A distinct lineage of influenza A virus from bats

Suxiang Tong[a,1], Yan Li[a], Pierre Rivailler[b], Christina Conrardy[a], Danilo A. Alvarez Castillo[c], Li-Mei Chen[b], Sergio Recuenco[d], James A. Ellison[d], Charles T. Davis[b], Ian A. York[b], Amy S. Turmelle[d], David Moran[c], Shannon Rogers[a], Mang Shi[a], Ying Tao[a], Michael R. Weil[e], Kevin Tang[f], Lori A. Rowe[f], Scott Sammons[f], Xiyan Xu[b], Michael Frace[f], Kim A. Lindblade[g], Nancy J. Cox[b], Larry J. Anderson[a], Charles E. Rupprecht[d,1], and Ruben O. Donis[b,1]

[a]Division of Viral Diseases, [b]Influenza Division, [d]Division of High Consequence Pathogens and Pathology, and [f]Division of Scientific Resources, Centers for Disease Control and Prevention, Atlanta, GA 30333; [e]Emory University, Atlanta, GA 30322; [c]Center for Health Studies, Universidad del Valle de Guatemala, 01015, Guatemala City, Guatemala; and [g]International Emerging Infections Program, Centers for Disease Control and Prevention Regional Office for Central America and Panama, 01015, Guatemala City, Guatemala

Edited by Robert A. Lamb, Northwestern University, Evanston, IL, and approved January 17, 2012 (received for review October 7, 2011)

Influenza A virus reservoirs in animals have provided novel genetic elements leading to the emergence of global pandemics in humans. Most influenza A viruses circulate in waterfowl, but those that infect mammalian hosts are thought to pose the greatest risk for zoonotic spread to humans and the generation of pandemic or panzootic viruses. We have identified an influenza A virus from little yellow-shouldered bats captured at two locations in Guatemala. It is significantly divergent from known influenza A viruses. The HA of the bat virus was estimated to have diverged at roughly the same time as the known subtypes of HA and was designated as H17. The neuraminidase (NA) gene is highly divergent from all known influenza NAs, and the internal genes from the bat virus diverged from those of known influenza A viruses before the estimated divergence of the known influenza A internal gene lineages. Attempts to propagate this virus in cell cultures and chicken embryos were unsuccessful, suggesting distinct requirements compared with known influenza viruses. Despite its divergence from known influenza A viruses, the bat virus is compatible for genetic exchange with human influenza viruses in human cells, suggesting the potential capability for reassortment and contributions to new pandemic or panzootic influenza A viruses.

evolution | host range | orthomyxoviridae | Chiroptera | Central America

Emerging infectious diseases and pandemics in humans often originate from pathogens transmitted from nonhuman animal reservoirs (1). The pandemics of severe acute respiratory syndrome, HIV, and 2009 H1N1 influenza illustrate the dramatic impact of viral host-switching on public health and the global economy (2–6). Early prediction, detection, characterization, and risk assessment of viruses in their animal hosts, before they spread into the human population, are critical to protect public health (4, 7, 8). Bats (order Chiroptera) are of particular interest, because they comprise nearly 1,200 species worldwide, accounting for approximately one-fourth of all mammal species, and their global distribution, abundance, ability to fly and migrate over large distances, and sociality favor the acquisition and spread of viruses (9). In recent history, bats have been sources of multiple pathogenic viruses for humans and domestic animals, including coronaviruses, filoviruses, henipaviruses, and lyssaviruses.

Consensus-degenerate RT-PCR/PCR primers designed to identify the most conserved genomic sequence regions of virus families, subfamilies, or genera is a powerful technology for the detection of novel pathogens (10) and has been used to identify a number of new viruses (11–13). We developed a pan-influenza virus RT-PCR that detects influenza virus signature sequences within polymerase basic protein 1 (PB1), the catalytic subunit of the RNA-dependent RNA polymerase, which is one of the most conserved proteins across RNA viruses (14). This approach allowed us to search for novel influenza viruses from fruit bats in Guatemala during 2009–2010.

Results

Identification of an Influenza A Virus in Bats. A total of 316 bats from 21 different species were captured from eight locations in southern Guatemala in two consecutive years (180 bats in May, 2009, and 136 bats in September, 2010) (Fig. 1 and *SI Appendix*, Table S1). Three of 316 bat rectal swabs were positive by the pan-influenza RT-PCR assay. All three were collected from little yellow-shouldered bats (*Sturnira lilium,* family Phyllostomidae), a frugivorous bat that is abundant throughout Central and South America. Two of the positive samples were from two of 15 little yellow-shouldered bats (bats GU09-153 and GU09-164) captured during 2009 at El Jobo, Guatemala, and the third was from one of 14 little yellow-shouldered bats (bat GU10-060) captured during 2010 from Agüero, Guatemala, located ~50 km from El Jobo (Fig. 1 and *SI Appendix*, Table S1). Each of the three samples was estimated by quantitative real-time RT-PCR (qRT-PCR) to have ~10^5–10^6 viral genome copies per 100 μL of rectal swab suspension. Of the other available specimens (i.e., liver, intestine, lung, and kidney tissues and an oral swab from bat GU09-164), the four tissue specimens were positive, but the oral swab specimen was negative. These data support an infectious process rather than transit of ingested infected material through the digestive tract as the source of viral RNA, particularly because this bat species does not feed on other vertebrates. The nucleotide sequences of the 250-bp diagnostic PB1 amplicons from three different bat rectal samples were very similar to each other and by BLASTn search were most closely related to influenza A virus PB1 genes.

