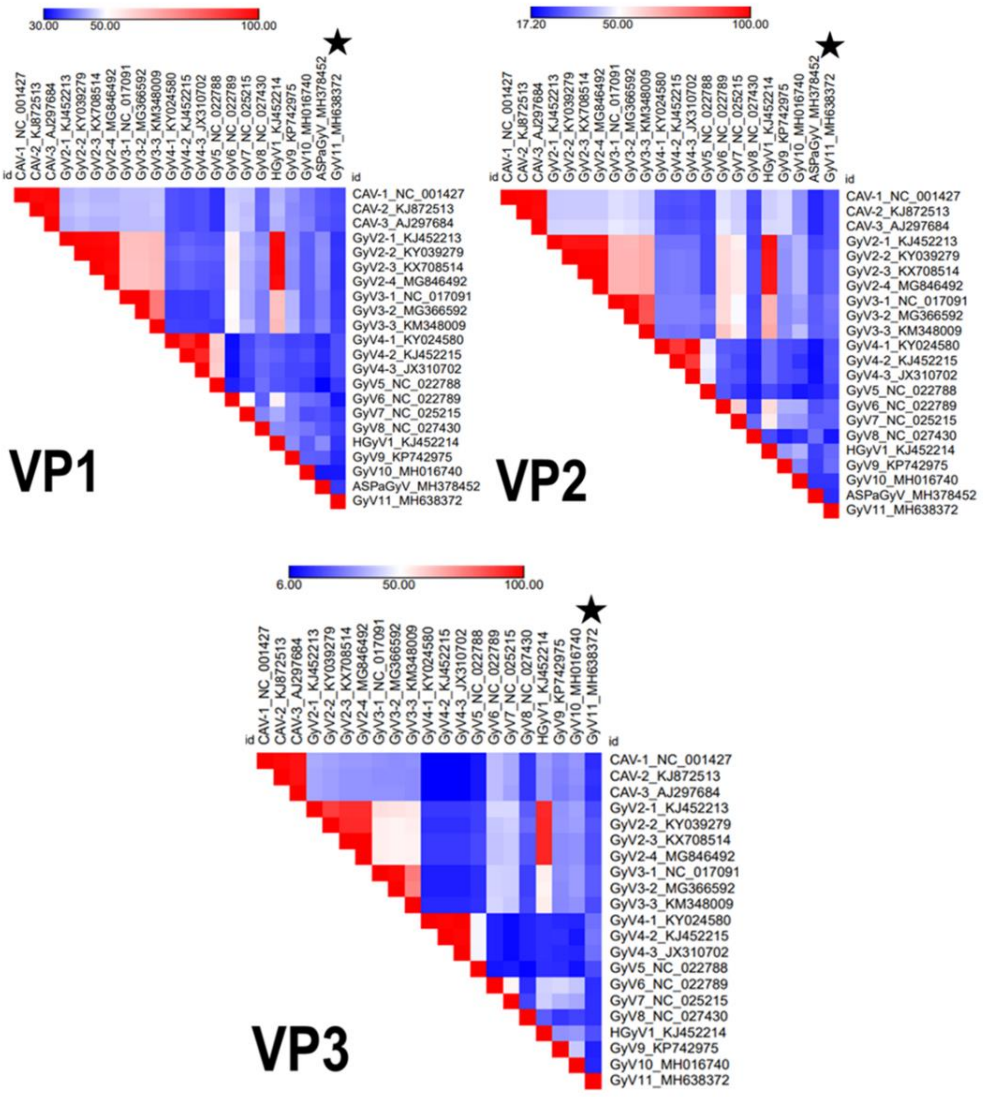


Figure 1. Heatmaps created from pairwise similarities for the three viral proteins (VP1, VP2, and VP3) among all representative gyroviruses. Legend values show in percentage the lowest (dark blue) and the highest (red) similarities in the comparison. CAV: chicken anemia virus; GyV: gyrovirus; HGyV: human gyrovirus.

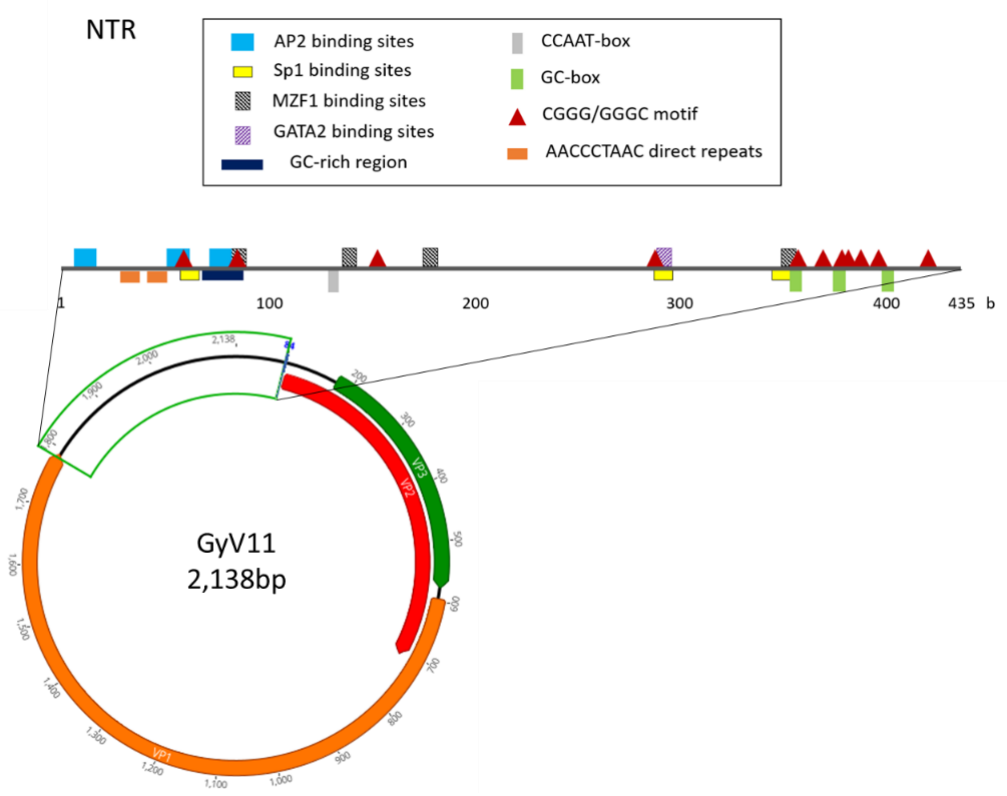


Figure 2. Genomic organization of the gyrovirus GyV11 with the three characteristic open reading frames of the members of the genus Gyrovirus: VP1 (capsid protein), VP2 (phosphatase), and VP3. The non-translated region (NTR) is highlighted and the positions of the putative regulatory motifs are shown. MZF, myeloid zinc finger; Sp1, zinc finger transcription factor.

Phylogenetic analysis

The Bayesian maximum clade credibility (MCC) trees and maximum likelihood phylogeny (not shown) of gyrovirus VP1, VP2, and putative VP3 amino acid sequences revealed two well supported lineages of gyroviruses (Figure 3), corresponding to gyrovirus groups A and B (Gia Phan et al. 2013). The molecular phylogenies showed that GyV11 belongs to the gyrovirus lineage B, with VP1 and VP2 sequences being markedly divergent and forming a basal node with respect to previously reported gyroviruses of that lineage, GyV4 and GyV5, detected in humans, ferrets, and fowl. The VP3 sequence is also most closely related to GyV4 and GyV5, with which it forms a sublclade that is sister to two gyroviruses found in sea birds (GyV8 and ASPaGyV). The most recent common ancestor of the two clades is supported by posterior and bootstrap values (Figure 3). The position of GyV11 in all trees showed a sister relationship and is consistent with the genetic equidistance to all viruses from the lineage B (Figures 1 and 3). In addition,

the phylogenetic analyses also revealed a basal position of all gyroviruses detected in wild birds, likely owing to ancestral origin and longtime evolution.

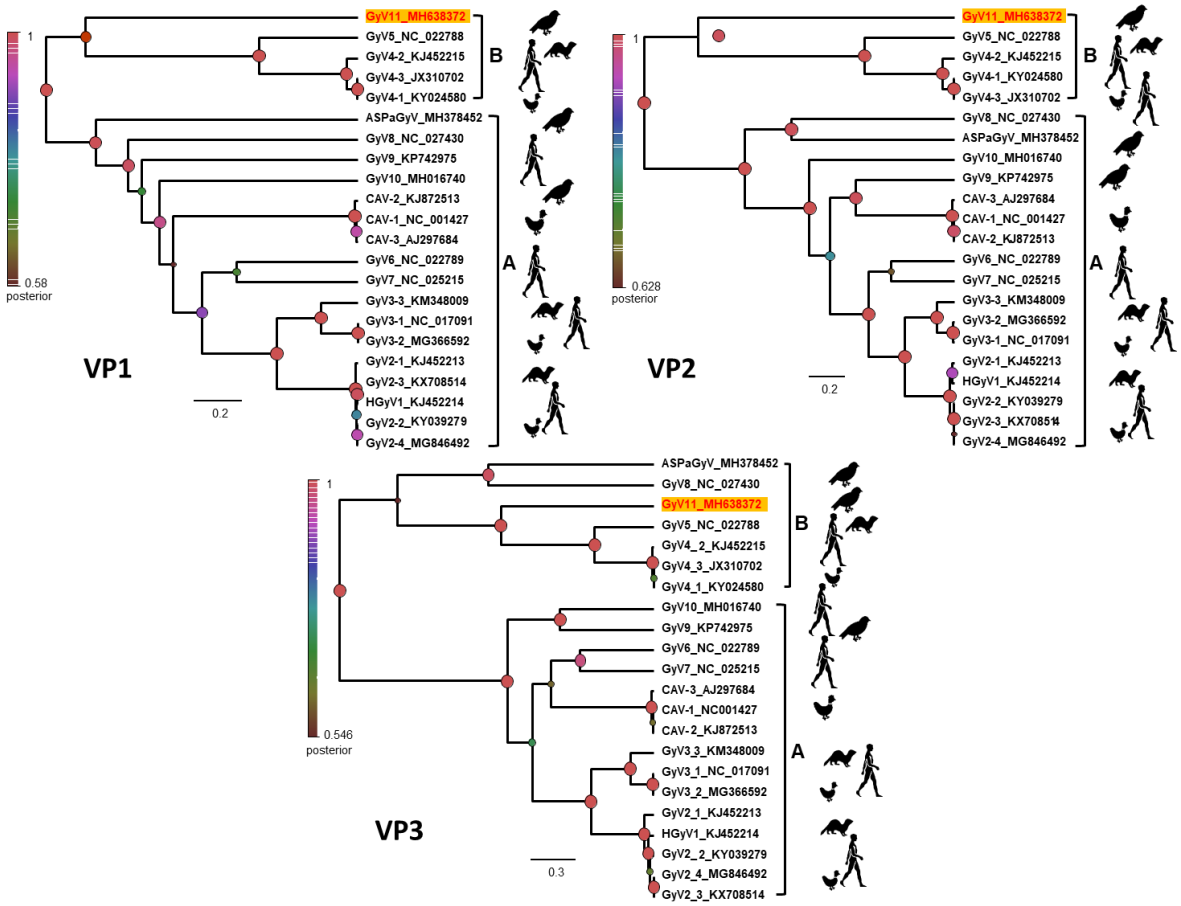


Figure 3. Bayesian maximum clade credibility (MCC) trees representing the phylogenetic reconstructions of the gyroviruses including the GyV11 from this study based on the amino acid sequences of VP1, VP2, and VP3 proteins. Bayesian posterior probabilities are indicated at the nodes. Strain names and GenBank accession numbers for sequences used to construct the trees are indicated on the branches. The scale bar indicates mean number of amino acid substitutions per site.

DISCUSSION

In this study, we have detected and characterized for the first time the presence of a novel gyrovirus, GyV11, in wild populations of birds in a remote area with practically no human influence. Owing to very low similarities with previously known gyroviruses, GyV11 could be designated as a novel gyrovirus species. Moreover, it is also the first time that the same gyrovirus is found in individuals from three different avian species. To date, the study of gyroviruses in birds has been almost exclusively limited to poultry (Schat 2009; Chu et al. 2012a; dos Santos et al. 2012), with GyV8, GyV10, and ASPaGyV being the only ones described exclusively in other avian species (*Fulmarus glacialis*, *Chauna torquata*, and *Oceanodroma homochroa*, respectively) (Li et al. 2015; Goldberg et al. 2018). Both GyV8 and GyV10 were identified in diseased birds, although GyV10 was also present in apparently healthy individuals (Goldberg et al. 2018). GyV8 was found in one northern fulmar showing ataxia and head tilt, but no other fulmar with similar clinical signs analyzed resulted positive for GyV8 (Li et al. 2015). ASPaGyV was discovered in cloacal samples from Ashy Storm Petrels (Waits et al. 2018). The presence of GyV11 was detected in three different birds from the Nouragues Natural Reserve, but only one of them was found to be positive in both cloacal and blood samples, thus supporting the idea of a possible systemic infection. The concurrent presence of CAV in blood and rectal content has been previously shown to occur in chicks inoculated with the virus at one day of age up to seven days after inoculation (Yuasa et al. 1983). The bird likely infected by GyV11 was also found to be positive for astrovirus in the cloacal sample, which may suggest a compromised immune system and a possible co-infection. Co-infections with Marek's disease virus and CAV and with GyV2 and Newcastle Disease Virus have been reported in chickens (Miles et al. 2001; Abolnik and Wandrag 2014). The other two birds from Nouragues resulted positive for GyV11 only in oral or cloacal samples, but not in blood, so the existence of infection cannot be confirmed. However, the presence of this new gyrovirus in oral and cloacal samples from different bird species strengthens the hypothesis that the fecal-oral route is likely the main route of transmission for gyroviruses (Yuasa et al. 1983). Similarly to mammals, the impact of gyroviruses in the bird's health is questionable as they can represent active infections or passive dietary transit (Phan et al. 2015). However, our study shows that GyV11 can be dispersed by the faeces of, at least, two different avian species (*Myrmoderus ferrugineus* and *Pithys albifrons*) and is likely circulating among different bird species within the same community. In fact, in a

closer area such as Brazil, the circulation of a phylogenetically close genotype (GyV4) has been detected in chickens and three other gyroviruses (CAV, GyV2, and GyV3) have been found in the fecal virome of wild birds (Lima et al. 2019; Duarte et al. 2019).

