

# SCIENTIFIC REPORTS



OPEN

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

Izaskun Fernández-Correa<sup>1</sup>, Daniel A. Truchado<sup>1,2</sup>, Esperanza Gomez-Lucia<sup>3</sup>, Ana Doménech<sup>3</sup>, Javier Pérez-Tris<sup>2</sup>, Jonas Schmidt-Chanasit<sup>4,5</sup>, Daniel Cadar<sup>4</sup> & Laura Benítez<sup>1</sup>

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.

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<sup>1</sup>. 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<sup>2</sup>. 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<sup>3</sup> and that mammal and avian species are their main hosts<sup>4,5</sup>, 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<sup>2,6</sup>. For example, the great diversity of the Neotropical fauna harbors the risk of emergence of new infections in domestic animals and

<sup>1</sup>Department of Genetics, Physiology and Microbiology, Faculty of Biology, Complutense University of Madrid, José Antonio Novais, 12, 28040, Madrid, Spain. <sup>2</sup>Department of Biodiversity, Ecology and Evolution, Faculty of Biology, Complutense University of Madrid, José Antonio Novais, 12, 28040, Madrid, Spain. <sup>3</sup>Department of Animal Health, Faculty of Veterinary Medicine, Complutense University of Madrid, Avda. Puerta de Hierro s/n, 28040, Madrid, Spain. <sup>4</sup>Bernhard-Nocht-Institut für 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. <sup>5</sup>University of Hamburg, Faculty of Mathematics, Informatics and Natural Sciences, Hamburg, Germany. Izaskun Fernández-Correa and Daniel A. Truchado contributed equally. Correspondence and requests for materials should be addressed to L.B. (email: [lbenitez@ucm.es](mailto:lbenitez@ucm.es))

humans that might be in contact with these birds or ecosystem because of human impact on nature. Conversely, the valuable biodiversity of these regions may be threatened by invasive viruses imported into the Neotropics<sup>7,8</sup>. There is growing evidence that the risk of pathogen transfer is increasing in both directions<sup>9–11</sup>, 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<sup>12,13</sup>. New astroviruses have been discovered as part of the intestinal virome of wildlife, including bats<sup>14</sup> and especially birds<sup>15</sup>, 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<sup>15–17</sup>, 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<sup>18</sup>. 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<sup>19</sup>. A viral genome-linked protein (VPg) with a key role in genome replication or protein synthesis is also encoded in the genome<sup>20,21</sup>. This protein is linked to the 5'-end of astrovirus genome and it possesses a TEEY-like residue that has been demonstrated to be related to viral infectivity in human astroviruses<sup>20</sup>.

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<sup>22,23</sup>. However, astrovirus infections usually occur asymptotically in this group of animals, only causing mild disease. The most frequent transmission route is fecal-oral<sup>19,24</sup>, 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<sup>25</sup>.

The classification of astroviruses has been redefined several times since they were first discovered in 1975<sup>13</sup>. According to ICTV, the current classification does not correspond to the phylogeny of this group of viruses, being 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<sup>13,24</sup>. 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<sup>13</sup>. 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<sup>26</sup>.

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.

## 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 samples 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). Attempts to isolate the novel astroviruses detected here were not possible due to very limited availability of the amount of samples.