Viral Genome Sequencing. The viral genomic sequences present in rectal swabs from bats GU09-153 and GU09-164 were determined by two next-generation sequencing approaches (Illumina GAIIx and 454 pyrosequencing, respectively) and were verified by the Sanger chain-termination method. The viral genomic sequences from bat GU10-060 were determined using direct PCR sequencing by the Sanger method. The 5′ and 3′ termini of all genome segments from bats GU09-164 and GU10-

Author contributions: S.T., C.E.R., and R.O.D. designed research; S.T., Y.L., P.R., C.C., D.A.A.C., L.-M.C., S. Recuenco, J.A.E., C.T.D., I.A.Y., A.S.T., D.M., M.S., Y.T., M.R.W., K.T., L.A.R., S.S., and X.X. performed research; S.T., Y.L., P.R., L.-M.C., I.A.Y., S. Rogers, M.S., C.E.R., and R.O.D. contributed new reagents/analytic tools; S.T., Y.L., P.R., C.C., D.A.A.C., L.-M.C., S. Recuenco, J.A.E., C.T.D., I.A.Y., A.S.T., D.M., M.S., Y.T., M.R.W., K.T., L.A.R., S.S., X.X., M.F., K.A.L., N.J.C., L.J.A., C.E.R., and R.O.D. analyzed data; and S.T., Y.L., P.R., L.-M.C., C.T.D., I.A.Y., M.S., M.R.W., L.J.A., C.E.R., and R.O.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. CY103873, CY103874, CY103875, CY103876, CY103877, CY103878, CY103879, CY103880, CY103881, CY103882, CY103883, CY103884, CY103885, CY103886, CY103887, CY103888, CY103889, CY103890, CY103891, CY103892, CY103893, CY103894, CY103895, and CY103896).

[1]To whom correspondence may be addressed. E-mail: sot1@cdc.gov, cyr5@cdc.gov, or rvd6@cdc.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1116200109/-/DCSupplemental.

MICROBIOLOGY

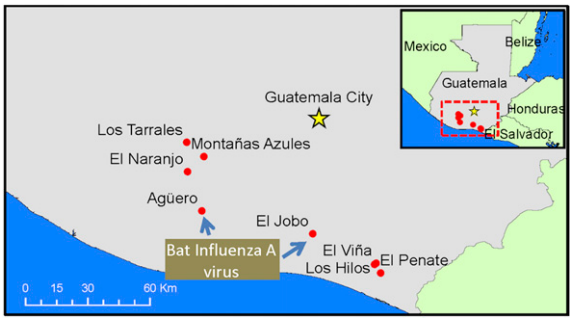

Fig. 1. Locations of bat-sampling sites in Guatemala. Geographic locations in Guatemala in which bats were captured are indicated with red dots. Locations of pan-Flu RT-PCR–positive bats are denoted by arrows. See *SI Appendix*, Table S1 for additional information.

060 were sequenced from 5′ and 3′ RACE amplicons. The genomes from bats GU09-153 and GU09-164 were nearly identical to each other (99.99% nucleotide identity) and were more distantly related to the genome from bat GU10-060, sampled the following year (96.1% nucleotide identity) (*SI Appendix*, Table S2). Nucleotide and protein sequence similarity searches (BLASTn, tBLASTx, BLASTp) of genetic sequence databases using these eight bat virus gene segments identified influenza type A sequences as the most significant matches. Each of the genomic segments of the bat viruses could be aligned to known influenza A virus genomic segments with few or no gaps, with the exception of the homolog of neuraminidase (NA), which required 16 gaps for alignment with known influenza A NA segments (Fig. 2 and *SI Appendix*, Fig. S1 and Table S3). Key structural features of the bat virus genomes are consistent with a type A influenza virus in that (*i*) they comprise eight RNA segments (based on next-generation deep sequencing and RACE PCR sequencing), six of which are monocistronic [polymerase basic protein 2, PB1, polymerase (PA), hemagglutinin (HA), NA, and nucleoprotein (NP)], whereas two (matrix (M) and nonstructural (NS)) contain an additional reading frame for expression of the M2 and the NS2/NEP protein homologs, respectively, via spliced mRNAs (*SI Appendix*, Table S3) (15, 16); (*ii*) the protein-coding regions are in the cRNA polarity; (*iii*) the coding region of each genomic RNA segment (vRNA) is flanked by noncoding regions (NCR) at the 3′ and 5′ ends that include sequences conserved among most segments, followed by segment-specific sequences (*SI Appendix*, Fig. S2) (17); (*iv*) the 12 nucleotides at the 3′ end and the 13 nucleotides at the 5′ end of each genome segment are nearly identical to those of other known influenza A viruses; there is one base difference at position 5 of the 3′ NCR region (*SI Appendix*, Fig. S2) and one base difference in the 5′ NCR of GU10-060 M (G) and NP (C) genes; (*v*) the terminal NCRs of each segment comprise complementary sequences predicted to form panhandle-like structures with putative transcription promoter function (*SI Appendix*, Fig. S2) (17, 18); (*vi*) the uridine-tract polyadenylation signals are located 16 nucleotides from the 5′ end of the vRNA (19) (*SI Appendix*, Fig. S2); and (*vii*) conserved sequence motifs in the bat virus genome and proteome are shared with known influenza A viruses (*SI Appendix*, Table S4 and Movies S1–S6). These findings indicated that the bat virus genomes are most closely related structurally to influenza A viruses. Here we refer to these viruses as "A/little yellow-shouldered bat/Guatemala/164/2009" (which we describe in this manuscript as the reference virus, abbreviated as "A/bat/Guat/09"), "A/little yellow-shouldered bat/Guatemala/153/2009," and "A/little yellow-shouldered bat/Guatemala/060/2010" (abbreviated to "A/bat/Guat/10"), according to standard influenza nomenclature (20).