The prevalence of GyV11 in the analyzed Neotropical birds (0.73%) is much lower than the prevalence of other gyroviruses observed in poultry (Roussan and Jordan 2006; dos Santos et al. 2012; Snoeck et al. 2012). A serological survey implemented in Japan showed that anti-CAV antibodies were present in chickens (60.2% seroprevalence) and quails (61.3%), but they were not found in blood samples from wild birds (Farkas et al. 1998). To our knowledge, there are no other prevalence studies of gyroviruses in wild bird populations. However, the proportion of positive birds in our study is similar to that of other avian viruses analyzed in the same Neotropical rainforest (Moens et al. 2018). This could be explained because the conditions in broiler farms—enclosures with a high density of individuals of the same species—likely favor a faster spread of any infection by avian pathogens than in the pristine rainforest studied here. In addition, it has been shown that CAV and GyV2 are present as contaminants in some commercial poultry vaccines (Varela et al. 2014), which could explain seroprevalence values as high as 100% in some chicken flocks for CAV (Roussan and Jordan 2006) and the widespread distribution of GyV2 (dos Santos et al. 2012).

Gyroviruses have also been detected in fecal samples from other vertebrates such as ferrets, cats, and humans (Chu et al. 2012a; Gia Phan et al. 2013; Fehér et al. 2014; Zhang et al. 2014). Prevalence studies carried out in human stools showed values more similar to those observed in ours (0.56 and 1.67% in children with diarrhea) (Gia Phan et al. 2013), although some others show higher prevalences (13.9%–18.9% in patients with diarrhea) (Chu et al. 2012a). However, the majority of these viruses showed high similarity to gyroviruses previously described in chickens, so the presence of these viruses in the feces is likely owing to consumption of infected chicken meat. This hypothesis is reinforced by two studies carried out to find gyroviruses in human blood in Italy and France (Maggi et al. 2012; Biagini et al. 2013). Recent studies found that HGyV1 was mainly present in immunocompromised patients and, though it was detectable among healthy individuals, its prevalence was low (0.85%). However, the discovery of new gyrovirus genomes may improve the molecular detection of different gyrovirus types in human samples and those prevalences could turn out to be higher.

Regarding the GyV11 genome, we found a high number of putative regulatory motifs, but only one of them (Sp1 binding sites) had been previously described in the NTR of other gyroviruses (Noteborn and Koch 1995). The number of direct repeats in the NTR is also different from that of, for example, CAV, where in most isolates four direct repeats with a 12-b insert can be found (Noteborn and Koch 1995). However, four putative stem-loop structures that could act as transcription enhancers were found in the NTR of GyV11. It is noteworthy that two 4-b motifs (CGGG and GGGC), common in the NTR of other anelloviruses, the torque teno viruses, have been found 11 times along this region, some of them adjacent to the stem-loop motifs, as it happens in torque teno viruses (de Villiers et al. 2011). The presence of five putative binding motifs for two transcription factors related to hematopoiesis such as MZF1 and GATA2 is also remarkable (Hromas et al. 1996; Tsai and Orkin 1997). Whether these differences cause an increase in viral fitness in regards to establishment and spatial diffusion requires further analysis.

GyV11 VP1 and VP2 are similar to those of other identified gyroviruses. Genetic analysis revealed that VP1 has a N-terminal region rich in basic amino acids, a characteristic shared by all members of the family *Anelloviridae* (Rosario et al. 2017), and that VP2 has the CX_5R and the $WX_7HX_3CXCX_5H$ motifs characteristic of protein tyrosine phosphatases, conserved in other gyroviruses and anelloviruses (Peters et al. 2002; Rosario et al. 2017). However, VP3 nucleotide and amino acid sequences showed little similarity to those of previously described gyroviruses. The protein encoded is comprised of 134 aa, and is the longest VP3 among gyroviruses described to date. This ORF overlapping VP2 has been identified in all gyroviruses except GyV4 (Li et al. 2015). In spite of this, the protein it encodes has some important motifs related to the apoptotic function of VP3 in CAV: a proline-rich region in the N-terminal, a nuclear localization sequence, and a putative nuclear export sequence (Noteborn et al. 1994; Danen-Van Oorschot et al. 2003). Taking this into account, we cannot rule out the hypothesis that the VP3 of GyV11 may also induce apoptosis or, at least, play a role in the cell nucleus, regardless of the dissimilarity its sequence has with the other VP3 described.

Phylogenetic and genetic distance indicated that GyV11 represents a new prototype for another gyrovirus species. This study further supports the existence of two distinct gyrovirus lineages, A and B. As previously described (Gia Phan et al. 2013), we also found several genomic characteristics supporting the presence of two distinct gyrovirus groups. The group A VP1s were substantially longer than those of group B. Furthermore,

all members of lineage B genomes contained a small putative VP3 in the same location as the VP3/Apoptin protein present in genome of all members of lineage A, but in different reading frames. Evolutionary relationships seem particularly uncertain when it comes to those gyroviruses found in other avian species apart from chickens, that is, GyV8, GyV9, GyV10, and GyV11. Thus, the discovery of new members of the Gyrovirus genus together with a phylogenetic consensus are needed to clarify evolutionary processes in this group of viruses. Phylogenetic analysis also indicates that most basal gyroviruses in both A and B lineages are found in wild bird species, suggesting that gyroviruses may have originated in wild bird populations before spreading to other organisms.

In conclusion, GyV11 represents a novel member of the genus *Gyrovirus*, which shows little similarity with other previously described members of the genus. It was detected in cloacal, oral, and blood samples from passerines in a remote rainforest in French Guiana. The presence of the GyV11 in faeces and blood could be considered as responsible for a possible immunosuppression status, as has been previously shown for other gyroviruses such as CAV and GyV10. The detection and genetic characterization of a novel gyrovirus (GyV11) from a remote area expands our knowledge about the geographic distribution, host range, ecology, diversity, genomic structure, and evolutionary relationships of gyroviruses.

SUPPLEMENTARY MATERIAL

Chapter 3: A novel and divergent gyrovirus with unusual genomic features detected in wild passerine birds from a remote rainforest in French Guiana

Table S3.1. List of species including the 50 individuals selected for deep sequencing and the 356 additional individuals tested for prevalence analysis.

Deep sequencing	
Species	**Number of individuals**
Glyphorynchus spirurus	8
Gymnopithys rufigularis	2
Hylophylax naevius	2
Lepidothrix serena	2
Leptotila rufaxilla	1
Mionectes macconnelli	2
Philydor erythrocercum	3
Bucco capensis	1
Chloroceryle inda	1
Corapipo gutturalis	1
Corythopis torquatus	1
Pseudopipra pipra	2
Myrmotherula longipennis	1
Pithys albifrons	2
Ramphocelus carbo	2
Cyanoloxia cyanoides	1
Formicarius analis	1
Hylophylax poecilinotus	1
Micrastur ruficollis ssp. *concentricus*	1
Myrmoderus ferrugineus	2
Isleria guttata	1
Ceratopipra erythrocephala	1
Platyrinchus coronatus	1
Platyrinchus saturatus	1
Thamnomanes ardesiacus	2
Tachyphonus surinamus	1
Terenotriccus erythrurus	1
Thamnomanes caesius	1
Epinecrophylla gutturalis	1
Turdus albicollis	1
Xenops minutus	1
Xiphorhynchus pardalotus	1

Prevalence analysis

Species	Number of individuals
Attila spadiceus	1
Automolus ochrolaemus	3
Celeus elegans	1
Cercomacroides tyrannina	1
Chloroceryle inda	2
Conopophaga aurita	2
Corapipo gutturalis	4
Corythopis torquatus	3
Cyanerpes caeruleus	1
Cyanoloxia cyanoides	1
Cyphorhinus arada	1
Deconychura longicauda	1
Pseudopipra pipra	25
Formicarius analis	1
Formicarius colma	2
Galbula albirostris	1
Glyphorynchus spirurus	30
Gymnopithys rufigularis	12
Hylophilus ochraceiceps	3
Hylophylax naevius	5
Hylophylax poecilinotus	22
Hypocnemis cantator	2
Lepidothrix serena	10
Leptotila rufaxilla	2
Lipaugus vociferans	2
Manacus manacus	5
Microbates collaris	2
Microcerculus bambla	1
Microrhopias quixensis	1
Mionectes macconnelli	18
Myiobius barbatus	1
Myrmoderus ferrugineus	1
Myrmornis torquata	3
Myrmotherula axillaris	8
Isleria guttata	5
Epinecrophylla gutturalis	6
Myrmotherula longipennis	12
Myrmotherula menetriesii	2
Onychorhynchus coronatus	1
Sporophila angolensis	1
Percnostola rufifrons	4
Philydor erythrocercum	2
Philydor pyrrohodes	2
Phoenicircus carnifex	1
Ceratopipra erythrocephala	6
Pithys albifrons	53
Platyrinchus coronatus	4

Species	Number of individuals
Platyrinchus saturatus	1
Progne chalybea	1
Ramphocaenus melanurus	2
Ramphocelus carbo	12
Rhynchocyclus olivaceus	1
Saltator maximus	1
Schiffornis turdina	3
Sclerurus caudacutus	3
Sclerurus rufigularis	9
Tachyphonus surinamus	4
Thamnomanes caesius	16
Thamnomanes ardesiacus	1
Thamnophilus murinus	2
Pheugopedius coraya	2
Tolmomyias poliocephalus	1
Turdus albicollis	7
Tyrannus melancholicus	1
Xenops minutus	3
Xiphorhynchus pardalotus	3

Chapter 4: Identifying divergent host roles for avian malaria in a community of Neotropical understory passerine birds

Identifying divergent host roles for avian malaria in a community of Neotropical understory passerine birds

Daniel A. Truchado[1,2], Carolina Remacha[1], Lucía Jiménez-Gallardo[1], Borja Milá[3], Michaël A.J. Moens[4], Laura Benítez[2] and Javier Pérez-Tris[1].

[1] Department of Biodiversity, Ecology and Evolution. Faculty of Biology. Complutense University of Madrid. José Antonio Novais, 12. 28040, Madrid, Spain

[2] Department of Genetics, Physiology and Microbiology. Faculty of Biology. Complutense University of Madrid. José Antonio Novais, 12. 28040, Madrid, Spain

[3] National Museum of Natural Sciences, Spanish Research Council (CSIC), Madrid 28006, Spain

[4] Fundación de Conservación Jocotoco, Lizardo García E9-104 y Andrés Xaura, 170143 Quito, Ecuador

ABSTRACT

Avian malaria is an important disease caused by parasites of the genus *Plasmodium* that affects birds all over the world with the potential to result in great fitness costs. However, information about the epidemiological role that each host species may play in the community is still incipient, especially in some undersampled regions such as the Neotropics. In this study, we studied the prevalence of *Plasmodium* infections in the passerine community of the understory of a remote Neotropical forest in the Nouragues Natural Reserve, French Guiana, using a screening method that detects *Plasmodium* with high sensitivity even in coinfection with other parasites. We observed a community prevalence of 10.8% for avian malaria, a higher value to those from other Neotropical regions. Neverteless, *Plasmodium* infection was unequally distributed among the different bird species. The species *Willisornis poecilinotus, Myrmotherula axillaris, Isleria guttata* and *Hylophylax naevius*, all of them belonging to the ant-followers guild (family Thamnophilidae), significantly concentrated the majority of the infections while only the species *Glyphorhynchus spiriurus* (family Furnariidae) seems to avoid the disease, scoring very low prevalence despite being one of the most abundant birds in the forest. Also, we identified 11 different *Plasmodium* lineages, four of which were new. Therefore, avian malaria did not affect equally to all species in the study area, which could greatly influence the transmission of *Plasmodium* parasites in the ecosystem of Nouragues.

INTRODUCTION

Haemosporidians are important vector-borne parasites of birds. There is a wide variety of lineages mainly belonging to the genera *Plasmodium*, *Haemoproteus* and *Leucocytozoon* infecting a diverse range of bird species in all ecosystems where dipteran vectors are present (Valkiunas 2004; Marzal 2012; Fecchio et al. 2020a; Nourani et al. 2020). Parasites of the genus *Plasmodium* are the most invasive and widespread and cause avian malaria in the strict sense (Valkiūnas et al. 2005). Although negative effects of haemosporidians on their hosts are difficult to study in most wild populations, some authors have shown associated costs mainly affecting survival and fitness of infected birds (Marzal et al. 2005; Martínez-De La Puente et al. 2010; Lachish et al. 2011). Even when the infection becomes low-level and chronic, it has long-term adverse effects for infected individuals mediated by telomere shortening (Asghar et al. 2015). Haemosporidian infection is especially detrimental for domestic birds (Williams 2005; Cellier-Holzem et al. 2010) and also for naïve wild bird populations whose immune system is not prepared to cope with the infection (Atkinson and Samuel 2010). It is in this group of naïve wild birds where haemosporidian infections can represent a threat for the conservation of their unique avifauna. Haemosporidian infections and its consequences also depend on the different host species. Host traits, such as mating system, flocking behavior, diet, immunocompetency or nest type, could determine the probability of being infected (Fecchio et al. 2020b), although there is controversy in some of the results. Moreover, the majority of the studies available have been carried out at a high geographical scale when studying the prevalence of avian malaria at lower scales gives us specific information about the roles that the different species of the community may play in the local transmission of the disease. It has been demonstrated that mosquito species transmitting avian malaria do not have any preference for a particular set of bird species when feeding despite the heterogeneous distribution of *Plasmodium* lineages among the different bird species (Medeiros et al. 2013). Therefore, host compatibility seems to be the key for this uneven distribution of the disease in avian communities. Taking all this into consideration, exhaustive sampling of wild bird communities to determine which species act as key hosts while others may avoid the infection is necessary in order to know the dynamics of the transmission of *Plasmodium* parasites.

Despite the importance of the disease they produce and the relatively high number of avian malaria studies available, not all regions on Earth have been analyzed with a similar

sampling effort. North America and Europe account for a great number of the existing articles, while other regions such as Australia or the tropics remain little studied (Clark et al. 2014). Tropical regions are of special interest as they foster the highest vertebrate species diversity in the world (Sherratt and Wilkinson 2009). Moreover, this increasing diversity gradient towards the tropics seems to be also true for *Plasmodium* lineages and most dipteran species that can act as vectors (Foley et al. 2007; Clark et al. 2014; Oakgrove et al. 2014). Within the tropics, research on avian malaria in the Neotropics is essentially informative as this region includes, also, several of the main hotspots of avian diversity in the world (Myers et al. 2000; Orme et al. 2005). French Guiana is located in one of such hotspots which harbors a unique avifauna composed of more than 700 bird species (Thiollay 2002). However, due to its remote location, pathogen studies in Guianan avifauna are very scarce. Studies conducted in other parts of the Neotropics showed that prevalence of avian haemosporidian infection was relatively low, being *Haemoproteus* and *Plasmodium* the main genera infecting birds (Moens and Pérez-Tris 2016; Fecchio et al. 2020b). Although prevalence is low on the whole in this region, they can reach up to more than 30% in the analyzed populations (Lacorte et al. 2013; Marzal et al. 2015; Gonzalez-Quevedo et al. 2016). Moreover, it seems that not all bird groups are infected with the same intensity and lineage diversity within the Neotropics. For example, hummingbirds in Ecuador showed very high infection intensities of *Haemoproteus witti*, a parasite whose intensity was extremely low in passerines, revealing the role of hummingbirds as key reservoirs of the parasite in the community (Moens et al. 2016). Another study, however, showed that avian haemosporidian infection was heterogeneous among the different bird species analyzed but it was not related to avian ecological traits, but on other factors such as vector abundance (Fecchio et al. 2017). These results evidence that, despite the available research carried out in the Neotropics to date, more effort should be done in order to clarify the complex factors involved in the host-parasite interactions of avian haemosporidians in such a diverse region (Fecchio et al. 2020b).