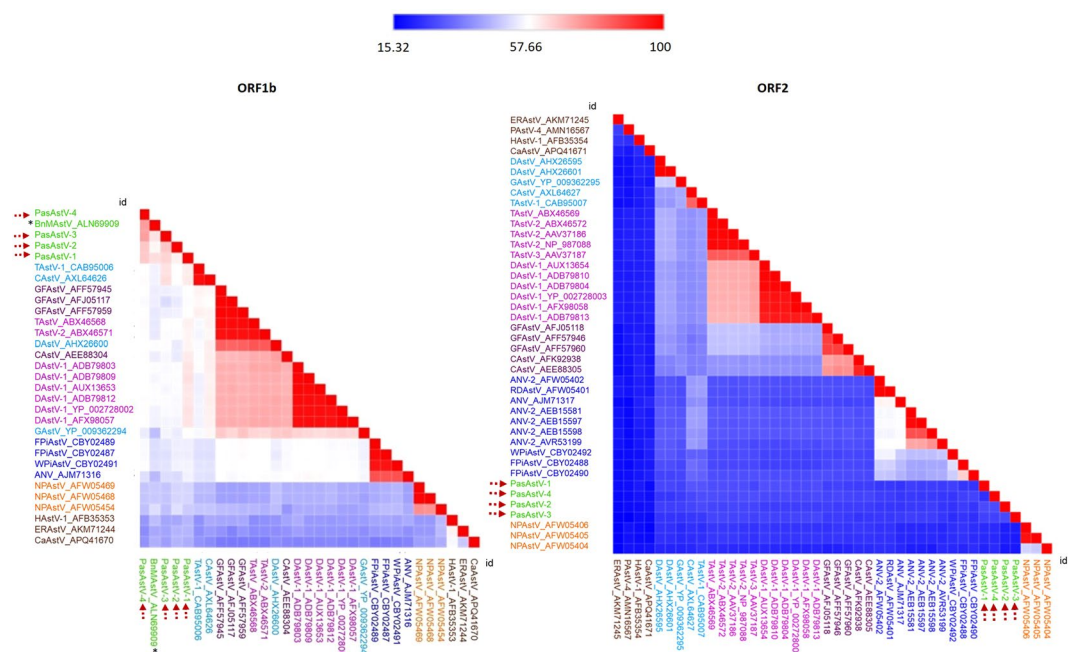
Identity matrices (Fig. 1) showed high divergences of the four novel astroviruses from French Guianan passerine birds, compared to previously known astroviruses. Amino acid sequence identity of capsid protein (ORF2) varied from 35% to 93% within the genus *Avastrovirus*, but we found lower identities 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 (Fig. 1). Genetic distance analysis also indicated that the passerine avastroviruses can be further subgrouped into four supported clades: *Passerine astrovirus 1* (PasAsTV-1), *Passerine astrovirus 2* (PasAsTV-2), *Passerine astrovirus 3* (PasAsTV-3) and *Passerine*

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

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

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



**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 Fig. 4). Difference in size between the two heatmaps is due to the different availability of sequences for both ORFs in Genbank.

*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.

**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 (Fig. 2A). None of them have exhibited the conserved genomic RNA promoter sequence at 5' UTR<sup>19</sup> as based on the Interproscan analyses results using Geneious v1.1. 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<sup>19</sup> (Fig. 2B). Distinctive sgRNA promoter sequences (5'-AUUUGGAGNGGNGGACCNAAN-3') were found at the 3' end of the four ORF1b<sup>19</sup>. A structure similar to the stem-loop 2-like (s2m) described for astroviruses<sup>19</sup> 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 S1) except for a hypervariable region in ORF1. The VPg sequences, with some amino acid variations, were conserved in all four novel astroviruses (Fig. 3).

**Phylogeny and taxonomy.** The analysis carried out to remove potential ORF2 recombinants from our dataset 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 (Fig. 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.*<sup>17</sup> formed a distinct, early diverging clade, which we named *Avastrovirus 4*. The four newly discovered passerine astroviruses 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).