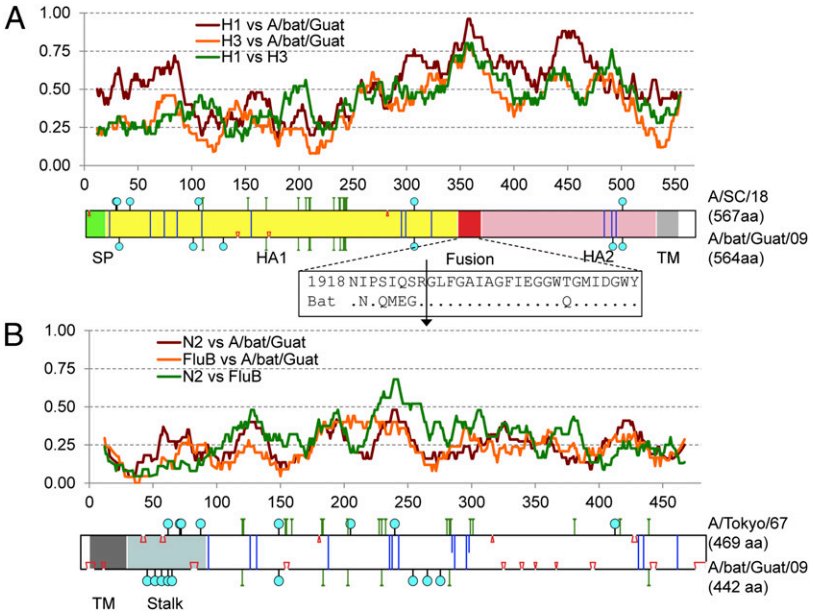

Fig. 2. Structural and functional elements of the A/bat/Guat/09 HA and NA-like proteins. HA and NA proteins from A/bat/Guat/09 and those of known influenza A viruses were aligned and compared. (*A*, *Upper*) Amino acid similarity plot of HA from A/South Carolina/1918(H1N1), A/Panama/2007/1999(H3N2), and A/bat/Guat/09 as shown in the legend. (*A*, *Lower*) Schematic diagram of aligned HA from influenza A/South Carolina/1918(H1N1) (upper half of the horizontal bar) and A/bat/Guat/09 (lower half), showing major structural sections: signal peptide (SP) in solid green, HA1 in yellow, fusion peptide in red, HA2 in pink, and transmembrane domain (TM) in gray; putative glycosylation sites (cyan lollipops); conserved disulphide bonds (blue lines); and residues involved in sialic acid binding (green pin). Receptor-binding residues (green pins) are shown for A/bat/Guat/09 if the residues are found in known influenza A viruses rather than being unique to A/bat/Guat/09. Gaps in the alignment are indicated with red triangles. (*B*, *Upper*) Amino acid similarity plot of NA from A/bat/Guat/09, A/Tokyo/3/1967(H2N2) and B/Lee/40 as shown in the legend. (*B*, *Lower*) Schematic diagram of aligned NA from A/Tokyo/3/1967(H2N2) and A/bat/Guat/09. Labels are as in *A*, except that green pins represent residues involved in sialidase catalytic activity.