In this study, we carried out a prevalence analysis of *Plasmodium* infection in an assemblage of wild passerines inhabiting the Nouragues Natural Reserve, in French Guiana. Our objective is to determine divergent host species roles ("key reservoir hosts" vs. "parasite avoiders") of a phylogenetically restricted avian group to understand the dilution/concentration effects that *Plasmodium* parasites may be facing in the analyzed community. Subsequently, we analyzed whether this infection was randomly distributed

among the different bird species and families sampled. These results will help understand the real impact of avian malaria in a pristine, remote tropical area while we will be able to know whether some bird species act like key hosts of this group of parasites in this ecosystem.

MATERIAL AND METHODS

Sample collection

A random sampling of understory passerines was carried out in The Nouragues Natural Reserve, in French Guiana (4º05'N, 52º40'W). The Nouragues Natural Reserve is located in a tropical rainforest in northern South America, where average annual temperature is near 26ºC. The climate is very wet on the whole (annual precipitation > 3,000 mm) with substantially less rainfall during the dry season between August and November. The highest elevation in the Reserve is 430 m above sea level (Grimaldi and Riera 2001).

Birds were mist-netted in Pararé and Inselberg camps, two locations separated 8 Km within Nouragues Natural Reserve, during the rainy (January) and dry seasons (October-November) of 2016. Birds were taken standard morphometric measurements and photographs and were ringed to avoid repeated sampling of the same individuals. Blood samples were taken by venipuncture and the volume of blood extracted depended on the body size of the bird (always < 1% of body mass) (Carpenter and Campbell 1988). Blood samples were preserved in pure ethanol at room temperature in the field and at -20ºC in the laboratory until processing for molecular analysis. Once the samples were taken, birds were released unharmed at the site of capture. A total of 437 individual samples belonging to 62 species distributed in 15 families were analyzed.

Screening of parasites

DNA was extracted from blood samples following a standard ammonium acetate protocol (Green et al. 2012). We measured the concentration of DNA in our samples and they were diluted to equal final DNA concentrations to 25 ng/µl. We tested the quality of the extracted DNA by amplifying bird sexing markers (Fridolfsson and Ellegren 1999). Then, we carried out the multiplex PCR designed by Ciloglu et al. (2019) to estimate the community prevalence of avian malaria. We selected this nested PCR as the best available

screening method in terms of sensitivity, especially in case of multi-genus infections (Ciloglu et al. 2019).

Each PCR reaction was set up in a final volume of 10 µl per tube, including 5 µl of 2× Qiagen Multiplex PCR Master Mix (Qiagen, Hilden, Germany), 0.2 µl of each primer (PMF/R, HMF/R and LMF/R) at 10 µM concentration, 1.8 µl of ddH_2O and 2 µl of DNA template (25 ng/µl). The PCR protocol included an initial step at 95ºC during 15 min to activate the HotStartTaq DNA Polymerase followed by 35 cycles of denaturation at 94ºC for 30 s, annealing at 59ºC for 90 s and extension at 72ºC for 30 s. The final extension took place at 72ºC for 10 min. Each PCR analysis included, at least, one negative control (ddH_2O) and a triple positive control (*Plasmodium*, *Haemoproteus* and *Leucocytozoon*) per 48 samples. To determine the identity of *Plasmodium* lineages, we carried out the nested PCR designed by Hellgren et al. (2004) with the positive individuals in the multiplex PCR. Positive samples in the nested PCR were sequenced to obtain the 479-bp consensus fragment of the *cytb* conventionally used as a DNA barcode to determine lineage identitiy. Parasites with at least one nucleotide substitution in this marker compared to reference sequences were considered as new lineages.

We visualized 4µl of PCR products run for 1h at 90 V in a 2% agarose gel stained with GelRed® 100×. Prevalence of *Plasmodium* parasite infections have been estimated as the number of infected individuals divided by the total at the species and family levels. We followed the avian taxonomy proposed in the IOC World Bird List 10.2 (https://www.worldbirdnames.org/) (Gill F. et al. 2020).

Statistical analysis

To assess whether a species showed significantly higher or lower prevalence than expected by chance, we performed 10,000 random distribution where we randomly shuffled the observed infections among the sampled individuals. After this, for each group (species or family) we calculated the number of times we obtained a prevalence higher or equal as the observed one (i.e., the species was identified as a "key host"), and similarly the opposite, the number of times we got a prevalence lower or equal as the observed one ("parasite avoiders"). Division of these frequencies by 10,000, the expected prevalence assuming that all species had the same probability of infection, give the statistical significance to host roles. The replicates of the random distribution were carried out using R software (R Development Core Team 2013).

RESULTS

We had a small sample size of five or less individuals in 41 species, while only 5 species had more than 20 individuals sampled (*Pithys albifrons, Glyphorynchus spirurus, Pseudopipra pipra, Willisornis poecilinotus* and *Mionectes macconnelli*; Table S4.1). Mean *Plasmodium* community prevalence was 10.8% (47 out of 437 birds) and a total of 23 species resulted positive for *Plasmodium* with at least one individual infected (Table S4.1).

Four species of Thamnophilidae (*Willisornis poecilinotus, Myrmotherula axillaris, Isleria guttata* and *Hylophylax naevius*) had higher prevalence than expected if parasite infections were randomly distributed among individual birds (Figure 1). Only these four avian species accounted for 38.3% of the total infected individuals (18 out of 47 birds). One species, *Glyphorynchus spirurus,* of the Furnariidae family had lower prevalence than expected by chance ($P < 0.009$, 0 infections out of 41 individuals).

At the family level, Thamnophilidae had significantly higher prevalence ($P < 0.003$), while Furnariidae had significantly lower prevalence than expected by chance ($P < 0.001$). The rest of species and families did not differ significantly from the prevalence expected by chance (Figure 1).

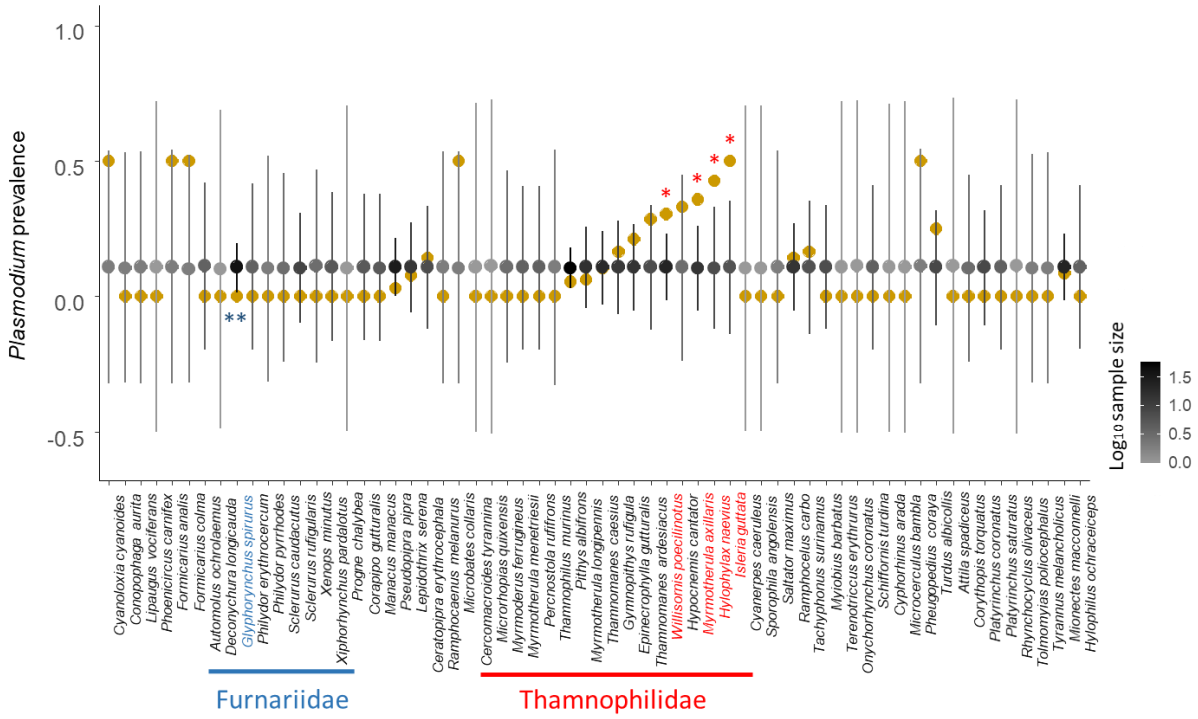


Figure 1. Mean *Plasmodium* prevalence by species in the understory bird community of Nouragues. Ochre circles represent mean observed prevalence while grey-black circles with range lines represent mean ± 95% CI random prevalence obtained after 10,000 random distributions of observed parasites among sampled birds. Sample size of each species is represented by a grey gradient in logarithmic scale, with lighter color for smaller sample sizes. Statistical significance of the difference between observed and expected prevalence for each species is indicated with stars (** $P < 0.01$, * $P < 0.05$). Species and families showing significantly higher prevalences ("key hosts") are highlighted in red while species and families with significantly lower prevalences are highlighted in blue.

Many *Plasmodium* infections could not be amplified with the nested PCR, and we could only obtain sequences from 17 individuals. We identified 11 different lineages, four of which were new (Table 1). One lineage involved in a multiple infection with *Haemoproteus* need further assessment to be classified as new. THACAE08 was the most frequent lineage (n = 5) infecting five different host species.

We obtained several *Haemoproteus* infections but no *Leucocytozoon* infections in the same community. However, we had important diagnostic problems with non-specific bands of the same size as *Haemoproteus* amplification products in some of the samples (see General methods) so we decided not to include *Haemoproteus* results in the analysis.

Table 1. Lineages of *Plasmodium* detected in this study and the number of individuals of each host species infected by them. The novel lineages are marked with a star. The lineage named "P_NEW?" was involved in a multiple infection with *Haemoproteus* and needs further assessment to be clearly classified as new.

Family	Host species	P_THACAE08	P_HYLNAE02*	P_WILPOE22*	P_CORPIL02	P_CYCYA01	P_MYRAXI09	P_TURALB06*	P_GYMRUF01*	P_New?*	P_THACAE08	P_THASAT10
Pipridae	*Ceratopipra erythrocephala*									1		
Polioptilidae	*Microbates collaris*					1						
Thamnophilidae	*Gymnopithys rufigula*	1							1			
Thamnophilidae	*Hylophylax naevius*	1	1									
Thamnophilidae	*Willisornis poecilinotus*			3								1
Thamnophilidae	*Hypocnemis cantator*	1										
Thamnophilidae	*Myrmotherula axillaris*		1				1					
Thamnophilidae	*Isleria guttata*		1								1	
Thamnophilidae	*Thamnomanes ardesiacus*	1										
Thamnophilidae	*Thamnomanes caesius*	1										
Thraupidae	*Ramphocelus carbo*				1							
Turdidae	*Turdus albicollis*							1				
	Total infected	5	3	3	1	1	1	1	1	1	1	1

DISCUSSION

In this study we describe the distribution of avian malaria prevalence among understory passerine species of a Neotropical rainforest bird community in French Guiana, revealing that not all bird species are equally affected by this disease. A small group of ant-followers (family Thamnophilidae) concentrated the majority of the infections while one species of the family Fuirnariidae appeared to avoid the disease.

The prevalence of *Plasmodium* infection we obtained (10.8 %) in the analyzed community of passerines was higher than the average *Plasmodium* infection in the

Neotropics (near 2%). The prevalence of avian malaria in the Neotropics seems to be lower at higher altitudes than our sampling area (Harrigan et al. 2014; Gonzalez-Quevedo et al. 2016) although the prevalence of haemosporidian parasites in general is usually low in this region (close to 10%) compared to other regions of the world (Bennett and Borrero 1976; Valkiunas et al. 2003; Fecchio et al. 2007, 2020b; Moens and Pérez-Tris 2016). However, the prevalence of *Plasmodium* infection in birds from Guyana, in a tropical lowland relatively closer to our study area, reached up to 24.6% of the analyzed birds (Durrant et al. 2006).

Interestingly, *Plasmodium* infections were not equally distributed among the species of passerines analyzed, revealing that some species act as "key reservoir hosts" by showing higher prevalences than expected by chance and some others act as "parasite avoiders" as they present lower prevalences than expected. There were 23 species confirmed to be *Plasmodium* hosts, four of which (*Isleria guttata*, *Myrmotherula axillaris*, *Hylophylax naevius* and *Willisornis poecilinotus*) could be classified as "key reservoir hosts" as their positive individuals represented 38.3% of all *Plasmodium* infections. Contrarily, it is also remarkable that a bird species seems to escape *Plasmodium* infections: *Glyphorynchus spirurus* with 41 analyzed individuals, all of them negative. This species is among the most abundant in the community of passerines of Nouragues (the second most abundant in our sample), which would imply that *Plasmodium* parasites would face a dilution effect if vectors carry them to these apparently incompetent, yet highly abundant hosts. Dilution effect hypothesis supports the idea that diverse communities prevent the spread of diseases by several mechanisms, such as interfering pathogen transmission (Civitello et al. 2015). If mosquitoes transmitting avian malaria bite birds randomly (Medeiros et al. 2013), the abundance of the species *Glyphorynchus spirurus*, a "parasite avoider", could be the key for the regulation of the transmission of *Plasmodium* infection in the community.

It is also noteworthy that the only avian family showing higher prevalence than expected by chance is the family Thamnophilidae, also known as ant-followers. Peters (2010) reported that African ant-following birds showed higher prevalences of *Plasmodium* infections (as well as other blood parasites) than other bird species. He also gave some possible explanations to that: from vectors tracking ant swarms to an increase in the rate of parasite transmission when ant-follower birds gather to feed or even the exposure to a greater diversity of parasites due to larger home ranges of ant-follower birds. Our study

in the Neotropics partially supports the hypothesis that ant-following birds may be more prone to become infected by *Plasmodium* than other birds with different foraging behavior. However, being an ant-follower is not sufficient to be a good reservoir in the analyzed community of passerines. One proof of this is the absence or minimum prevalence of infection observed in three well-sampled ant-following species (*Myrmotherula longipennis*, *Epinecrophylla gutturalis* and *Pithys albifrons*), some of which are congeneric to species identified as key reservoirs.