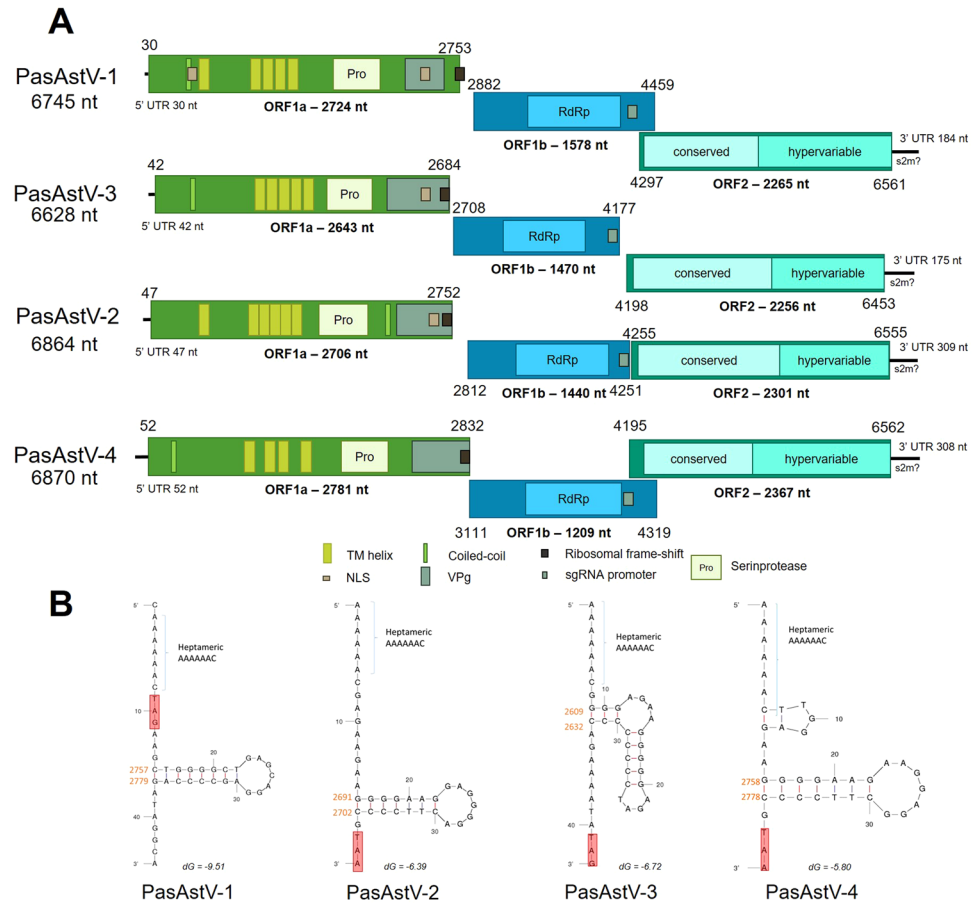
## Discussion

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<sup>12,13</sup>. 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.

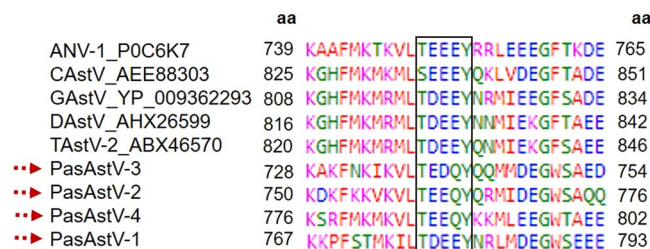
Our results significantly expanded the known range of hosts of avastroviruses and they provide further insight into genetic diversity, evolution and population structure of these viruses. Earlier reports on astroviruses from passerines were scarce or absent, with only one representative member detected in a black-napped monarch (*Hypothymis azurea*) from Cambodia<sup>27</sup>. 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<sup>28,29</sup>. 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<sup>17</sup>). 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.*<sup>27</sup>. This represents 0.8% prevalence, comparable with 5.1% of passerines infected in French Guiana (20 positive out of 395 tested).