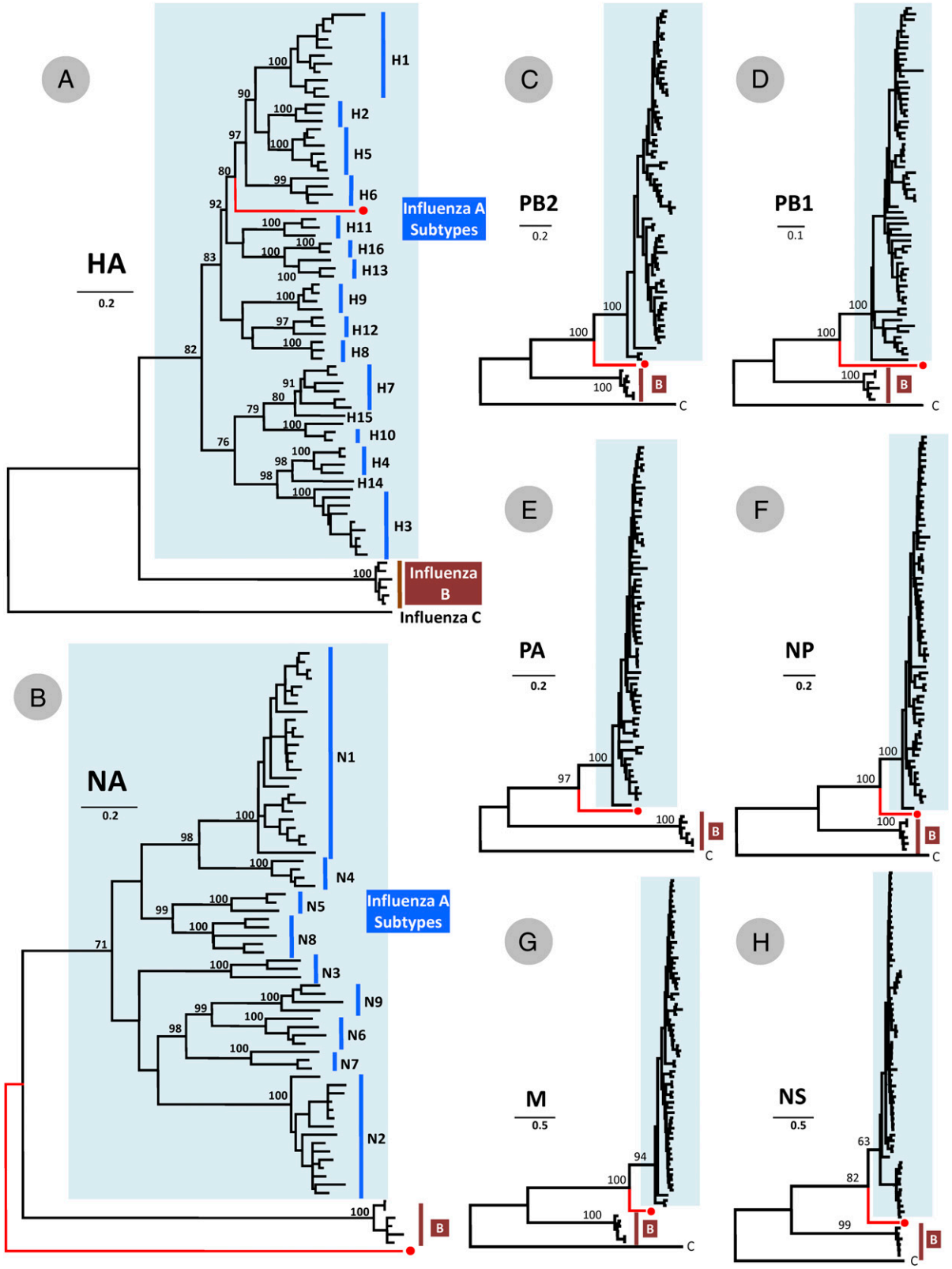

Fig. 3. Evolution of the A/bat/Guat/09 genome. Phylogenetic trees of the eight gene segments of the bat virus genomes (represented by a red dot and red branch) were inferred by the maximum-likelihood method using the dataset described in *SI Appendix*, Table S7. Bootstrap values are shown above branches to the left of major nodes denoting influenza A (shaded in light blue), influenza B (indicated in purple), and A/bat/Guat/09 (indicated by a red dot) divergences. (Scale bars, number of expected nucleotide substitutions per site.) Trees are shown for (*A*) HA, (*B*) NA, (*C*) PB2, (*D*) PB1, (*E*) PA, (*F*) NP, (*G*) M, and (*H*) NS segments.

HA and NA Gene Products. The HA and NA proteins of influenza viruses serve critical biological functions and also determine the subtype of each type A virus. On average, the A/bat/Guat/09 HA has 45% amino acid sequence identity to HAs from all known influenza A subtypes, similar to the 49% mean pairwise identity observed among the 16 known HA subtypes (*SI Appendix*, Table S5*A*). Sequence motifs conserved among influenza A virus HAs, such as the sialic acid receptor-binding site, can be identified in the HA-like protein of the bat virus (Fig. 2 and *SI Appendix*, Fig. S1*A*), although changes at positions that modulate the specificity for the galactose–sialic acid linkage suggest a distinct ligand preference (*SI Appendix*, Table S6*A*). In contrast, the protein encoded by the NA gene shows extensive divergence from known influenza virus NAs in amino acid residues comprising the canonical sialic acid-binding/catalytic site (*SI Appendix*, Table S6*B*). In addition, the A/bat/Guat/09 NA-like protein has only 24% amino acid sequence identity to other influenza A NA subtypes (*SI Appendix*, Table S5*B*); this similarity is lower than the similarity between NAs from influenza A and influenza B. Despite these differences, the NA homolog retains major overall structural components typical of influenza A and B NA proteins (Fig. 2 and *SI Appendix*, Fig. S1*B*).