In conclusion, *Plasmodium* infection distribution among understory passerine species in Nouragues is not random. Some species (various but not all ant-followers) concentrate a great part of the infections, while others seem to avoid the disease, such as we could determine for *Glyphorynchus spirurus*. This woodcreeper dwells in the lowest level of the forest, where ant-followers most typically forage, meaning that the difference is unlikely driven by habitat choice in relation to vector distribution. Further research is needed in order to clarify the transmission dynamics in this and other local bird communities, a necessary first step to fully understanding the patterns of variation in haemosporidian prevalence across scales.

SUPPLEMENTARY MATERIAL

Chapter 4: Identifying divergent host roles for avian malaria in a community of Neotropical understory passerine birds

Table S4.1. Species of passerines analyzed in this study and the number of them infected by *Plasmodium* parasites

Bird species	Family	n	infected
Cyanoloxia cyanoides	Cardinalidae	2	1
Conopophaga aurita	Conopophagidae	2	0
Lipaugus vociferans	Cotingidae	2	0
Phoenicircus carnifex	Cotingidae	1	0
Formicarius analis	Formicariidae	2	1
Formicarius colma	Formicariidae	2	1
Glyphorynchus spirurus	Furnariidae	41	0
Sclerurus rufigularis	Furnariidae	9	0
Xiphorhynchus pardalotus	Furnariidae	5	0
Automolus ochrolaemus	Furnariidae	4	0
Philydor erythrocercum	Furnariidae	4	0
Sclerurus caudacutus	Furnariidae	3	0
Xenops minutus	Furnariidae	3	0
Philydor pyrrhodes	Furnariidae	2	0
Deconychura longicauda	Furnariidae	1	0
Progne chalybea	Hirundinidae	1	0
Pseudopipra pipra	Pipridae	30	1
Lepidothrix serena	Pipridae	13	1
Ceratopipra erythrocephala	Pipridae	7	1
Corapipo gutturalis	Pipridae	5	0
Manacus manacus	Pipridae	5	0
Microbates collaris	Polioptilidae	2	1
Ramphocaenus melanurus	Polioptilidae	2	0
Pithys albifrons	Thamnophilidae	56	3
Willisornis poecilinotus	Thamnophilidae	23	7
Thamnomanes caesius	Thamnophilidae	19	2
Myrmotherula longipennis	Thamnophilidae	16	1
Myrmotherula axillaris	Thamnophilidae	14	5
Epinecrophylla gutturalis	Thamnophilidae	14	3

Bird species	Family	n	infected
Gymnopithys rufigula	Thamnophilidae	12	2
Hylophylax naevius	Thamnophilidae	7	3
Thamnomanes ardesiacus	Thamnophilidae	7	2
Isleria guttata	Thamnophilidae	6	3
Myrmotherula menetriesii	Thamnophilidae	4	0
Percnostola rufifrons	Thamnophilidae	4	0
Hypocnemis cantator	Thamnophilidae	3	1
Myrmoderus ferrugineus	Thamnophilidae	3	0
Thamnophilus murinus	Thamnophilidae	2	0
Cercomacroides tyrannina	Thamnophilidae	1	0
Microrhopias quixensis	Thamnophilidae	1	0
Ramphocelus carbo	Thraupidae	14	2
Tachyphonus surinamus	Thraupidae	6	1
Saltator maximus	Thraupidae	2	0
Cyanerpes caeruleus	Thraupidae	1	0
Sporophila angolensis	Thraupidae	1	0
Myiobius barbatus	Tityridae	7	0
Schiffornis turdina	Tityridae	4	0
Onychorhynchus coronatus	Tityridae	1	0
Terenotriccus erythrurus	Tityridae	1	0
Pheugopedius coraya	Troglodytidae	2	1
Cyphorhinus arada	Troglodytidae	1	0
Microcerculus bambla	Troglodytidae	1	0
Turdus albicollis	Turdidae	8	2
Mionectes macconnelli	Tyrannidae	23	2
Platyrinchus coronatus	Tyrannidae	8	0
Platyrinchus saturatus	Tyrannidae	4	0
Corythopis torquatus	Tyrannidae	3	0
Tolmomyias poliocephalus	Tyrannidae	2	0
Tyrannus melancholicus	Tyrannidae	2	0
Attila spadiceus	Tyrannidae	1	0
Rhynchocyclus olivaceus	Tyrannidae	1	0
Hylophilus ochraceiceps	Vireonidae	4	0

GENERAL DISCUSSION

In this thesis, we shed some light on the cloacal virome and the blood parasite community of wild birds from a remote region in a context where the search for novel pathogens in wildlife is one of the main objectives of global human health. Using a discovery-driven approach, we utilized NGS to characterize the cloacal virome of wild bird populations dominated by passerines inhabiting a pristine, remote tropical rainforest in French Guiana. Moreover, we described the cloacal virome of another wild bird community where passerines prevail in a more studied geographic area (in central Spain) in order to study the viorme of passerines. Also, we compared the similarity of the contigs obtained from Nouragues and La Herrería to test whether wildlife viruses in remote regions are more divergent than wildlife viruses circulating in a more studied area. On the other hand, we carried out a prevalence analysis of avian haemosporidian parasites in the Nouragues avian community to identify the genera infecting birds in this area and to know the scope of the infection in this unique ecosystem.

Comparative composition of Nouragues/La Herrería cloacal viromes

We focused our study on cloacal samples and oral samples were only tested for the presence of those viruses detected in cloaca that could be transmitted by the fecal-oral route. Regarding cloacal virome, the great majority of the contigs obtained from the deep sequencing analysis showed no similarity to reference sequences in public databases, something common in this type of studies, where up to 90% of viral reads can be considered as "viral dark matter" (Krishnamurthy and Wang 2017) as they do not align to any known viral sequence, especially in the case of highly divergent viruses and short fragments. This high number of unclassified contigs can be also the result of analyzing the virome of non-traditional hosts. Research on virology has been traditionally biased towards human and other mammalian viruses (Huson et al. 2007; Krishnamurthy and Wang 2017), so the probability of finding closer reference sequences when working with wild bird viruses is very low, thus highlighting the need for searching in more non-canonical hosts in order to expand the available knowledge on animal viruses.

When comparing similarity parameters (bit score and identitity) of the Blastn search with the contigs obtained in Nouragues and La Herrería we found that bit scores and identities were significantly lower in Nouragues. This results would imply that viruses in the cloaca of wild birds in Nouragues are genetically less similar to known viruses than the viruses

carried by wild birds living in La Herrería, a more studied ecosystem. However, using Blastx, we obtained the opposite result, suggesting that amino acidic sequences of viruses from Nouragues are more similar to those in databases than amino acidic sequences of viruses from La Herrería. Furthermore, although we found higher diversity of virus families in La Herrería, the singularity of these viruses was lower compared to Nouragues, where all complete or almost complete viral genomes we obtained were very divergent to their closest relatives in the phylogeny. Taking all this into consideration, we can suggest that wild birds in Nouragues carry genetically more singular viruses belonging to fewer virus families than wild birds in La Herrería. More comparative studies are needed in order to clarify if this trend is true also for other remote regions of the planet.

Novel putative avian viruses found in Nouragues/La Herrería

Among all putative avian viruses we sequenced, we focused on the four complete genomes of astroviruses (PasAstV-1-4) and the gyrovirus (GyV11) as they are phylogenetically close to important known avian pathogens involved in intestinal disorders and their presence in different bird species of a remote region is of special interest. The four novel astroviruses, provisionally named PasAst-1-4, are genetically distinct among them. The ORF layout is different in PasAstV-2 and PastAstV-4 compared to PasAstV-1 and PasAstV-3, the latter two being the only ones sharing the same arrangement of ORFs. The phylogenetic analysis of ORF2 amino acid sequences placed PasAstV-1-4 together forming a distinct novel clade within the genus *Avastrovirus* (provisionally named *Avastrovirus 5*). However, *Avastrovirus 5* had very long internal branches compared to other groups and they scored the highest intra-specific genetic distance on average, which may suggest that an unknown diversity of avian astroviruses is yet to be discovered in wild birds of this region.

Community-level astrovirus prevalence in Nouragues showed that passerines play an important role in the circulation of these viruses, as 4.9% of 406 analyzed birds were positive for astroviruses, all of them passerines. This prevalence is higher than the 1.7% reported in a previous analysis carried out in cloacal swabs of wild birds in Cambodia and Hong Kong (Chu et al. 2012b). This difference becomes even higher if we restrict it to passerines, as the only community-level study on astroviruses in passerines showed a prevalence of 0.8% (Mendenhall et al. 2015), far behind the 5.1% of passerines positive for astorivrus in Nouragues. Our results show that astroviruses are present in a relatively high prevalence in the avian community of Nouragues, playing passerines an important

role in their circulation. Moreover, the four novel astroviruses we obtained expand our knowledge about the phylogeny, geographic distribution and ecological relationships of these group of avian viruses and pave the road for further research on viruses of wild birds.

Another putative avian pathogen circulating among wild birds in Nouragues is Gyrovirus 11 (GyV11). It was detected in cloacal, oral and blood samples of three passerine species, which expands the known host range of gyroviruses and strengthens the hypothesis that the fecal-oral route is likely the main route of transmission for this group of viruses (Yuasa et al. 1983). Also, the presence of GyV11 in the blood of one ferruginous-backed antbird supports the idea of a possible systemic infection. It is interesting that this individual was also positive for astrovirus infection, which would suggest a compromised immune system and a possible co-infection. Co-infections with other gyroviruses have been previously reported in chickens infected simultaneously with Chicken anemia virus (CAV) and Marek's disease virus and GyV2 and Newcastle disease virus (Miles et al. 2001; Abolnik and Wandrag 2014; Rivero and Gandon 2018). Community-level prevalence of GyV11 is 0.73%, much lower than the prevalences of other gyroviruses in poultry studies (Roussan and Jordan 2006; dos Santos et al. 2012; Snoeck et al. 2012). However, this prevalence is more similar to that of other avian viruses analyzed in the same Neotropical rainforest (Moens et al. 2018). This could be explained by two factors. Firstly, the conditions in poultry enclosures, where chickens live in large groups of individuals in close contact with each other, favor the rapid spread of any viral disease in comparison with the pristine Nouragues rainforest. Secondly, CAV and GyV2, the most common gyroviruses in chickens, have been reported as contaminants in some poultry vaccines (Varela et al. 2014), being likely the cause of the high seroprevalence observed in some chicken flocks for CAV (Roussan and Jordan 2006) or the widespread distribution of GyV2 (dos Santos et al. 2012).

Some parts of the GyV11 genome show little similarity with other previously described members of the genus *Gyrovirus*. For example, its NTR presents several putative binding motifs for regulatory proteins from which only one (Sp1 binding motif) had been previously reported in other gyroviruses (Noteborn and Koch 1995). Also, two 4-b motifs (CGGG and GGGC) typical in other anelloviruses (de Villiers et al. 2011) but not in gyroviruses were found 11 times along this region. Furthermore, although VP1 and VP2 of GyV11 are similar to VP1 and VP2 of other gyroviruses, VP3 amino acid sequence

shows no similarity to other gyrovirus' VP3. However, it has important motifs in common with the VP3 of CAV that have been related to its apoptotic functions (Noteborn et al. 1994; Danen-Van Oorschot et al. 2003), suggesting that the VP3 of GyV11 may also have this function. The singularity of GyV11 is also supported by genetic distance and phylogenetic analyses, as the former showed that similarity to other gyroviruses' proteins is low and the latter placed GyV11 in a divergent clade within the phylogeny of the genus *Gyrovirus*. However, the phylogenetic relationships of GyV11 and the other gyroviruses found in wild birds was not clear, so further research on this group of viruses in wild bird populations is needed in order to clarify their evolutionary history.

Also in Nouragues, French Guiana picornavirus (FGPV) is another putative important avian pathogen found in the clocal sample of a Rufous-throated Antbird (*Gymnophitys rufigula*). The family *Picornaviridae* is the most diverse within ssRNA (+) viruses, infecting mainly mammals and birds (Zell et al. 2017). However, it seems that picornavirus diversity is far from being completely discovered as several novel picornaviruses have been recently reported (Li et al. 2020; Ao et al. 2020; Scherbatskoy et al. 2020; Dastjerdi et al. 2020) and, among those infecting birds, only five out of 18 picornaviruses have been reported in wild birds (Woo et al. 2010; Boros et al. 2013, 2017; Phan et al. 2013; Pankovics et al. 2018). The Rufous-throated Antbird (*Gymnopithys rufigula*) positive to FGPV was also positive for astrovirus, suggesting a possible coinfection as in the case of the Ferruginous-Backed Antbird (*Myrmoderus ferrugineus*) positive for GyV11 and also for astrovirus. Unfortunately, picornavirus infections are usually asymptomatic so it is difficult to determine the concrete effect of FGPV on the bird.

In La Herrería, we partially sequenced another putative important avian virus: La Herrería adenovirus (LHAdV). Adenoviruses are pathogens of different species of vertebrates, mainly causing infections of the respiratory and digestive systems. They are frequently found in deep sequencing analyses of cloacal samples of birds (Phan et al. 2013; Lima et al. 2017; Zhao et al. 2018) as they can be transmitted by the fecal-oral route. LHAdV was detected in a pool formed only by passerines, and phylogenetic analysis placed it in a sister clade to *Great tit siadenovirus A* and *Raptor siadenovirus A*, as a putative novel species in the genus *Siadenovirus*. *Great tit siadenovirus A* was found in a dead passerine bird showing acute enteritis (Kovács et al. 2010) and *Raptor siadenovirus A* was detected in different dead hawks eagle owls showing diverse pathologies (Zsivanovits et al. 2006).

Thus, LHAdV could be an important pathogen for wild passerines circulating in La Herrería forest that may cause similar clinical conditions as its closer relatives. However, we cannot prove this hypothesis given the nature of our sampling method.

Other highly divergent viruses identified in the cloacal viromes

Among all the novel viruses we sequenced in the cloacal viromes of Nouragues and La Herrería, not all of them could be clearly classified as putative avian viruses. This is the case of the novel French Guiana reovirus (FGRV) we found in Nouragues and the two novel hepe-like viruses, French Guiana hepevirus (FGHEV) and La Herrería hepevirus (LHHEV), we discovered in both locations.

FGRV is a novel reovirus sequenced in the cloacal sample of an insectivorous Neotropical passerine, a Wedge-billed Woodcreeper (*Glyphorynchus spirurus*). Reoviruses are dsRNA viruses infecting a wide variety of hosts from plants to vertebrates and insects. In vertebrates, reovirus infection usually cause gastroenteritis and respiratory diseases. FGRV is a very divergent reovirus whose closest relative is Cimodo virus, a reovirus detected in Africa likely infecting mosquitoes (Hermanns et al. 2014). Both reoviruses show little similarity both with the rest of reoviruses and with one another, but phylogenetic analysis suggest they form a new genus within the subfamily *Spinareovirinae*. As FGRV was detected in the cloaca of an insectivorous bird and its closest relative is a reovirus infecting mosquitoes, we could not clearly classify it as a pathogen of birds or a pathogen of an insect the Wedge-billed Woodcreeper had fed on.