The four novel astroviruses found in French Guiana substantially contribute to 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)<sup>30</sup>. 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<sup>31–33</sup>, although this matter is not addressed in these studies. 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





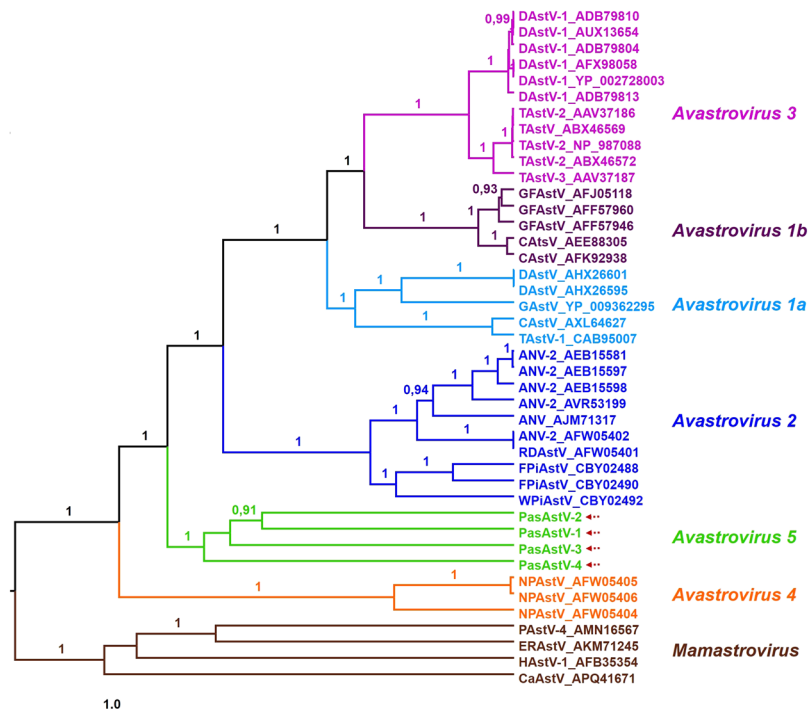
**Figure 2.** (A) Schematic representation of the genomic organization of the four novel PasAstV. NLS, nuclear localization sequence; VPg, viral protein associated with the genome; RdRp, RNA-dependent RNA polymerase; sgRNA, subgenomic RNA; s2m, stem-loop 2-like. (B) Prediction of secondary RNA structure 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 avastroviruses (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 TEEYY motif (black rectangle).

our local study of birds from the Guianan shield. The diversity of *Avastrovirus* can be further broadened by the recognition of *Avastrovirus 4*, a group of early diverging viruses found in a relatively well-sampled group of birds (ducks). Avastroviruses included in *Avastrovirus 4* already appeared forming a different clade when they were first described in both RdRp and ORF2 phylogenetic trees<sup>17</sup>. 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* (avastroviruses from Northern pintail) and *Avastrovirus 5* (avastroviruses 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 avastroviruses 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, an



**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 ( $\geq 90\%$ ) 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.

especial characteristic that only occurs in the DAstV isolate described by Fu *et al.*<sup>34</sup>, since astroviruses have usually two ORFs on the same reading frame. PasAstV-4 has ORF1b in +1 frame, the same as ANV-1<sup>24</sup>. 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 TEEY-like domain in the VPg putative protein. The ORF sizes of astroviruses from French Guiana are more similar to those from *Mamastrovirus* than from *Avastrovirus*. The 5' ends of the four passerine astroviruses are within *Astroviridae* range (11–85 nt)<sup>35</sup>, 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<sup>36</sup>. It is also interesting the *Mamastrovirus*-like structure present on the ribosomal frame-shift signal located between the ORF1a and 1b of PasAstV-1. Stop-codon location between AAAAAAC heptamer and the stem-loop in the ORF1a suggests that the ancestor of this astrovirus likely originated from a recombination event between *Mamastrovirus* and *Avastrovirus*, even though recombinations between the two genera have not been described so far<sup>13</sup>. Likewise, only one case of recombination in a human astrovirus between ORF1a and ORF1b has been described<sup>37</sup>. 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<sup>36</sup>.

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<sup>19,38</sup>, 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.