Phylogenetic Analysis. The alignment and phylogenetic analysis of the bat viruses included a representative sample comprising 66 known influenza A viruses, six influenza B viruses, and one influenza C virus. Phylogenetic analyses of the eight gene segments using maximum-likelihood methods (Fig. 3 *A*–*H*) and Bayesian inferences (*SI Appendix*, Fig. S3 *A*–*H*) showed three distinctive types of relationships between genes of A/bat/Guat/09 and the known influenza viruses. Analysis of influenza A and B HA genes suggests that the A/bat/Guat/09 HA is more closely related to the Group 1 HAs (subtypes H1, 2, 5, 6, 8, 9, 11, 12, 13, and 16) than to the Group 2 HAs (H3, 4, 7, 10, 14, and 15) (Fig. 3*A* and *SI Appendix*, Fig. S3*A* and Table S5*A*) (21), sharing ancestry with a monophyletic clade containing H1, H2, H5, and H6 subtypes. Therefore, the A/bat/Guat/09 HA gene is likely to be representative of a monophyletic group that merits a provisional subtype classification of H17, pending definitive taxonomic classification (20, 22). Analysis of the A/bat/Guat/09 NA indicated that it was highly divergent from both influenza A and B NA genes sharing an older ancestral relationship to known influenza viruses. (Fig. 3*B* and *SI Appendix*, Fig. S3*B* and Table S5*B*). A subtype nomenclature for this gene will require elucidation of its evolution within the Orthomyxoviridae. In contrast, the six so-called "internal" genes (PB2, PB1, PA, NP, M, and NS) of A/bat/Guat/09 were clustered outside the influenza A and B gene branches. Their positions in the phylogenetic tree were between the influenza A and B split but were related more closely to the type A viral genes (Fig. 3 *C*–*H* and *SI Appendix*, Fig. S3 *C*–*H*).

Minigenome Reporter Assays. Attempts to propagate A/bat/Guat/09 from RT-PCR–positive samples in 11-d-old embryonated chicken eggs and in several mammalian cells (including bat cells) have been unsuccessful to date. To test whether the A/bat/Guat/09 polymerase complex (PB2, PB1, PA, and NP) proteins function within human cells, we used a minigenome reporter replicon system driven by transient expression from transfected plasmid DNA (23). In this assay, a polymerase complex (PB2, PB1, PA, and NP) from one virus is cotransfected with a reporter plasmid in which luciferase expression is driven by NCR regions from the same or a different virus. The amount of luciferase expression indicates the degree of compatibility of the NCR with the polymerase complex and with cellular components required for efficient expression. Efficient expression of the luciferase reporter flanked by NCR of the A/bat/Guat/09 NS gene was observed when the homologous PB2, PB1, PA, and NP proteins (the minimal functional polymerase complex) were cotransfected into human lung cells (A549 cell line) (Fig. 4), indicating that human cells are compatible with these functions of the bat virus. Furthermore, A/bat/Guat/09 polymerase complex proteins recognized a reporter minigenome from a human influenza A virus (A/WSN/33) with high efficiency (Fig. 4), demonstrating that the bat polymerase complex is compatible with human virus NCRs.

Discussion

Wild populations of waterfowl are considered hosts to the most diverse influenza A viruses, including 16 distinct alleles of the HA gene and nine alleles of the NA gene. Sporadically, these viruses infect mammals, and in rare instances these events lead to sustained transmission in the new mammalian host. Human, swine, and equine influenza viruses are thought to have emerged directly or indirectly from this avian reservoir. The influenza viruses of domesticated animals, particularly those of swine, are thought to pose the greatest risk for zoonotic spread to humans and the generation of pandemic or panzootic viruses. Here we demonstrate that little yellow-shouldered bats in Central America also are hosts of an influenza A virus and constitute a potential sylvatic mammalian reservoir of influenza. This virus is significantly divergent from known influenza A viruses, such that its HA can be classified as a separate subtype, designated as "H17" in keeping with current influenza A virus subtyping classification. Although the NA phylogeny and sequence divergence from known influenza viruses indicates a more ancient relationship, these comparisons, together with a high degree of structural homology to other influenza viruses, suggests the need for further refinement of this NA classification. Additional comparison with members of the Orthomyxoviridae family, as well as other future sequences from bat viruses, may help resolve this issue.