Something similar occurs in the case of FGHEV and LHHEV. Hepeviruses are important zoonotic ssRNA (+) viruses causing hepatitis E and splenomegaly with high mortality rates among vertebrates. FGHEV and LHEHV show some genomic features that diverge from the canonical architecture of the family *Hepeviridae* such as their ORF arrangement or the unusually long genome in the case of FGHEV (Purdy et al. 2017). Phylogenetic analyses placed LHHEV within the hepe-like viruses, a recently described group closely related to hepeviruses whose effect on the host is still unknown (Wu et al. 2018). Furthermore, hepe-like viruses have only been detected in invertebrates (Shi et al. 2016; Wu et al. 2018; Dong et al. 2020) or in the feces of vertebrates feeding on them (Williams et al. 2018; Reuter et al. 2018, 2020), so they seem to be likely viruses of invertebrates. Thus, we suggest that LHHEV is a new member of the hepe-like group, but we ignore whether it is a putative avian pathogen or a virus infecting invertebrates as it was detected

in the cloaca of an insectivorous bird. The same situation happens with FGHEV. However, its phylogenetic position is ambiguous and cannot be classified clearly as a hepe-like virus or a divergent hepevirus.

If FGHEV and LHHEV were viruses infecting invertebrates, we were able to sequence the whole genome of FGHEV and almost the complete genome of LHHEV in a cloacal sample of two passerines. This fact argues in favor that both viruses could have maintained their integrity until the end of the avian digestive tract so birds feeding on invertebrates could act as dispersers of their pathogens. Then, their putative invertebrate host could get infected when they come in contact with avian feces. This viral transmission route has been already described in the gypsy moth (Reilly and Hajek 2012) and could be also the explanation of the high number of contigs belonging to the family *Polycipiviridae* found in the cloacal virome of Nouragues, as the vast majority of the viruses of this family have been detected in invertebrates.

Leaving aside whether FGRV and FGHEV infect birds and/or invertebrates, they are two examples of highly divergent viruses circulating in wild bird populations in a remote forest whose host range, pathogenity and phylogentic history is still unknown. This strengthens the need for further investigation of the virome of wildlife in order to clarify their ecology, epidemiology and closer relatives in the phylogeny.

Regarding avian malaria, we discovered various new lineages of parasites of the genus *Plasmodium* and *Haemoproteus*, which expands the known diversity of these parasites, particularly in the Neotropics. The genus *Leucocytozoon* had very low prevalence (if it was present) in this bird community, as we failed to detect infections despite of using the most sensitive screening method available for these parasites. One advantage of the method used is its improved sensitivity in the detection of mixed infections of parasites of different genera. We found various multiple infections of *Haemoproteus* and *Plasmodium* that would have remained undetected with the most commonly used screening methods, the nested PCR. However, our analysis of non-canonical hosts (in this case a great diversity of birds of families that are rare or absent in temperate areas) revealed various drawbacks of the screening method used. Most worryingly, various species produced bands of the diagnostic length of parasites, a problem which precluded the analysis of *Haemoproteus* infections in our study. But on a positive side, our observation immediately suggested recommendations for the improvement of the method. In particular, the use of genus-specific primers may be compulsory to do a screening of

some species that are prone to produce misleading unspecific amplifications with these primers.

Our results on avian malaria also reveals that not all species in the community play the same role in the transmission of the disease. Some species (four in our study area) concentrate a great part of *Plasmodium* infections ("key reservoir hosts") while some others (only one in our case) seem to avoid the disease ("parasite avoider"). It is remarkable that the four species of "key reservoir hosts" belong to the guild of the ant-following birds (family Thamnophilidae), while the "parasite avoider" (*Glyphorhynchus spirurus*) was the second most abundant in our sample. This is of great importance for transmission dynamics as avian malaria parasites may be affected by the dilution effect in avian communities where one of the most abundant species is also an incompetent host. The dilution effect prevents the spread of diseases in diverse communities due to different mechanisms (Civitello et al. 2015). One of such mechanisms are encounter reduction and transmission reduction (Keesing et al. 2006), two situations that *Plasmodium* parasites might be facing in the passerine community of Nouragues.

CONCLUSIONS

I. Extending the focus to non-traditional hosts is essential in the discovery of novel animal viruses. Our research analyzing the virome of wild bird populations dominated by passerines revealed that those fully-sequenced viruses we obtained were completely unknown and sufficiently different to reference sequences to be considered novel species or even genera. Carrying out discovery-driven approaches in hosts other than poultry and waterfowl is crucial for expanding the knowledge in the field of avian virology.

II. Wildlife of remote regions harbor an unknown viral diversity that is yet to be described. The cloacal virome of wild birds of Nouragues contained viruses genetically more dissimilar to reference sequences in public databases than the cloacal virome of wild birds of La Herrería. Therefore, studying the diversity and ecological relations of viruses in the wild proves instrumental for uncovering the true diversity of viruses that, given favourable conditions, could cause emerging infectious diseases that may threaten humans, domestic animals or wildlife. Also,

we obtain precious information about novel virus-host relationships and virus ecology.

III. Discovery-driven approaches with a specific design may speed up the full description of the virome of wildlife and substantially expand the knowledge in animal virology. Deep sequencing analysis proves a powerful and useful tool to characterize the viral diversity of wild animals in this type of studies.

IV. Our study of wild bird communities of Nouragues shows that Astrovirus prevalence may be locally higher than previous studies had suggested (4.3 % higher than the only previous study in passerines). Moreover, passerines are important hosts of astroviruses in the primary tropical forests in the Guianan shield.

V. The four novel astroviruses described in Neotropical passerines, PasAstV-1-4, contribute to broaden the phylogeny of this family and the host range of astroviruses as this is the first report of this group of viruses in birds of the families Thamnophilidae, Cardinalidae, Conopophagidae, Furnariidae, Tyraniidae and Turdidae.

VI. GyV11 is a novel divergent member of the genus *Gyrovirus* showing genomic features different from those of other gyroviruses described so far. GyV11 is present at a low prevalence in the wild bird communities of Nouragues, is likely transmitted by the fecal-oral route, infects (at least) the species *Myrmoderus ferrugineus* and, according to our observations, is candidate to cause immunosuppression.

VII. Avian malaria in Nouragues does not affect all bird species equally. Four species of the family Thamnophilidae concentrated the majority of the infections while the species *Glyphorhynchus spirurus* (family Furnariidae) seemed to avoid the infection. These differences prove the existence of divergent host species roles in the community, including key reservoirs and less competent hosts.

FUTURE RESEARCH

In this thesis, we have demonstrated that both, understudied hosts and remote regions harbor a yet unknown pathogen diversity that can be properly characterized by discovery-driven approaches. Also, we have observed that not all bird species present the same set of pathogens, but it appears that some avian groups are key for the circulation of certain pathogens. For example, passerines of Nouragues seem to play an important role in the circulation of astroviruses while some species of ant-following birds are the main concentrators of malaria infections. More specifically, some individuals present more sensitivity to coinfections than others from the same species, as in the case of coinfections with two putative avian viruses or coinfections of haemosporidians. In the case of avian malaria, we were able to identify those species that could play an important role in the transmission of the disease either because they concentrated the majority of the infections or because they seemed to avoid the disease. It would be interesting to focus the discovery-driven efforts to those individuals with multiple infections in the case of viruses as well. Thus, we would be able to know if those species/individuals are especially suitable for a wide variety of pathogens, what factors are influencing this patterns or what role those species/individuals play in the dynamics of pathogen transmission within the community.

REFERENCES

Abolnik C, Wandrag DBR (2014) Avian gyrovirus 2 and avirulent Newcastle disease virus coinfection in a chicken flock with neurologic symptoms and high mortalities. Avian Dis 58:90–94. https://doi.org/10.1637/10657-090313-Reg.1

Adlard RD, Miller TL, Smit NJ (2015) The butterfly effect: parasite diversity, environment, and emerging disease in aquatic wildlife. Trends Parasitol 31:160–166. https://doi.org/10.1016/j.pt.2014.11.001

Ao Y, Zhou Y, Li D, Duan Z (2020) A novel picornavirus identified in wild *Macaca mulatta* in China. Arch Virol 165:495–504. https://doi.org/10.1007/s00705-019-04442-3

Asghar M, Hasselquist D, Hansson † B, et al (2015) Hidden costs of infection: chronic malaria accelerates telomere degradation and senescence in wild birds. Science 347:436–438. https://doi.org/10.5061/dryad.d04h0

Atkinson CT, Samuel MD (2010) Avian malaria *Plasmodium relictum* in native Hawaiian forest birds: epizootiology and demographic impacts on àapapane *Himatione sanguinea*. J Avian Biol 41:357–366. https://doi.org/10.1111/j.1600-048X.2009.04915.x

Bahl J, Pham TT, Hill NJ, et al (2016) Ecosystem interactions underlie the spread of avian influenza A viruses with pandemic potential. PLoS Pathog 12:. https://doi.org/10.1371/journal.ppat.1005620

Baird JK (2009) Malaria zoonoses. Travel Med Infect Dis 7:269–277. https://doi.org/10.1016/j.tmaid.2009.06.004

Barrowclough GF, Cracraft J, Klicka J, Zink RM (2016) How many kinds of birds are there and why does it matter? PLoS One 11:e0166307. https://doi.org/10.1371/journal.pone.0166307

Bartlett CM, Bain ODILE (1987) New avian filarioids (Nematoda: Splendidofilariinae): *Dessetfilaria guianensis* gen. n. sp. n., *Andersonfilaria africanus* gen. n., sp. n., and *Splendidofilaria chandenieri* sp. n. Proc Helminthol Soc Wash 54:1–14

Bascompte J (2010) Structure and dynamics of ecological networks. Science 329:765–

766. https://doi.org/10.1126/science.1194255

Bell JA, Weckstein JD, Fecchio A, Tkach V V. (2015) A new real-time PCR protocol for detection of avian haemosporidians. Parasites and Vectors 8:383. https://doi.org/10.1186/s13071-015-0993-0

Bennett GF, Borrero JI (1976) Blood parasites of some birds from Colombia. J Wildl Dis 12:454–458. https://doi.org/10.7589/0090-3558-12.3.454

Bensch S, Hellgren O, Pérez-Tris J (2009) MalAvi: A public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. Mol Ecol Resour 9:1353–1358. https://doi.org/10.1111/j.1755-0998.2009.02692.x

Bensch S, Stjernman M, Hasselquist D, et al (2000) Host specificity in avian blood parasites: a study of *Plasmodium* and *Haemoproteus* mitochondrial DNA amplified from birds. Proc R Soc B Biol Sci 267:1583–1589. https://doi.org/10.1098/rspb.2000.1181

Benskin CMWH, Wilson K, Jones K, Hartley IR (2009) Bacterial pathogens in wild birds: a review of the frequency and effects of infection. Biol Rev 84:349–373. https://doi.org/10.1111/j.1469-185X.2008.00076.x

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580. https://doi.org/10.1093/nar/27.2.573

Biagini P, Bédarida S, Touinssi M, et al (2013) Human gyrovirus in healthy blood donors, France. Emerg. Infect. Dis. 19:1014–1015

Bodewes R (2018) Novel viruses in birds: flying through the roof or is a cage needed? Vet J 233:55–62. https://doi.org/10.1016/j.tvjl.2017.12.023

Boros Á, Kiss T, Kiss O, et al (2013) Genetic characterization of a novel picornavirus distantly related to the marine mammal-infecting aquamaviruses in a long-distance migrant bird species, European roller (*Coracias garrulus*). J Gen Virol 94:2029–2035. https://doi.org/10.1099/vir.0.054676-0

Boros Á, Pankovics P, Mátics R, et al (2017) Genome characterization of a novel megrivirus-related avian picornavirus from a carnivorous wild bird, western marsh harrier (*Circus aeruginosus*). Arch Virol 162:2781–2789.

https://doi.org/10.1007/s00705-017-3403-4

Boujon CL, Koch MC, Wüthrich D, et al (2017) Indication of cross-species transmission of astrovirus associated with encephalitis in sheep and cattle. Emerg Infect Dis 23:1604–1608. https://doi.org/10.3201/eid2309.170168

Breman JG (2001) The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. Am J Trop Med Hyg 64:1–11. https://doi.org/10.4269/ajtmh.2001.64.1

Brown VR, Bevins SN (2017) A review of virulent Newcastle disease viruses in the United States and the role of wild birds in viral persistence and spread. Vet Res 48:1–15. https://doi.org/10.1186/s13567-017-0475-9

Brûlé S, Touroult J (2014) Insects of French Guiana: a baseline for diversity and taxonomic effort. Zookeys 111–130. https://doi.org/10.3897/zookeys.434.7582

Carpenter JW, Campbell TW (1988) Avian hematology and cytology. J Zoo Anim Med 19:244. https://doi.org/10.2307/20094898

Carroll D, Daszak P, Wolfe ND, et al (2018) The Global Virome Project. Science 359:872–874. https://doi.org/10.1126/science.aap7463

Carver T, Thomson N, Bleasby A, et al (2009) DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 25:119–120. https://doi.org/10.1093/bioinformatics/btn578

Cellier-Holzem E, Esparza-Salas R, Garnier S, Sorci G (2010) Effect of repeated exposure to *Plasmodium relictum* (lineage SGS1) on infection dynamics in domestic canaries. Int J Parasitol 40:1447–1453. https://doi.org/10.1016/j.ijpara.2010.04.014

Chan JFW, To KKW, Chen H, Yuen KY (2015) Cross-species transmission and emergence of novel viruses from birds. Curr Opin Virol 10:63–69. https://doi.org/10.1016/j.coviro.2015.01.006

Chu DKW, Poon LLM, Chiu SSS, et al (2012a) Characterization of a novel gyrovirus in human stool and chicken meat. J Clin Virol 55:209–213. https://doi.org/10.1016/j.jcv.2012.07.001

Chu, Leung CYH, Perera HKK, et al (2012b) A novel group of avian astroviruses in wild

aquatic birds. J Virol 86:13772–13778. https://doi.org/10.1128/jvi.02105-12

Ciloglu A, Ellis VA, Bernotienė R, et al (2019) A new one-step multiplex PCR assay for simultaneous detection and identification of avian haemosporidian parasites. Parasitol Res 118:191–201. https://doi.org/10.1007/s00436-018-6153-7

Civitello DJ, Cohen J, Fatima H, et al (2015) Biodiversity inhibits parasites: broad evidence for the dilution effect. Proc Natl Acad Sci U S A 112:8667–8671. https://doi.org/10.1073/pnas.1506279112

Clark NJ, Clegg SM, Lima MR (2014) A review of global diversity in avian haemosporidians (*Plasmodium* and *Haemoproteus*: Haemosporida): New insights from molecular data. Int J Parasitol 44:329–338. https://doi.org/10.1016/j.ijpara.2014.01.004