## 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 (4°05'N, 52°40'W). The sampling area was located in a tropical rainforest, where average temperature is around 26 °C 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. The VTM were centrifuged at 13,000 rpm for 1 min to pellet out epithelial cells. VTM were removed and the pellets resuspended in 250 µl of PBS. Then, samples were subjected to 2 freeze-thaw cycles at –80 °C 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 after 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, with Phred quality score < 20 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<sup>39</sup> and CLC workbench and compared with a non-redundant and viral proteome database (NCBI) 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<sup>40</sup>.

**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.*<sup>23</sup> 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<sup>-1</sup> 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<sub>2</sub>O up to 25 µl using a cycling condition as follows: synthesis of cDNA at 50 °C 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 minute. 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<sup>41</sup> was employed for prediction of transmembrane helices in proteins; cNLS Mapper<sup>42</sup> 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<sup>43</sup> for locating the viral protein associated with the genome (VPg), and by Multicoil scoring form<sup>44</sup> for locating the coiled-coil regions. Hairpin loops were analysed using Mfold<sup>45</sup>.

**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<sup>46</sup>.

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<sup>47</sup>. 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<sup>31</sup>. The analysis was performed under the best fit amino acid substitution model identified as LG + F + I + F using Bayesian Information Criterion implemented as the model selection framework in jModel-Test<sup>48</sup> and MEGA X<sup>49</sup>. Monte Carlo Markov Chains (MCMC) were run for 10<sup>7</sup> generations, sampling every 1,000 trees. Traces were inspected for convergence with Tracer 1.5<sup>50</sup>. The 10,000 resulting trees were summarized with TreeAnnotator v2.1.2<sup>50</sup> 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 1000 replicates and partial deletion (95%). The identity matrices were generated with the MUSCLE tool<sup>51</sup> (<https://www.ebi.ac.uk/Tools/msa/muscle/>), and transformed to HeatMaps with Morpheus software (<https://software.broadinstitute.org/morpheus>).

**Ethics statement.** All methods were carried out in accordance with European Union and national French regulations. Capture, sampling and transport of samples were authorised by the Service of Natural Environments, Biodiversity, Sites and Landscapes, Regional Directorate for the Environment, Planning and Housing at French Guiana (license 030418). The experimental protocols were approved by the Committee on Animal Testing of Comptense University (CEA-UCM, authorisation number 44-2016).