The genomic sequence of this virus sheds light on the ancestry and evolution of influenza A viruses. Phylogenetic analysis showed that the six internal genes of A/bat/Guat/09 diverged from those of known influenza A viruses after the split of influenza A and B but before the divergence of the internal genes of known influenza A viruses (Fig. 3 *C*–*H* and *SI Appendix*, Fig. S3 *C*–*H*). The HA, on the other hand, diverged after the split of influenza A

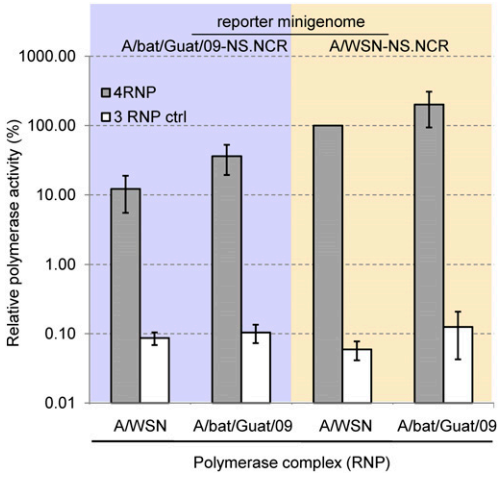

Fig. 4. RNA-dependent RNA polymerase activity of A/bat/Guat/09 ribonucleoprotein (RNP) complex proteins (PB2, PB1, PA, and NP). A549 human lung cells were transfected with pPol1-A/bat-NS.NCR-Renilla or pPol1-A/WSN-NS.NCR-Renilla and pSV40-Luc reporter plasmids together with plasmids expressing PB2, PB1, PA, and NP from either A/WSN/33 or A/bat/Guat/09 viruses (4RNP; solid bars) or without the PB1 expression plasmid (3 RNP ctrl; open bars). Values shown represent the activities of each RNP and reporter relative to that of WSN virus (100%). Error bars indicate 95% confidence intervals. Experiments were performed three times independently.

HA subtypes into Group 1 and 2. Although different in their topologies, the HA and the internal genes of A/bat/Guat/09 have distinct but relatively close evolutionary relationships to known influenza genes, indicating that these genes could have diverged from ancestral influenza viruses at approximately the same time. The topological differences between the HA and internal genes of A/bat/Guat/09 could be explained by the greater diversity of HA gene pools compared with internal genes. Immunologic selective pressure on virion surface genes has resulted in multiallelic populations, whereas the population of internal genes remained relatively homogenous, perhaps because of the "hitchhiking" of internal genes with surface genes during selective sweeps (24, 25). The phylogenetic location of the internal genes of A/bat/Guat/09 is consistent with this hypothesis and further suggests a significant loss of diversity among internal genes during the process of evolution. It is possible that the internal genes of A/bat/Guat/09 are representatives of otherwise extinct gene lineages that circulated more widely in previous centuries.

Unlike HA and internal genes, the NA gene of A/bat/Guat/09 is extraordinarily divergent from known influenza viruses. The origin of this gene is difficult to determine because of the long branch length and absence of closely related genes. One possibility is that the A/bat/Guat/09 NA descends from a much older ancestor than the other genes, as the result of reassortment with yet-to-be-discovered or extinct highly divergent bat influenza viruses. Alternatively, the NA gene could have diverged with the HA and internal genes but evolved at a faster rate, perhaps under different selection pressures, in agreement with the higher synonymous/nonsynonymous substitution (dN/dS) value of the 2009 and 2010 bat influenza NA sequences relative to other genes (*SI Appendix*, Table S2). A better understanding of the evolutionary history of the bat influenza NA gene would be facilitated by identifying additional related viruses in bats.

Based on minigenome reporter assays, the A/bat/Guat/09 genome replication complex was able to function in human cells, because the combination of PB2, PB1, PA, and NP was able to drive luciferase expression efficiently from the bat NCR in human cells (A549 cells). Moreover, the A/bat/Guat/09 replication complex was capable of efficiently driving expression from a human virus NCR, demonstrating that the bat polymerase complex is compatible with human virus NCRs and suggesting that reassortants of bat and human viruses may not be incompatible.

The extent to which A/bat/Guat/09-like viruses circulate in bat populations worldwide is unknown. Serologic studies are ongoing to establish the seroprevalence in bats from Central America and other regions. The genomic detection of this virus in 1% of the total bat population tested (equivalent to 13% and 7% of the little yellow-shouldered bat populations tested at El Jobo and Agüero, respectively) is consistent with the frequency of influenza virus detection in wild bird surveillance projects (26). Virus detection in animals sampled at two distinct locations and in different years indicates that its presence is unlikely to be an incidental interspecies transfer, and the amount of variation between the three viruses detected is not unusual for RNA viruses circulating within a significant population.

Influenza type A viruses have established sustained circulation in multiple animal populations, including birds as well as some mammalian species. Bats may now be added to the list of mammalian hosts of influenza A viruses. Considering that swine, horses, and dogs may serve as bridging hosts of other viruses associated with bats (e.g., filoviruses, paramyxoviruses, and others), their potential permissiveness for bat influenza viruses should be investigated. Clearly, the identification of influenza A viruses in bats expands the repertoire of likely reservoirs of influenza viruses and raises further questions: How are influenza viruses maintained in bat populations? What are the public health and agricultural ramifications attributable to this reservoir of influenza? Are there yet-to-be-discovered reservoirs of ancestral influenza viruses in other parts of the animal kingdom?