Cortez V, Meliopoulos VA, Karlsson EA, et al (2017) Astrovirus biology and pathogenesis. Annu Rev Virol 4:327–348. https://doi.org/10.1146/annurev-virology

Cox FE (2010) History of the discovery of the malaria parasites and their vectors. Parasit Vectors 3:1–9. https://doi.org/10.1186/1756-3305-3-5

Cracraft J, Barker FK (2009) Passerine birds (Passeriformes). In: Timetree of Life. Oxford University Press, New York, pp 423–431

Cunningham AA, Daszak P, Wood JLN (2017) One health, emerging infectious diseases and wildlife: two decades of progress? Philos. Trans. R. Soc. B Biol. Sci. 372:20160167

Danen-Van Oorschot AAAM, Zhang YH, Leliveld SR, et al (2003) Importance of nuclear localization of apoptin for tumor-specific induction of apoptosis. J Biol Chem 278:27729–27736. https://doi.org/10.1074/jbc.M303114200

Darriba D, Taboada GL, Doallo R (2012) jModelTest 2: more models, new heuristics and high-performance computing. Nat Methods 9:772. https://doi.org/10.1038/nmeth.2109

Dastjerdi A, Benfield C, Everest D, et al (2020) Novel enteric viruses in fatal enteritis of grey squirrels. J Gen Virol jgv001431. https://doi.org/10.1099/jgv.0.001431

Daszak P, Cunningham AA, Hyatt AD (2000) Emerging infectious diseases of wildlife -

threats to biodiversity and human health. Science 287:443–449. https://doi.org/10.1126/science.287.5452.443

Day JM, Ballard LL, Duke M V., et al (2010) Metagenomic analysis of the turkey gut RNA virus community. Virol J 7:313. https://doi.org/10.1186/1743-422X-7-313

de Souza WM, Fumagalli MJ, Martin MC, et al (2019) Pingu virus: A new picornavirus in penguins from Antarctica. Virus Evol 5:vez047. https://doi.org/10.1093/ve/vez047

de Villiers E-M, Borkosky SS, Kimmel R, et al (2011) The diversity of torque teno viruses: in vitro replication leads to the formation of additional replication-competent subviral molecules. J Virol 85:7284–7295. https://doi.org/10.1128/JVI.02472-10

Delwart E (2012) Animal virus discovery: Improving animal health, understanding zoonoses, and opportunities for vaccine development. Curr Opin Virol 2:344–352. https://doi.org/10.1016/j.coviro.2012.02.012

Devlin JM, Vaz PK, Coppo MJC, Browning GF (2016) Impacts of poultry vaccination on viruses of wild bird. Curr Opin Virol 19:23–29. https://doi.org/10.1016/j.coviro.2016.06.007

Dhama K, Mahendran M, Tomar S (2008) Pathogens transmitted by migratory birds: threat perceptions to poultry health and production. Int J Poult Sci 7:516–525

Domermuth CH, Harris JR, Gross WB, Dubose RT (1979) A naturally occurring infection of chickens with a hemorrhagic enteritis/marble spleen disease type of virus. Avian Dis 23:479–484. https://doi.org/10.2307/1589578

Donato C, Vijaykrishna D (2017) The broad host range and genetic diversity of mammalian and avian astroviruses. Viruses 9:102. https://doi.org/10.3390/v9050102

Dong X, Hu T, Liu Q, et al (2020) A novel hepe-like virus from farmed giant freshwater prawn *Macrobrachium rosenbergii*. Viruses 12:323. https://doi.org/10.3390/v12030323

dos Santos HF, Knak MB, de Castro FL, et al (2012) Variants of the recently discovered avian gyrovirus 2 are detected in Southern Brazil and The Netherlands. Vet

Microbiol 155:230–236. https://doi.org/10.1016/j.vetmic.2011.09.021

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214. https://doi.org/10.1186/1471-2148-7-214

Duarte MA, Silva JMF, Brito CR, et al (2019) Faecal virome analysis of wild animals from Brazil. Viruses 11:803. https://doi.org/10.3390/v11090803

Durrant KL, Beadell JS, Ishtiaq F, et al (2006) Avian hematozoa in South America: a comparison of temperate and tropical zones. Ornithol Monogr 98–111

El Zowalaty ME, Järhult JD (2020) From SARS to COVID-19: a previously unknown SARS- related coronavirus (SARS-CoV-2) of pandemic potential infecting humans – Call for a One Health approach. One Heal 9:100124. https://doi.org/10.1016/j.onehlt.2020.100124

Eltahir YM, Qian K, Jin W, Qin A (2011) Analysis of chicken anemia virus genome: evidence of intersubtype recombination. Virol J 8:512. https://doi.org/10.1186/1743-422X-8-512

Fallon SM, Ricklefs RE, Swanson BL, Bermingham E (2003) Detecting avian malaria: an improved polymerase chain reaction diagnostic. J Parasitol 89:1044–1047. https://doi.org/10.1645/GE-3157

Farkas T, Maeda K, Sugiura H, et al (1998) A serological survey of chickens, Japanese quail, pigeons, ducks and crows for antibodies to chicken anaemia virus (CAV) in Japan. Avian Pathol 27:316–320. https://doi.org/10.1080/03079459808419344

Fawaz M, Vijayakumar P, Mishra A, et al (2016) Duck gut viral metagenome analysis captures snapshot of viral diversity. Gut Pathog 8:30. https://doi.org/10.1186/s13099-016-0113-5

Fecchio A, Bell JA, Bosholn M, et al (2020a) An inverse latitudinal gradient in infection probability and phylogenetic diversity for *Leucocytozoon* blood parasites in New World birds. J Anim Ecol 89:423–435. https://doi.org/10.1111/1365-2656.13117

Fecchio A, Chagas CRF, Bell JA, Kirchgatter K (2020b) Evolutionary ecology, taxonomy, and systematics of avian malaria and related parasites. Acta Trop 204:105364. https://doi.org/10.1016/j.actatropica.2020.105364

Fecchio A, Ellis VA, Bell JA, et al (2017) Avian malaria, ecological host traits and mosquito abundance in southeastern Amazonia. Parasitology 144:1117–1132. https://doi.org/10.1017/S003118201700035X

Fecchio A, Marini MÂ, Martins Braga É (2007) Baixa prevalência de hemoparasitos em aves silvestres no Cerrado do Brasil Central. Neotrop Biol Conserv 2:127–135

Fehér E, Pazár P, Kovács E, et al (2014) Molecular detection and characterization of human gyroviruses identified in the ferret fecal virome. Arch Virol 159:3401–3406. https://doi.org/10.1007/s00705-014-2203-3

Fernández-Correa I, Truchado DA, Gomez-Lucia E, et al (2019) A novel group of avian astroviruses from Neotropical passerine birds broaden the diversity and host range of *Astroviridae*. Sci Rep 9:1–9. https://doi.org/10.1038/s41598-019-45889-3

Ferreira-Junior FC, de Angeli Dutra D, Silveira P, et al (2018) A new pathogen spillover from domestic to wild animals: *Plasmodium juxtanucleare* infects free-living passerines in Brazil. Parasitology 145:1949–1958. https://doi.org/10.1017/S003118201800077X

Fitzgerald SD, Rautenschlein S, Mahsoub HM, et al (2020) Adenovirus infections. In: Diseases of poultry. Springer New York, pp 321–363

Foley DH, Rueda LM, Wilkerson RC (2007) Insight into global mosquito biogeography from country species records. J Med Entomol 44:554–567. https://doi.org/10.1093/jmedent/44.4.554

Font I (2000) Climatología de España y Portugal. Universidad de Salamanca. Salamanca

François S, Pybus OG (2020) Towards an understanding of the avian virome. J Gen Virol jgv001447. https://doi.org/10.1099/jgv.0.001447

Fridolfsson AA, Ellegren H (1999) A simple and universal method for molecular sexing of non-ratite birds. J Avian Biol 30:116–121. https://doi.org/10.2307/3677252

Fu Y, Pan M, Wang X, et al (2009) Complete sequence of a duck astrovirus associated with fatal hepatitis in ducklings. J Gen Virol 90:1104–1108. https://doi.org/10.1099/vir.0.008599-0

Fuentes C, Bosch A, Pintó RM, Guix S (2012) Identification of human astrovirus

genome-linked protein (VPg) essential for viral infectivity. J Virol 86:10070–10078. https://doi.org/10.1128/JVI.00797-12

Fuller T, Bensch S, Müller I, et al (2012) The ecology of emerging infectious diseases in migratory birds: an assessment of the role of climate change and priorities for future research. Ecohealth 9:80–88. https://doi.org/10.1007/s10393-012-0750-1

Gallana M, Ryser-Degiorgis MP, Wahli T, Segner H (2013) Climate change and infectious diseases of wildlife: altered interactions between pathogens, vectors and hosts. Curr Zool 59:427–437. https://doi.org/10.1093/czoolo/59.3.427

Garcia SC, Navarro Lopez R, Morales R, et al (2013) Molecular epidemiology of newcastle disease in Mexico and the potential spillover of viruses from poultry into wild bird species. Appl Environ Microbiol 79:4985–4992. https://doi.org/10.1128/AEM.00993-13

Gia Phan T, Phung Vo N, Sdiri-Loulizi K, et al (2013) Divergent gyroviruses in the feces of Tunisian children. Virology 446:346–348. https://doi.org/10.1016/j.virol.2013.08.020

Gill F., Donsker D, Rasmussen P. (2020) IOC World Bird List (v10.2). https://doi.org/10.14344/IOC.ML.10.2.

Giotis ES, Skinner MA (2019) Spotlight on avian pathology: fowlpox virus. Avian Pathol 48:87–90. https://doi.org/10.1080/03079457.2018.1554893

Goldberg TL, Clyde VL, Gendron-Fitzpatrick A, et al (2018) Severe neurologic disease and chick mortality in crested screamers (*Chauna torquata*) infected with a novel Gyrovirus. Virology 520:111–115. https://doi.org/10.1016/j.virol.2018.05.014

Gonzalez-Quevedo C, Pabón A, Rivera-Gutierrez HF (2016) Prevalence of haemosporidians in a Neotropical endemic bird area. Avian Conserv Ecol 11:. https://doi.org/10.5751/ace-00834-110107

Goodfellow I (2011) The genome-linked protein VPg of vertebrate viruses - A multifaceted protein. Curr Opin Virol 1:355–362. https://doi.org/10.1016/j.coviro.2011.09.003

Grabherr MG, Haas BJ, Yassour M, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644.

https://doi.org/10.1038/nbt.1883

Gray GC, Kayali G (2009) Facing pandemic influenza threats: the importance of including poultry and swine workers in preparedness plans. Poult Sci 88:880–884. https://doi.org/10.3382/ps.2008-00335

Green MR, Hughes H, Sambrook J, MacCallum P (2012) Molecular cloning: a laboratory manual. In: Molecular cloning: a laboratory manual. pp 1890–1890

Greene HW (2005) Organisms in nature as a central focus for biology. Trends Ecol Evol 20:23–27. https://doi.org/10.1016/j.tree.2004.11.005

Grimaldi M, Riera B (2001) Geography and Climate. In: Monographiae Biologicae. pp 9–18

Guernier V, Hochberg ME, Guégan JF (2004) Ecology drives the worldwide distribution of human diseases. PLoS Biol 2:e141. https://doi.org/10.1371/journal.pbio.0020141

Guindon S, Dufayard JF, Lefort V, et al (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321. https://doi.org/10.1093/sysbio/syq010

Guix S, Bosch A, Pintó RM (2012) Astrovirus taxonomy. In: Astrovirus Research. Springer New York, New York, NY, pp 97–118

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: Nucleic acids symposium series. [London]: Information Retrieval Ltd., c1979-c2000., pp 95–98

Harrigan RJ, Sedano R, Chasar AC, et al (2014) New host and lineage diversity of avian haemosporidia in the northern andes. Evol Appl 7:799–811. https://doi.org/10.1111/eva.12176

Hellgren O, Waldenström J, Bensch S (2004) A new PCR assay for simultaneous studies of *Leucocytozoon*, *Plasmodium*, and *Haemoproteus* from avian blood. J Parasitol 90:797–802. https://doi.org/10.1645/ge-184r1

Hermanns K, Zirkel F, Kurth A, et al (2014) Cimodo virus belongs to a novel lineage of reoviruses isolated from African mosquitoes. J Gen Virol 95:905–909. https://doi.org/10.1099/vir.0.062349-0

Honkavuori KS, Briese T, Krauss S, et al (2014) Novel coronavirus and astrovirus in Delaware Bay shorebirds. PLoS One 9:e93395. https://doi.org/10.1371/journal.pone.0093395

Hromas R, Davis B, Rauscher FJ, et al (1996) Hematopoietic transcriptional regulation by the myeloid zinc finger gene, MZF-1. Curr Top Microbiol Immunol 211:159–164. https://doi.org/10.1007/978-3-642-85232-9_16

Hu D, Zhu C, Wang Y, et al (2017) Virome analysis for identification of novel mammalian viruses in bats from Southeast China. Sci Rep 7:10917. https://doi.org/10.1038/s41598-017-11384-w

Hulse-Post DJ, Sturm-Ramirez KM, Humberd J, et al (2005) Role of domestic ducks in the propagation and biological evolution of highly pathogenic H5N1 influenza viruses in Asia. Proc Natl Acad Sci 102:10682–10687. https://doi.org/10.1073/pnas.0504662102

Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. Genome Res 17:377–386. https://doi.org/10.1101/gr.5969107

Huson DH, Beier S, Flade I, et al (2016) MEGAN Community Edition - Interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput Biol 12:e1004957. https://doi.org/10.1371/journal.pcbi.1004957

Imada T, Yamaguchi S, Mase M, et al (2000) Avian nephritis virus (ANV) as a new member of the family *Astroviridae* and construction of infectious ANV cDNA. J Virol 74:8487–8493. https://doi.org/10.1128/JVI.74.18.8487-8493.2000

Jiang B, Monroe SS, Koonint E V, et al (1993) RNA sequence of astrovirus: distinctive genomic organization and a putative retrovirus-like ribosomal frameshifting signal that directs the viral replicase synthesis. Proc Natl Acad Sci USA 90:10539–10543. https://doi.org/10.1073/pnas.90.22.10539

Jones KE, Patel NG, Levy MA, et al (2008) Global trends in emerging infectious diseases. Nature 451:990–993. https://doi.org/10.1038/nature06536

Kalra BS, Chawla S, Gupta P, Valecha N (2006) Screening of antimalarial drugs: an overview. Indian J Pharmacol 38:5. https://doi.org/10.4103/0253-7613.19846

Kapgate SS, Barbuddhe SB, Kumanan K (2015) Next-generation sequencing

technologies: tool to study avian virus diversity. Acta Virol 59:3–13. https://doi.org/10.4149/av_2015_01_3

