# Supplementary Information:

# Ranking the risk of animal-to-human spillover for newly discovered viruses

## Authors

Zoë L. Grange[1], Tracey Goldstein[1$], Christine K. Johnson[1$], Simon Anthony[1,2,3,4], Kirsten Gilardi[1], Peter Daszak[2], Kevin J. Olival[2], Tammie O'Rourke[5], Suzan Murray[6], Sarah H. Olson[7], Eri Togami[1], Gema Vidal[1], Expert Panel[8], PREDICT Consortium[9], and Jonna A.K. Mazet[1*]

## Author Affiliations

[1]One Health Institute and Karen C. Drayer Wildlife Health Center, School of Veterinary Medicine, University of California, Davis, California 95616, USA

[2]EcoHealth Alliance, New York, 1001, New York, USA

[3]Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, New York 10032, USA

[4]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York,10032, USA

[5]Metabiota, Inc. 7-1611 Bowen Road, Nanaimo, BC V9S 1G5, Canada

[6]Global Health, Smithsonian Conservation Biology Institute, Washington, DC 20008, USA

[7]Wildlife Conservation Society, New York, New York, 10460, USA

[$]Co-second authors

**Corresponding Author:** *jkmazet@ucdavis.edu One Health Institute and Karen C. Drayer Wildlife Health Center, School of Veterinary Medicine, University of California, Davis, California 95616, USA, +1 530 752 7526

**This file includes:**

**Other supplementary materials for this manuscript include the following:**

All datasets along with the R code and R package dependencies needed to fully replicate and evaluate these analyses are provided on the Open Science Framework at the following link https://osf.io/mb6qn/?view_only=f6326d48d7d941afa7af02714819a1a2 The top-level README.txt file in the directory details the file structure and metadata provided. The R code contains functions to extract information for risk factors, if applicable from static data sources. The functions are designed so that data sources can be updated in the future, as and when new information becomes available.

For future additions of viruses into the database, the SpillOver website uses the R codes, functions, datasets and methods described below as a basis to automatically provide answers. Automatic answers, were applicable are editable by the submitter and website administrators to improve accuracy.

# Supplementary Information Text

## SI Methods

### Viral Risk Ranking Framework

Detailed methodologies for each of the 31 risk factors used in the risk ranking assessment are described below. Eleven risk factors assessed by experts as having low or no contribution to spillover risk, evidenced by a *Risk Factor Influence* score of below two (see manuscript), were not included in the relative risk but methods for their classification are described for reference and tagged with 'NOT INCLUDED IN TOP 31 RISK FACTORS' in case they are useful for inclusion later in the risk ranking framework, as the scientific process reveals more information on the relative importance of risk factors for estimating spillover risk for potentially zoonotic viruses.

#### *Reference Data Sources for Risk Factors and Viruses and Viruses and Viruses*

The International Committee on Taxonomy of Viruses (ICTV) 2015 Master Species List (MSL30) version 5/24/2016 (referred to as ICTV list) was downloaded from the ICTV online database (https://talk.ictvonline.org/) and used as a reference of known virus species within 26 target viral families of concern to human health (*Anelloviridae, Arenaviridae*, *Bornaviridae*, *Bunyaviridae*, *Coronaviridae*, *Filoviridae*, *Flaviviridae*, *Hepadnaviridae, Hepeviridae*, *Orthomyxoviridae*, *Paramyxoviridae*, *Picornaviridae*, *Reoviridae*, *Retroviridae*, *Rhabdoviridae*, *Togaviridae*, *Astroviridae*, *Caliciviridae*, *Picobirnaviridae*, *Arteriviridae, Herpesviridae*, *Poxviridae*, *Parvoviridae, Adenoviridae*, *Papillomaviridae*, *Polyomaviridae*)(1). Plant virus species were removed from the list and not assessed or included in any analyses, resulting in a master ICTV list of 1,126 virus species.

A literature review of three reference virology textbooks(2-4), four human/zoonotic virus databases(5-8) and a literature search of all databases in the Web of Science using the search term: 'Virus Name' AND 'Humans' was conducted for each virus species in the ICTV list to characterize zoonotic potential and obtain information required for several risk factors, as well as to provide auto-populated data for the risk ranking application. For each virus species, we assessed its ability to infect and pathogenicity in a taxonomic group (human, terrestrial mammals, birds, other animals), duration of infection in humans, transmissibility between humans, zoonotic ability and whether the virus has caused an epidemic or pandemic in humans/animals. Further details are provided within each risk factor method described below.

Additionally, we extracted virus detection data from the NCBI nucleotide database (ftp://ftp.ncbi.nlm.nih.gov/ncbi-asn1/) on June 8th, 2018. Attributes including accession number, organism name, host species, sample description and country of detection were extracted from Genbank files using a custom python script in Anaconda Jupyter software(9) using Biopython Entrez and SeqIO(10). A list of taxon IDs was extracted from the NCBI taxonomy database for each virus species in the ICTV virus list (described above) and used to select all NCBI nucleotide records associated with a given virus species designation and its synonyms. Due to ambiguity of data sources, such as experimental infections and inaccuracies in reported location, data without references (a Pubmed ID) were removed from the database. Host species data were cleaned using accepted scientific names and taxonomy in the IUCN Red List(11), and when not available, the Catalogue of life(12). Country names/designations were corrected as per the IUCN Red List database(11) based on ISO 3166 list of countries. Reports of unknown species origin and viral sequences originating from domestic species (animals that have been domesticated to be kept as pets or food), environmental samples, cell lines, cloned sequences, laboratory isolates and/or synthetic constructs were removed from the dataset. When available, the longitude and latitude of the host species where sampled were extracted from the public resources. If not available, the central coordinates of the reported nearest village or town were used as a reference point. If location was only provided at the state district or provincial level, the geographic coordinates were classified as unknown.