## References

1. Bascompte, J. Structure and dynamics of ecological networks. *Science* **329**, 765–766 (2010).
2. Carroll, D. *et al.* The Global Virome Project. *Science* **359**, 872–874 (2018).
3. Woolhouse, M. E. J. & Gowtage-Sequeria, S. Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.* **11**, 1842–1847 (2005).
4. Taylor, L. H., Latham, S. M. & Woolhouse, M. E. Risk factors for human disease emergence. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **356**, 983–989 (2001).
5. Morse, S. S. *et al.* Prediction and prevention of the next pandemic zoonosis. *The Lancet* **380**, 1956–1965. [https://doi.org/10.1016/S0140-6736\(12\)61684-5](https://doi.org/10.1016/S0140-6736(12)61684-5) (2012).
6. Moens, M. A. J. & Pérez-Tris, J. Discovering potential sources of emerging pathogens: South America is a reservoir of generalist avian blood parasites. *Int. J. Parasitol.* **46**, 41–49 (2016).
7. Keesing, F. *et al.* Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* **468**, 647–652 (2010).
8. Daszak, P., Cunningham, A. A. & Hyatt, A. D. Emerging infectious diseases of wildlife - threats to biodiversity and human health. *Science* **287**, 443–449 (2000).
9. Komar, O. *et al.* West Nile virus transmission in resident birds, Dominican Republic. *Emerg. Infect. Dis.* **9**, 1299–302 (2003).
10. Pereda, A. J. *et al.* Avian influenza virus isolated in wild waterfowl in Argentina: Evidence of a potentially unique phylogenetic lineage in South America. *Virology* **378**, 363–370 (2008).
11. Ferreira-Junior, F. C. *et al.* A new pathogen spillover from domestic to wild animals: *Plasmodium juxtanculare* infects free-living passerines in Brazil. *Parasitology* **145**, 1949–1958 (2018).
12. Cortez, V. *et al.* Astrovirus biology and pathogenesis. *Annu. Rev. Virol.* **4**, 327–348 (2017).
13. Donato, C. & Vijaykrishna, D. The broad host range and genetic diversity of mammalian and avian astroviruses. *Viruses* **9**, 102. <https://doi.org/10.3390/v9050102> (2017).
14. Hu, D. *et al.* Virome analysis for identification of novel mammalian viruses in bats from Southeast China. *Sci. Rep.* **7**, 10917 (2017).
15. Honkavuori, K. S. *et al.* Novel coronavirus and astrovirus in Delaware Bay shorebirds. *PLoS One* **9**, e93395 (2014).
16. Kofstad, T. & Jonassen, C. M. Screening of feral and wood pigeons for viruses harbouring a conserved mobile viral element: Characterization of novel astroviruses and picornaviruses. *PLoS One* **6**, e25964 (2011).
17. Chu, D. K. W. *et al.* A novel group of avian astroviruses in wild aquatic birds. *J. Virol.* **86**, 13772–13778 (2012).
18. Jiang, B., Monroe, S. S., Koonin, E. V., Stine, S. E. & Glass, R. I. 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 (1993).
19. Pantin-Jackwood, M., Todd, D. & Koci, M. D. Avian astroviruses. In *Astrovirus Research: Essential Ideas, Everyday Impacts, Future Directions* (ed. Schultz-Cherry, S.) 151–180, [https://doi.org/10.1007/978-1-4614-4735-1\\_9](https://doi.org/10.1007/978-1-4614-4735-1_9) (Springer New York, 2013).
20. Fuentes, C., Bosch, A., Pintó, R. M. & Guix, S. Identification of human astrovirus genome-linked protein (VPg) essential for viral infectivity. *J. Virol.* **86**, 10070–10078 (2012).
21. Goodfellow, I. The genome-linked protein VPg of vertebrate viruses - A multifaceted protein. *Current Opinion in Virology* **1**, 355–362 (2011).
22. Imada, T. *et al.* Avian Nephritis Virus (ANV) as a new member of the family *Astroviridae* and construction of infectious ANV cDNA. *Journal of virology* **74**, 8487–8493 (2000).
23. Todd, D. *et al.* Identification of chicken enterovirus-like viruses, duck hepatitis virus type 2 and duck hepatitis virus type 3 as astroviruses. *Avian Pathol.* **38**, 21–29 (2009).
24. Guix, S., Bosch, A. & Pintó, R. M. Astrovirus taxonomy. in *Astrovirus Research* 97–118, [https://doi.org/10.1007/978-1-4614-4735-1\\_6](https://doi.org/10.1007/978-1-4614-4735-1_6) (Springer New York, 2012).
25. Boujon, C. L. *et al.* Indication of cross-species transmission of astrovirus associated with encephalitis in sheep and cattle. *Emerg. Infect. Dis.* **23**, 1604–1608 (2017).
26. Fuller, T. *et al.* 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 (2012).
27. Mendenhall, I. H. *et al.* Detection of a novel astrovirus from a black-naped monarch (*Hypothymis azurea*) in Cambodia. *Virol. J.* **12**, 182 (2015).
28. Thiollay, J.-M. Avian diversity and distribution in French Guiana: patterns across a large forest landscape. *J. Trop. Ecol.* **18**, 471–498 (2002).
29. Brûlé, S. & Touroult, J. Insects of French Guiana: a baseline for diversity and taxonomic effort. *Zookeys* 111–30, <https://doi.org/10.3897/zookeys.434.7582> (2014).
30. Simmonds, P. *et al.* Consensus statement: Virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* **15**, 161. <https://doi.org/10.1038/nrmicro.2016.177> (2017).
31. Zhang, Y., Wang, F., Liu, N., Yang, L. & Zhang, D. Complete genome sequence of a novel avastrovirus in goose. *Arch. Virol.* **162**, 2135–2139 (2017).
32. Liu, N., Wang, F. & Zhang, D. Complete sequence of a novel duck astrovirus. *Arch. Virol.* **159**, 2823–2827 (2014).
33. Liu, N., Jiang, M., Dong, Y., Wang, X. & Zhang, D. Genetic characterization of a novel group of avastroviruses in geese. *Transbound. Emerg. Dis.* 1–6, <https://doi.org/10.1111/tbed.12873> (2018).
34. Fu, Y. *et al.* Complete sequence of a duck astrovirus associated with fatal hepatitis in ducklings. *J. Gen. Virol.* **90**, 1104–1108 (2009).