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

Keesing F, Belden LK, Daszak P, et al (2010) Impacts of biodiversity on the emergence and transmission of infectious diseases. Nature 468:647–652. https://doi.org/10.1038/nature09575

Keesing F, Holt RD, Ostfeld RS (2006) Effects of species diversity on disease risk. Ecol Lett 9:485–498. https://doi.org/10.1111/j.1461-0248.2006.00885.x

Kelley LA, Mezulis S, Yates CM, et al (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858. https://doi.org/10.1038/nprot.2015.053

Kim BY, Lee DH, Kim MS, et al (2012) Exchange of Newcastle disease viruses in Korea: The relatedness of isolates between wild birds, live bird markets, poultry farms and neighboring countries. Infect Genet Evol 12:478–482. https://doi.org/10.1016/j.meegid.2011.12.004

Kofstad T, Jonassen CM (2011) Screening of feral and wood pigeons for viruses harbouring a conserved mobile viral element: characterization of novel astroviruses and picornaviruses. PLoS One 6:e25964. https://doi.org/10.1371/journal.pone.0025964

Komar O, Robbins MB, Klenk K, et al (2003) West Nile virus transmission in resident birds, Dominican Republic. Emerg Infect Dis 9:1299–302. https://doi.org/10.3201/eid0910.030222

Koonin E V, Gorbalenyat AE, Purdy MA, et al (1992) Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: delineation of an additional group of positive-strand RNA plant and animal viruses. Proc Natl Acad Sci 89:8259–8263. https://doi.org/10.1073/pnas.89.17.8259

Kosugi S, Hasebe M, Tomita M, Yanagawa H (2009) Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of

composite motifs. Proc Natl Acad Sci U S A 106:10171–10176. https://doi.org/10.1073/pnas.0900604106

Kovács ER, Jánoska M, Dán Á, et al (2010) Recognition and partial genome characterization by non-specific DNA amplification and PCR of a new siadenovirus species in a sample originating from *Parus major*, a great tit. J Virol Methods 163:262–268. https://doi.org/10.1016/j.jviromet.2009.10.007

Krishnamurthy SR, Wang D (2017) Origins and challenges of viral dark matter. Virus Res 239:136–142. https://doi.org/10.1016/j.virusres.2017.02.002

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315

Kumar S, Stecher G, Li M, et al (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096

Kwok KTT, Nieuwenhuijse DF, Phan MVT, Koopmans MPG (2020) Virus metagenomics in farm animals: a systematic review. Viruses 12

La Cour T, Kiemer L, Mølgaard A, et al (2004) Analysis and prediction of leucine-rich nuclear export signals. Protein Eng Des Sel 17:527–536. https://doi.org/10.1093/protein/gzh062

Lachish S, Knowles SCL, Alves R, et al (2011) Fitness effects of endemic malaria infections in a wild bird population: the importance of ecological structure. J Anim Ecol 80:1196–1206. https://doi.org/10.1111/j.1365-2656.2011.01836.x

Lacorte GA, Flix GMF, Pinheiro RRB, et al (2013) Exploring the diversity and distribution of neotropical avian malaria parasites - A molecular survey from Southeast Brazil. PLoS One 8:. https://doi.org/10.1371/journal.pone.0057770

LaDeau SL, Kilpatrick AM, Marra PP (2007) West Nile virus emergence and large-scale declines of North American bird populations. Nature 447:710–713. https://doi.org/10.1038/nature05829

Lai KY, Ng GWY, Wong KF, et al (2013) Human H7N9 avian influenza virus infection: a review and pandemic risk assessment. Emerg Microbes Infect 2:e48.

https://doi.org/10.1038/emi.2013.48

Land H, Van Riper Iii C, Van Riper SG, et al (1986) The epizootiology and ecological significance of malaria in Hawaiian land birds. Source Ecol Monogr 56:327–344. https://doi.org/10.2307/1942550

Lebov J, Grieger K, Womack D, et al (2017) A framework for One Health research. One Heal 3:44–50. https://doi.org/10.1016/j.onehlt.2017.03.004

Levin II, Zwiers P, Deem SL, et al (2013) Multiple lineages of avian malaria parasites (*Plasmodium*) in the Galapagos Islands and evidence for arrival via migratory birds. Conserv Biol 27:1366–1377. https://doi.org/10.1111/cobi.12127

Li L, Pesavento PA, Gaynor AM, et al (2015) A gyrovirus infecting a sea bird. Arch Virol 160:2105–2109. https://doi.org/10.1007/s00705-015-2468-1

Li L, Victoria JG, Wang C, et al (2010) Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. J Virol 84:6955–6965. https://doi.org/10.1128/jvi.00501-10

Li M, Liu H, Bi Y, et al (2017) Highly pathogenic avian influenza A(H5N8) virus in wild migratory birds, Qinghai Lake, China. Emerg Infect Dis 23:637–641. https://doi.org/10.3201/eid2304.161866

Li Y, Wang K, Yu K, et al (2020) Identification and genome characterization of a novel picornavirus from ducks in China. Arch Virol 1–3. https://doi.org/10.1007/s00705-020-04691-7

Lima DA, Cibulski SP, Finkler F, et al (2017) Faecal virome of healthy chickens reveals a large diversity of the eukaryote viral community, including novel circular ssDNA viruses. J Gen Virol 98:690–703. https://doi.org/10.1099/jgv.0.000711

Lima DA, Cibulski SP, Tochetto C, et al (2019) The intestinal virome of malabsorption syndrome-affected and unaffected broilers through shotgun metagenomics. Virus Res 261:9–20. https://doi.org/10.1016/j.virusres.2018.12.005

Lina PHC, van Noord MJ, de Groot FG (1973) Detection of virus in squamous papillomas of the wild bird species *Fringilla coelebs*. J Natl Cancer Inst 50:567–571. https://doi.org/10.1093/jnci/50.2.567

Liu N, Jiang M, Dong Y, et al (2018) Genetic characterization of a novel group of avastroviruses in geese. Transbound Emerg Dis 65:927–932. https://doi.org/10.1111/tbed.12873

Liu N, Wang F, Zhang D (2014) Complete sequence of a novel duck astrovirus. Arch Virol 159:2823–2827. https://doi.org/10.1007/s00705-014-2141-0

Luiz Proença-Módena J, Santos Macedo I, Arruda E (2007) H5N1 avian influenza virus: an overview. Brazilian J Infect Dis 11:125–133. https://doi.org/10.1590/S1413-86702007000100027

Ma C, Rogers J, Zhou S, et al (2020) Modern pandemics: recession and recovery. Available SSRN 1–1

Maggi F, Macera L, Focosi D, et al (2012) Human gyrovirus DNA in human blood, Italy. J Virol 18:956. https://doi.org/10.1128/jvi.00639-11

Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16:562–563. https://doi.org/10.1093/bioinformatics/16.6.562

Martínez-De La Puente J, Merino S, Tomás G, et al (2010) The blood parasite *Haemoproteus* reduces survival in a wild bird: a medication experiment. Biol Lett 6:663–665. https://doi.org/10.1098/rsbl.2010.0046

Marzal A (2012) Recent advances in studies on avian malaria parasites. In: Malaria Parasites. InTech, pp 135–158

Marzal A, De Lope F, Navarro C, Møller AP (2005) Malarial parasites decrease reproductive success: an experimental study in a passerine bird. Oecologia 142:541–545. https://doi.org/10.1007/s00442-004-1757-2

Marzal A, García-Longoria L, Cárdenas Callirgos JM, Sehgal RNM (2015) Invasive avian malaria as an emerging parasitic disease in native birds of Peru. Biol Invasions 17:39–45. https://doi.org/10.1007/s10530-014-0718-x

McWilliam H, Li W, Uludag M, et al (2013) Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res 41:W597-600. https://doi.org/10.1093/nar/gkt376

Medeiros MCI, Hamer GL, Ricklefs RE (2013) Host compatibility rather than vector-host-encounter rate determines the host range of avian *Plasmodium* parasites. Proc

R Soc B Biol Sci 280:. https://doi.org/10.1098/rspb.2012.2947

Mendenhall IH, Yaung KN, Joyner PH, et al (2015) Detection of a novel astrovirus from a black-naped monarch (*Hypothymis azurea*) in Cambodia. Virol J 12:182. https://doi.org/10.1186/s12985-015-0413-2

Méndez E, Murillo A, Velázquez R, et al (2012) Replication cycle of astroviruses. In: Astrovirus Research. Springer New York, New York, NY, pp 19–45

Michel F, Fischer D, Eiden M, et al (2018) West Nile virus and Usutu virus monitoring of wild birds in Germany. Int J Environ Res Public Health 15:171. https://doi.org/10.3390/ijerph15010171

Mihalov-Kovács E, Fehér E, Martella V, et al (2014) The fecal virome of domesticated animals. VirusDisease 25:150–157. https://doi.org/10.1007/s13337-014-0192-1

Miles AM, Reddy SM, Morgan RW (2001) Coinfection of specific-pathogen-free chickens with Marek's disease virus (MDV) and chicken infectious anemia virus: effect of MDV pathotype. Avian Dis 45:9–18

Moens MAJ, Pérez-Tris J (2016) Discovering potential sources of emerging pathogens: South America is a reservoir of generalist avian blood parasites. Int J Parasitol 46:41–49. https://doi.org/10.1016/j.ijpara.2015.08.001

Moens MAJ, Pérez-Tris J, Cortey M, Benítez L (2018) Identification of two novel CRESS DNA viruses associated with an Avipoxvirus lesion of a blue-and-gray Tanager (*Thraupis episcopus*). Infect Genet Evol 60:89–96. https://doi.org/10.1016/j.meegid.2018.02.015

Moens MAJ, Valkiūnas G, Paca A, et al (2016) Parasite specialization in a unique habitat: hummingbirds as reservoirs of generalist blood parasites of Andean birds. J Anim Ecol 85:1234–1245. https://doi.org/10.1111/1365-2656.12550

Morgan ER, Lundervold M, Medley GF, et al (2006) Assessing risks of disease transmission between wildlife and livestock: the saiga antelope as a case study. Biol Conserv 131:244–254. https://doi.org/10.1016/j.biocon.2006.04.012

Morse SS, Mazet JAK, Woolhouse M, et al (2012) Prediction and prevention of the next pandemic zoonosis. Lancet 380:1956–1965. https://doi.org/10.1016/S0140-6736(12)61684-5

Myers N, Mittermeier RA, Mittermeier CG, et al (2000) Biodiversity hotspots for conservation priorities. Nature 403:853–858. https://doi.org/10.1038/35002501

Ng TFF, Mesquita JR, Nascimento MSJ, et al (2014) Feline fecal virome reveals novel and prevalent enteric viruses. Vet Microbiol 171:102–11. https://doi.org/10.1016/j.vetmic.2014.04.005

Noteborn MHM, Koch G (1995) Chicken anaemia virus infection: molecular basis of pathogenicity. Avian Pathol 24:11–31. https://doi.org/10.1080/03079459508419046

Noteborn MHM, Todd D, Verschueren CAJ, et al (1994) A single chicken anemia virus protein induces apoptosis. J Virol 68:346–351

Nourani L, Zakeri S, Dinparast Djadid N (2020) Dynamics of prevalence and distribution pattern of avian *Plasmodium* species and its vectors in diverse zoogeographical areas - A review. Infect Genet Evol 81:104244. https://doi.org/10.1016/j.meegid.2020.104244

Oakgrove KS, Harrigan RJ, Loiseau C, et al (2014) Distribution, diversity and drivers of blood-borne parasite co-infections in Alaskan bird populations. Int J Parasitol 44:717–727. https://doi.org/10.1016/j.ijpara.2014.04.011

Olival KJ, Latinne A, Islam A, et al (2020) Population genetics of fruit bat reservoir informs the dynamics, distribution and diversity of Nipah virus. Mol Ecol 29:970–985. https://doi.org/10.1111/mec.15288

Orme CDL, Davies RG, Burgess M, et al (2005) Global hotspots of species richness are not congruent with endemism or threat. Nature 436:1016–1019. https://doi.org/10.1038/nature03850

Pankovics P, Boros Á, Kiss T, et al (2015) Detection of a mammalian-like astrovirus in bird, European roller (*Coracias garrulus*). Infect Genet Evol 34:114–121. https://doi.org/10.1016/J.MEEGID.2015.06.020

Pankovics P, Boros Á, Phan TG, et al (2018) A novel passerivirus (family *Picornaviridae*) in an outbreak of enteritis with high mortality in estrildid finches (*Uraeginthus* sp.). Arch Virol 163:1063–1071. https://doi.org/10.1007/s00705-017-3699-0

Pantin-Jackwood M, Todd D, Koci MD (2013) Avian astroviruses. In: Schultz-Cherry S

(ed) Astrovirus Research: Essential Ideas, Everyday Impacts, Future Directions. Springer New York, New York, NY, pp 151–180

Peiris JSM, De Jong MD, Guan Y (2007) Avian influenza virus (H5N1): a threat to human health. Clin Microbiol Rev 20:243–267. https://doi.org/10.1128/CMR.00037-06

Pereda AJ, Uhart M, Perez AA, et al (2008) Avian influenza virus isolated in wild waterfowl in Argentina: evidence of a potentially unique phylogenetic lineage in South America. Virology 378:363–370. https://doi.org/10.1016/J.VIROL.2008.06.010

Pérez-Tris J, Bensch S (2005) Dispersal increases local transmission of avian malarial parasites. Ecol Lett 8:838–845. https://doi.org/10.1111/j.1461-0248.2005.00788.x

Peters MA, Jackson DC, Crabb BS, Browning GF (2002) Chicken anemia virus VP2 is a novel dual specificity protein phosphatase. J Biol Chem 277:39566–39573. https://doi.org/10.1074/jbc.M201752200

Peters MK (2010) Ant-following and the prevalence of blood parasites in birds of African rainforests. J Avian Biol 41:105–110. https://doi.org/10.1111/j.1600-048X.2010.04896.x

Phan TG, da Costa AC, Zhang W, et al (2015) A new gyrovirus in human feces. Virus Genes 51:132–135. https://doi.org/10.1007/s11262-015-1210-0

Phan TG, Vo NP, Boros Á, et al (2013) The viruses of wild pigeon droppings. PLoS One 8:e72787. https://doi.org/10.1371/journal.pone.0072787

Plowright RK, Foley P, Field HE, et al (2011) Urban habituation, ecological connectivity and epidemic dampening: The emergence of hendra virus from flying foxes (*Pteropus* spp.). Proc R Soc B Biol Sci 278:3703–3712. https://doi.org/10.1098/rspb.2011.0522

Prilusky J, Felder CE, Zeev-Ben-Mordehai T, et al (2005) FoldIndex(C): a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 21:3435–3438. https://doi.org/10.1093/bioinformatics/bti537