Materials and Methods

Sample Collection. Guatemala was chosen as one major comparative New World study location in Central America as part of the Centers for Disease Control and Prevention (CDC) Global Disease Detection Program. Detailed information on bat capture and handling are available in *SI Appendix, Supplementary Materials and Methods*. Project protocols for animal capture and use were approved by the CDC Animal Care and Use Committee and the Ethics and Animal Care and Use Committee of the Universidad del Valle de Guatemala (Guatemala City, Guatemala). Representative samples at each site consisted of adults and juveniles of both sexes. After bats were killed, a complete necropsy was performed on all bats in compliance with approved field protocols. Samples included blood, major organs (liver, intestine, lung, and kidney), and rectal and oral swabs. Total nucleic acids (TNA) were extracted from 200 μL of PBS suspension of each swab by using the QIAamp MinElute Virus Spin kit (Qiagen) and then were stored at −80 °C.

Pan-Influenza RT-PCR. TNA extracted from the rectal swabs (n = 316) were screened for the presence of influenza virus RNA using pan-influenza (pan-Flu) RT-PCR with consensus degenerate primers targeted at a conserved region of PB1 in the influenza virus genome (primer sequences are available upon request). Standard precautions were taken to avoid cross-contamination of samples before and after RNA extraction and amplification. Each of the positive results was repeated and confirmed from different TNA aliquots of the original bat rectal swab eluate. RT-PCR methods are described in detail in the *SI Appendix, Supplementary Materials and Methods*. Purified DNA amplicons (both strands) were sequenced with the RT-PCR primers on an ABI Prism 3130 automated capillary sequencer (Applied Biosystems).

Complete Genome Sequencing. We used two sequencing strategies: high-throughput next-generation sequencing and amplicon-based Sanger sequencing. Sequencing methods are described in detail in *SI Appendix, Supplementary Materials and Methods*. Briefly, the pan-Flu RT-PCR–positive rectal samples were subjected to high-throughput sequencing by Illumina GAIIx (Illumina) or 454 GS-FLX pyrosequencing (Roche). To increase the reliability of the sequence data from high-throughput sequencing, amplicons of the eight viral segments from two samples (GU09-153 and GU09-164) also were analyzed by Sanger sequencing.

One sample (GU10-060) was analyzed only by Sanger sequencing. Amplicons from the eight viral segments (ranging in size from 800 bp to 2.3 kb) were subcloned into the pCR-XL-TOPO vector (Invitrogen). The 3′ end and 5′ end sequences of each segment from two bats were determined using the 5′/3′ RACE kit (Roche) according to the manufacturer's instructions. Automated sequencing was performed on an ABI Prism 3130 automated sequencer using cycle sequencing dye terminator chemistry (Applied Biosystems) according to the manufacturer's instructions. Sequences were assembled and analyzed using Sequencher software (Genecodes). Consensus gene sequences were compared with those from the high-throughput next-generation sequencing methods.

qRT-PCR Assay. The primers and TaqMan probe for a qRT-PCR assay were designed targeting the NP of the novel bat virus. The qRT-PCR assays were performed on the MX3005P (Agilent Technologies) using the SuperScript III Platinum One-step Quantitative RT-PCR system (Invitrogen). Detailed methods are available in the *SI Appendix, Supplementary Materials and Methods*. Negative controls without template were run on each plate using nuclease-free water. All standards, samples, and controls were carried out in triplicate. The standard template was constructed from the cloned NP segment where the primers and probe are located. Unknown sample concentrations were calculated from a standard curve of templates of known concentration analyzed on the same plate. A sample was considered positive when more than one of three reactions was positive.

Sequence Analysis and Phylogeny Inference. Detailed descriptions of the procedures used for sequence analysis and phylogenetic inference are provided in *SI Appendix, Supplementary Materials and Methods*. Briefly, sequences of influenza A and B viruses were downloaded from the Global Initiative on Sharing Avian Influenza Data database (http://platform.gisaid.org). Large datasets containing all publicly available sequences were used to generate consensus sequences and to identify bat residues never/rarely seen in known influenza sequences. Phylogenetic analyses were based on a smaller representative dataset containing 66 influenza A viruses, six influenza B

viruses, and one influenza C virus (*SI Appendix*, Table S7). Amino acid sequences were aligned using MAFFT (27), and alignment of each gene was mapped manually onto the nucleotide sequences. Phylogenetic trees of the eight gene segments of the bat virus genome were inferred by maximum likelihood using PAUP* v4.0b10 (28).

Minigenome Reporter Assays. A549 human lung carcinoma cells (American Type Culture Collection) were cotransfected with pPol1-NS-Renilla encoding an influenza A reporter minigenome under transcriptional control of the human RNA polymerase I, pSV-Luc encoding firefly luciferase under SV40 virus RNA polymerase II promoter control, and four plasmids expressing viral PB2, PB1, PA, and NP from the strain of interest (23). Twenty-four hours after transfection, cell lysates were harvested and further diluted to perform a dual luciferase assay according to the manufacturer's protocol (Promega). The influenza polymerase catalytic activity derived from the Renilla luciferase plasmid (pPol1-NS-Renilla) was corrected to account for well-to-well differences in transfection efficiency using the firefly luciferase activity values from the pSV-Luc plasmid. More details are is available in *SI Appendix, Supplementary Materials and Methods*.