35. Méndez, E., Murillo, A., Velázquez, R., Burnham, A. & Arias, C. F. Replication cycle of astroviruses. In *Astrovirus Research* 19–45, [https://doi.org/10.1007/978-1-4614-4735-1\\_2](https://doi.org/10.1007/978-1-4614-4735-1_2) (Springer New York, 2012).
36. Pankovics, P., Boros, Á., Kiss, T., Delwart, E. & Reuter, G. Detection of a mammalian-like astrovirus in bird, European roller (*Coracias garrulus*). *Infect. Genet. Evol.* **34**, 114–121 (2015).
37. Wolfaardt, M., Kiulia, N. M., Mwenda, J. M. & Taylor, M. B. Evidence of a recombinant wild-type human astrovirus strain from a Kenyan child with gastroenteritis. *J. Clin. Microbiol.* **49**, 728–31 (2011).
38. Ng, T. F. F. *et al.* Feline fecal virome reveals novel and prevalent enteric viruses. *Vet. Microbiol.* **171**, 102–11 (2014).
39. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644 (2011).
40. Huson, D. H. *et al.* MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* **12**, e1004957 (2016).
41. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
42. Kosugi, S., Hasebe, M., Tomita, M. & Yanagawa, H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. USA* **106**, 10171–10176 (2009).
43. Prilusky, J. *et al.* FoldIndex(C): a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435–3438 (2005).
44. Wolf, E., Kim, P. S. & Berger, B. MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci.* **6**, 1179–1189 (1997).
45. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
46. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic acids symposium series* **41**, 95–98 ([London]: Information Retrieval Ltd., c1979-c2000, 1999).
47. Martin, D. & Rybicki, E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562–563 (2000).
48. Darriba, D., Taboada, G. L. & Doallo, R. jModelTest 2: more models, new heuristics and high-performance computing. *Nat. Methods* **9**, 772 (2012).
49. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
50. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
51. McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41**, W597–W600 (2013).

## Acknowledgements

The authors thank to Borja Milá, Michaël Moens and Álvaro Veiga for invaluable help during fieldwork, and the staff of the Nouragues Research Field Station for their support. Special thanks to Alexandra Bialonski and Heike Baum for their assistance with next-generation sequencing analyses. This study was funded by the Spanish National Research Agency (projects CGL2013-41642-P/BOS and CGL2017-82117-P, co-funded by European Regional Development Fund) and the Investissement d'Avenir grants of the ANR (CEBA:ANR-10-LABX-25-01). DAT was supported by a PhD fellowship funded by UCM (CT27/16-CT28/16).

## Author Contributions

E.G.L., A.D., J.P.T. and L.B. conceived and designed the research study; D.A.T., J.P.T. and L.B. provided the avian samples; I.F.C. and D.A.T. performed the experiments. I.F.C., D.A.T., E.G.L., A.D., J.P.T., J.S.C. D.C. and L.B. analyzed the data. I.F.C., D.A.T., E.G.L., A.D., J.P.T., D.C. and L.B. wrote the paper. All authors read and approved the final version of the paper.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-45889-3>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019