Prosser DJ, Hungerford LL, Erwin RM, et al (2013) Mapping avian influenza transmission risk at the interface of domestic poultry and wild birds. Front Public Heal 1:28. https://doi.org/10.3389/fpubh.2013.00028

Purdy MA, Harrison TJ, Jameel S, et al (2017) ICTV virus taxonomy profile: *Hepeviridae*. J Gen Virol 98:2645–2646. https://doi.org/10.1099/jgv.0.000940

R Development Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Ramírez-Martínez LA, Loza-Rubio E, Mosqueda J, et al (2018) Fecal virome composition of migratory wild duck species. PLoS One 13:e0206970. https://doi.org/10.1371/journal.pone.0206970

Reed KD, Meece JK, Henkel JS, Reed K (2002) Birds, migration and emerging zoonoses: West Nile virus, Lyme disease, influenza A and enteropathogens. Clin Med Res 1:5–12. https://doi.org/10.3121/cmr.1.1.5

Reilly JR, Hajek AE (2012) Prey-processing by avian predators enhances virus transmission in the gypsy moth. Oikos 121:1311–1316. https://doi.org/10.1111/j.1600-0706.2011.19752.x

Reuter G, Boros Á, Mátics R, et al (2020) Detection and complete genome characterization of a novel RNA virus related to members of the Hepe-Virga clade in bird species, hoopoe (*Upupa epops*). Infect Genet Evol 81:104236. https://doi.org/10.1016/j.meegid.2020.104236

Reuter G, Boros Á, Tóth Z, et al (2018) Detection of a novel RNA virus with hepatitis E virus-like non-structural genome organization in amphibian, agile frog (*Rana dalmatina*) tadpoles. Infect Genet Evol 65:112–116. https://doi.org/10.1016/j.meegid.2018.07.029

Rijsewijk FAM, dos Santos HF, Teixeira TF, et al (2011) Discovery of a genome of a distant relative of chicken anemia virus reveals a new member of the genus Gyrovirus. Arch Virol 156:1097–1100. https://doi.org/10.1007/s00705-011-0971-6

Rivero A, Gandon S (2018) Evolutionary ecology of avian malaria: past to present. Trends Parasitol 34:712–726. https://doi.org/10.1016/j.pt.2018.06.002

Rizotto LS, Simão RM, Scagion GP, et al (2019) Detection of avian metapneumovirus subtype a from wild birds in the state of São Paulo, Brazil. Pesqui Vet Bras 39:209–213. https://doi.org/10.1590/1678-5150-PVB-6021

Rohaim MA, El Naggar RF, Helal AM, et al (2017) Reverse spillover of avian viral

vaccine strains from domesticated poultry to wild birds. Vaccine 35:3523–3527. https://doi.org/10.1016/j.vaccine.2017.05.033

Roossinck MJ, Bazán ER (2017) Symbiosis: viruses as intimate partners. Annu Rev Virol 4:123–139. https://doi.org/10.1146/annurev-virology-110615-042323

Rosario K, Breitbart M, Harrach B, et al (2017) Revisiting the taxonomy of the family *Circoviridae*: establishment of the genus *Cyclovirus* and removal of the genus *Gyrovirus*. Arch Virol 162:1447–1463. https://doi.org/10.1007/s00705-017-3247-y

Rosseel T, Ozhelvaci O, Freimanis G, Van Borm S (2015) Evaluation of convenient pretreatment protocols for RNA virus metagenomics in serum and tissue samples. J Virol Methods 222:72–80. https://doi.org/10.1016/j.jviromet.2015.05.010

Roussan DA, Jordan P (2006) Serological survey on the prevalence of chicken infectious anemia virus in commercial broiler chicken flocks in Northern Jordan. Int J Poult Sci 5:544–546

Schat KA (2009) Chicken anemia virus. In: Current Topics in Microbiology and Immunology. Springer, Berlin, Heidelberg, pp 151–183

Scherbatskoy EC, Subramaniam K, Al-Hussinee L, et al (2020) Characterization of a novel picornavirus isolated from moribund aquacultured clownfish. J Gen Virol jgv001421. https://doi.org/10.1099/jgv.0.001421

Şekercioğlu ÇH, Wenny DG, Whelan CJ, Floyd C (2016) Why birds matter: avian ecological function and ecosystem services. University of Chicago Press.

Sherratt TN, Wilkinson DM (2009) Why are the tropics so diverse? Oxford University Press, pp 111–135

Shi M, Lin XD, Tian JH, et al (2016) Redefining the invertebrate RNA virosphere. Nature 540:539–543. https://doi.org/10.1038/nature20167

Simmonds P, Adams MJ, Benk M, et al (2017) Consensus statement: virus taxonomy in the age of metagenomics. Nat Rev Microbiol 15:161–168. https://doi.org/10.1038/nrmicro.2016.177

Slusher MJ, Wilcox BR, Lutrell MP, et al (2014) Are passerine birds reservoirs for influenza a viruses? J Wildl Dis 50:792–809. https://doi.org/10.7589/2014-02-043

Smith GJD, Vijaykrishna D, Bahl J, et al (2009) Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza a epidemic. Nature 459:1122–1125. https://doi.org/10.1038/nature08182

Snoeck CJ, Komoyo GF, Mbee BP, et al (2012) Epidemiology of chicken anemia virus in Central African Republic and Cameroon. Virol J 9:1–10. https://doi.org/10.1186/1743-422X-9-189

Sun J, He WT, Wang L, et al (2020) COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. Trends Mol Med 26:483–495. https://doi.org/10.1016/j.molmed.2020.02.008

Swayne DE (2008) Avian influenza. Blackwell Pub

Taylor LH, Latham SM, Woolhouse ME (2001) Risk factors for human disease emergence. Philos Trans R Soc Lond B Biol Sci 356:983–9. https://doi.org/10.1098/rstb.2001.0888

Temmam S, Hul V, Bigot T, et al (2019) A novel *Polycipiviridae* virus identified in *Pteropus lylei* stools. Microbiol Resour Announc 8:. https://doi.org/10.1128/mra.01662-18

Thiel T, Whiteman NK, Tirapé A, et al (2005) Characterization of canarypox-like viruses infecting endemic birds in the Galápagos Islands. J Wildl Dis. https://doi.org/10.7589/0090-3558-41.2.342

Thiollay JM (2002) Avian diversity and distribution in French Guiana: patterns across a large forest landscape. J Trop Ecol 471–498. https://doi.org/10.1017/S026646740200233X

Thompson RCA, Lymbery AJ, Smith A (2010) Parasites, emerging disease and wildlife conservation. Int J Parasitol 40:1163–1170. https://doi.org/10.1016/j.ijpara.2010.04.009

Todd D, Smyth VJ, Ball NW, et al (2009) Identification of chicken enterovirus-like viruses, duck hepatitis virus type 2 and duck hepatitis virus type 3 as astroviruses. Avian Pathol 38:21–29. https://doi.org/10.1080/03079450802632056

Truchado DA, Diaz-Piqueras JM, Gomez-Lucia E, et al (2019) A novel and divergent gyrovirus with unusual genomic features detected in wild passerine birds from a

remote rainforest in French Guiana. Viruses 11:1148. https://doi.org/10.3390/v11121148

Tsai FY, Orkin SH (1997) Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. Blood 89:3636–3643. https://doi.org/10.1182/blood.v89.10.3636.3636_3636_3643

Tsunoda T, Takagi T (1999) Estimating transcription factor bindability on DNA. Bioinformatics 15:622–630. https://doi.org/10.1093/bioinformatics/15.7.622

Valkiunas G (2004) Avian malaria parasites and other haemosporidia. CRC press

Valkiūnas G, Anwar AM, Atkinson CT, et al (2005) What distinguishes malaria parasites from other pigmented haemosporidians? Trends Parasitol 21:357–358. https://doi.org/10.1016/j.pt.2005.06.001

Valkiunas G, Salaman P, Iezhova TA (2003) Paucity of hematozoa in Colombian birds. J Wildl Dis 39:445–448. https://doi.org/10.7589/0090-3558-39.2.445

van Riper III C, van Riper SG, Hansen WR (2002) Epizootiology and effect of avian pox on Hawaiian forest birds. Auk 119:929–942. https://doi.org/10.1642/0004-8038(2002)119[0929:EAEOAP]2.0.CO;2

Varela APM, dos Santos HF, Cibulski SP, et al (2014) Chicken anemia virus and avian gyrovirus 2 as contaminants in poultry vaccines. Biologicals 42:346–350. https://doi.org/10.1016/j.biologicals.2014.08.002

Verhagen JH, Herfst S, Fouchier RA (2015) How a virus travels the world. Science 347:616–617. https://doi.org/10.1126/science.1259924

Viana DS, Santamaría L, Figuerola J (2016) Migratory birds as global dispersal vectors. Trends Ecol Evol 31:763–775. https://doi.org/10.1016/j.tree.2016.07.005

Vibin J, Chamings A, Collier F, et al (2018) Metagenomics detection and characterisation of viruses in faecal samples from Australian wild birds. Sci Rep 8:1–23. https://doi.org/10.1038/s41598-018-26851-1

Vilibic-Cavlek T, Savic V, Petrovic T, et al (2019) Emerging trends in the epidemiology of West Nile and Usutu virus infections in Southern Europe. Front. Vet. Sci. 6

Waits K, Bradley RW, Warzybok P, et al (2018) Genome sequence of a gyrovirus associated with ashy storm-petrel. Microbiol Resour Announc 7:e00958-18. https://doi.org/10.1128/MRA.00958-18

Waldenström J, Bensch S, Hasselquist D, Östman Ö (2004) A new nested polymerase chain reaction method very efficient in detecting *Plasmodium* and *Haemoproteus* infections from avian blood. J Parasitol 90:191–194. https://doi.org/10.1645/ge-3221rn

Wang H, Yang S, Shan T, et al (2019a) A novel picornavirus in feces of a rainbow lorikeet (*Trichoglossus moluccanus*) shows a close relationship to members of the genus *Avihepatovirus*. Arch Virol 164:1911–1914. https://doi.org/10.1007/s00705-019-04246-5

Wang, Yang S, Liu D, et al (2019b) The fecal virome of red-crowned cranes. Arch Virol 164:3–16. https://doi.org/10.1007/s00705-018-4037-x

Warner RE (1968) The role of introduced diseases in the extinction of the endemic Hawaiian avifauna. Source: The Condor 70:101–120

Webster RG, Yakhno M, Hinshaw VS, et al (1978) Intestinal influenza: replication and characterization of influenza viruses in ducks. Virology 84:268–278. https://doi.org/10.1016/0042-6822(78)90247-7

WHO (2019) Ten threats to global health in 2019. World Heal. Organ.

Wille M, Eden JS, Shi M, et al (2018) Virus–virus interactions and host ecology are associated with RNA virome structure in wild birds. Mol Ecol 27:5263–5278. https://doi.org/10.1111/mec.14918

Wille M, Shi M, Klaassen M, et al (2019) Virome heterogeneity and connectivity in waterfowl and shorebird communities. ISME J 13:2603–2616. https://doi.org/10.1038/s41396-019-0458-0

Williams RB (2005) Avian malaria: clinical and chemical pathology of *Plasmodium gallinaceum* in the domesticated fowl *Gallus gallus*. Avian Pathol 34:29–47. https://doi.org/10.1080/03079450400025430

Williams SH, Che X, Garcia JA, et al (2018) Viral diversity of house mice in New York City. MBio 9:e01354-17. https://doi.org/10.1128/mBio

Wolf E, Kim PS, Berger B (1997) MultiCoil: a program for predicting two- and three-stranded coiled coils. Protein Sci 6:1179–89. https://doi.org/10.1002/pro.5560060606

Wolfaardt M, Kiulia NM, Mwenda JM, Taylor MB (2011) Evidence of a recombinant wild-type human astrovirus strain from a Kenyan child with gastroenteritis. J Clin Microbiol 49:728–31. https://doi.org/10.1128/JCM.01093-10

Woo PCY, Lau SKP, Huang Y, et al (2010) Comparative analysis of six genome sequences of three novel picornaviruses, turdiviruses 1, 2 and 3, in dead wild birds, and proposal of two novel genera, *Orthoturdivirus* and *Paraturdivirus*, in the family *Picornaviridae*. J Gen Virol 91:2433–2448. https://doi.org/10.1099/vir.0.021717-0

Woolhouse MEJ, Gowtage-Sequeria S (2005) Host range and emerging and reemerging pathogens. Emerg Infect Dis 11:1842–7. https://doi.org/10.3201/eid1112.050997

World Health Organization (2019) World malaria report 2019. World Health Organization

Wu N, Zhang P, Liu W, Wang X (2018) Sogatella furcifera hepe-like virus: first member of a novel *Hepeviridae* clade identified in an insect. Virus Res 250:81–86. https://doi.org/10.1016/j.virusres.2018.03.018

Wu Z, Ren X, Yang L, et al (2012) Virome analysis for identification of novel mammalian viruses in bat species from Chinese provinces. J Virol 86:10999–11012. https://doi.org/10.1128/jvi.01394-12

Xiang B, Han L, Gao P, et al (2017) Spillover of Newcastle disease viruses from poultry to wild birds in Guangdong province, southern China. Infect Genet Evol 55:199–204. https://doi.org/10.1016/j.meegid.2017.09.020

Yao S, Gao X, Tuo T, et al (2017) Novel characteristics of the avian gyrovirus 2 genome. Sci Rep 7:41068. https://doi.org/10.1038/srep41068

Yuasa N, Taniguchi T, Imada T, Hihara H (1983) Distribution of chicken anemia agent (CAA) and detection of neutralizing antibody in chicks experimentally inoculated with CAA. Natl Inst Anim Health Q (Tokyo) 23:78–81

Zell R, Delwart E, Gorbalenya AE, et al (2017) ICTV virus taxonomy profile: *Picornaviridae*. J Gen Virol 98:2421–2422. https://doi.org/10.1099/jgv.0.000911

Zhang X, Liu Y, Ji J, et al (2014) Identification of a chicken anemia virus variant-related gyrovirus in stray cats in China, 2012. Biomed Res Int. https://doi.org/10.1155/2014/313252

Zhang Y, Wang F, Liu N, et al (2017) Complete genome sequence of a novel avastrovirus in goose. Arch Virol 162:2135–2139. https://doi.org/10.1007/s00705-017-3297-1

Zhao L, Niu Y, Lu T, et al (2018) Metagenomic analysis of the Jinding duck fecal virome. Curr Microbiol 75:658–665. https://doi.org/10.1007/s00284-018-1430-3

Zsivanovits P, Monks DJ, Forbes NA, et al (2006) Presumptive identification of a novel adenovirus in a Harris hawk (*Parabuteo unicinctus*), a Bengal eagle owl (*Bubo bengalensis*), and a Verreaux's eagle owl (*Bubo lacteus*). J Avian Med Surg 20:105–112. https://doi.org/10.1647/2005-008.1

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–15. https://doi.org/10.1093/nar/gkg595