ACKNOWLEDGMENTS. We thank Luis Escobar, Alejandra Estevez, María Reneé López, Ramon Medrano, Maria E. Morales, and María Luisa Muller from Center for Health Studies, Universidad del Valle de Guatemala; Julio Martinez from the Guatemala Ministry of Agriculture-Animal Health Department; and Rafael Ciraiz from Guatemala Ministry of Public Health and Social Assistance for excellent technical and logistical assistance. We thank Amanda Balish, Thomas Rowe, Wendy Sessions, James Stevens, Taronna Maines, and Angela Yang for preliminary studies; Yunho Jang for technical assistance; and Peter Palese and Erich Hoffmann for plasmids. This study was supported in part by Technical Support Corps funds from the Global Disease Detection Program of the Centers for Disease Control and Prevention, Atlanta, GA.

1. Reid AH, Taubenberger JK (2003) The origin of the 1918 pandemic influenza virus: A continuing enigma. *J Gen Virol* 84:2285–2292.
2. Marx PA, Apetrei C, Drucker E (2004) AIDS as a zoonosis? Confusion over the origin of the virus and the origin of the epidemics. *J Med Primatol* 33:220–226.
3. Gao F, et al. (1999) Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* 397:436–441.
4. Murphy FA (2008) Emerging zoonoses: The challenge for public health and biodefense. *Prev Vet Med* 86:216–223.
5. Li W, et al. (2005) Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310: 676–679.
6. Garten RJ, et al. (2009) Antigenic and genetic characteristics of swine-origin 2009 A (H1N1) influenza viruses circulating in humans. *Science* 325:197–201.
7. Barrette RW, et al. (2009) Discovery of swine as a host for the Reston ebolavirus. *Science* 325:204–206.
8. Kuiken T, et al. (2005) Public health. Pathogen surveillance in animals. *Science* 309: 1680–1681.
9. Turmelle AS, Olival KJ (2009) Correlates of viral richness in bats (order Chiroptera). *EcoHealth* 6:522–539.
10. Rose TM, Henikoff JG, Henikoff S (2003) CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Res* 31:3763–3766.
11. Simons JN, et al. (1995) Isolation of novel virus-like sequences associated with human hepatitis. *Nat Med* 1:564–569.
12. Ksiazek TG, et al.; SARS Working Group (2003) A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 348:1953–1966.
13. Tong S, et al. (2009) Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerg Infect Dis* 15:482–485.
14. Poch O, Sauvaget I, Delarue M, Tordo N (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J* 8:3867–3874.
15. Lamb RA, Choppin PW, Chanock RM, Lai CJ (1980) Mapping of the two overlapping genes for polypeptides NS1 and NS2 on RNA segment 8 of influenza virus genome. *Proc Natl Acad Sci USA* 77:1857–1861.
16. Lamb RA, Lai CJ, Choppin PW (1981) Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: Colinear and interrupted mRNAs code for overlapping proteins. *Proc Natl Acad Sci USA* 78:4170–4174.
17. Desselberger U, Racaniello VR, Zazra JJ, Palese P (1980) The 3′ and 5′-terminal sequences of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. *Gene* 8:315–328.
18. Lee YS, Seong BL (1998) Nucleotides in the panhandle structure of the influenza B virus virion RNA are involved in the specificity between influenza A and B viruses. *J Gen Virol* 79:673–681.
19. Luo GX, Luytjes W, Enami M, Palese P (1991) The polyadenylation signal of influenza virus RNA involves a stretch of uridines followed by the RNA duplex of the panhandle structure. *J Virol* 65:2861–2867.
20. WHO (1980) A revision of the system of nomenclature for influenza viruses: A WHO memorandum. *Bull World Health Organ* 58:585–591.
21. Kovácová A, Ruttkay-Nedecký G, Haverlík IK, Janecek S (2002) Sequence similarities and evolutionary relationships of influenza virus A hemagglutinins. *Virus Genes* 24:57–63.
22. Röhm C, Zhou N, Süss J, Mackenzie J, Webster RG (1996) Characterization of a novel influenza hemagglutinin, H15: Criteria for determination of influenza A subtypes. *Virology* 217:508–516.
23. Chen LM, Davis CT, Zhou H, Cox NJ, Donis RO (2008) Genetic compatibility and virulence of reassortants derived from contemporary avian H5N1 and human H3N2 influenza A viruses. *PLoS Pathog* 4:e1000072.
24. Chen R, Holmes EC (2010) Hitchhiking and the population genetic structure of avian influenza virus. *J Mol Evol* 70:98–105.
25. Dugan VG, et al. (2008) The evolutionary genetics and emergence of avian influenza viruses in wild birds. *PLoS Pathog* 4:e1000076.
26. Munster VJ, et al. (2007) Spatial, temporal, and species variation in prevalence of influenza A viruses in wild migratory birds. *PLoS Pathog* 3:e61.
27. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30: 3059–3066.
28. Swofford DL (2001) *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods) 4.0 Beta* (Sinauer Associates, Sunderland, MA).