## Risk Factor Methods

### ***Host Risk Factors***

### Host plasticity - No. of species / No. of orders

Host plasticity is a measure of the diversity (number of species) and breadth (number of orders) of hosts the virus is known to infect. Host species are limited to the taxonomic classes *Mammalia*, *Aves*, *Reptilia* and *Amphibia*, excluding (excluding humans (*Homo sapiens*) and domestic species (animals that have been domesticated to be kept as pets or food). Host taxonomy are formatted in the first instance using accepted scientific names in the IUCN Red List(11), and when not available, the Catalogue of life database (12), last downloaded using an API in October 2020.

In order to keep information up to date, the administrators of the SpillOver website will download new versions of databases on a regular basis and update the database, which will recalculate the ranking of viruses in the database available in the webtool.

Using the cleaned NCBI nucleotide database extraction for viruses in the ICTV list described above, the total number of hosts species and orders identified for each virus (Fig. S3) were used to inform categorization of the two host plasticity risk factors into the following groups: 1 species, 2 species, 3-5 species, 6-10 species, and more than 10 species.

*Risk Levels* were assigned for a virus by counting the total number of unique host species/orders in the database, assigning *Risk Level* as shown below. Cases for which a host species was identified only at the genus level were not included in calculations of host diversity, unless the virus had only been recorded in an unidentified host species, then a *Risk Level Score* of 1 was assigned.

**Table S1:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Host plasticity'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| 1 species / order | 1 |
| 2 species / order | 2 |
| 3 - 5 species / order | 3 |
| 6 - 10 species / order | 4 |
| > 10 species / order | 5 |

NOTE: *Risk Level* is grouped by virus

Virus Data Risk calculation:

Host plasticity - No. of species: *Risk Factor Influence* (2.632707775) x *Risk Level Score*

Host plasticity - No. of orders: *Risk Factor Influence* (2.768817204) x *Risk Level Score*

## Host sample type diversity – NOT INCLUDED IN TOP 31 RISK FACTORS

Samples were classified into body system categories: Cardiovascular system (i.e. blood, heart), Digestive system (i.e. intestine), Endocrine system (i.e. skin, hair), Integumentary system (i.e. skin, hair), Urogenital system (i.e. urine, urogenital swab), Lymphatic system (i.e. lymph node), Muscular Skeletal System (i.e. muscle tissue, bone), Nervous system (i.e. brain, nerve), Respiratory system (i.e. oral or nasal swab, lung). Fecal (i.e. stool) samples are assigned their own category. *Risk Levels* categories were calculated as the total number of unique body system categories the virus is known to infect across the host species in the database. Virus sample diversity is the number of unique body system categories listed within the database listed.

**Table S2:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Host sample type diversity'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| 1 body system category | *0.5* |
| 2 body system categories | *1.0* |
| 3 body system categories | *1.5* |
| 4 body system categories | *2.0* |
| 5 body system categories | *2.5* |
| 6 body system categories | *3.0* |
| 7 body system categories | *3.5* |
| 8 body system categories | *4.0* |
| 9 body system categories | *4.5* |
| 10 body system categories | *5.0* |
| Unknown | 2.5 |

NOTE: *Risk Level* 'Unknown' with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

*Risk Factor Influence* (1.943) x *Risk Level Score*

## Conservation status of the host species – NOT INCLUDED IN TOP 31 RISK FACTORS

A custom function "HostTaxonomy" was created within R software(13) using rredlist(14) and sqldf(15). For IUCN Red List-recognized host species, conservation measures (Least Concern or Lower Risk – least concern, Lower Risk – conservation dependent, Near Threatened or Lower Risk – near threatened, Vulnerable, Endangered, Critically Endangered, Extinct in the Wild, Extinct, Data Deficient, Unknown) were extracted from the IUCN Red List database using a personal API token(11). The information in this study is valid as of October 2020. For future updates to the risk ranking, the SpillOver website has the functionality to update from the IUCN Red List database on a regular basis. Species not assessed by the IUCN, are defaulted as 'Unknown' conservation status.

**Table S3:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Conservation status of host species'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| Extinct | *0* |
| Extinct in the wild | *1* |
| Endangered | *3* |
| Vulnerable | *3* |
| Near Threatened | *3* |
| Lower Risk - near threatened | *3* |
| Data Deficient | *3* |
| Lower Risk - conservation dependent | *4* |
| Critically Endangered | *5* |
| Least concern | *5* |
| Lower Risk - least concern | *5* |

| | |
|---|---|
| Unknown | 2.5 |

*NOTE: If multiple risk levels were present for a virus, the risk level with the highest score was used for the risk calculation. Risk Level* 'Unknown' with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

*Risk Factor Influence* (1.387) x *Risk Level Score*

## Habitat breadth of the host(s) – NOT INCLUDED IN TOP 31 RISK FACTORS

Habitat is the place or environment where a plant or animal naturally or normally lives and grows. A custom function "HostHabitat" was created within R software(13) using rredlist(14) and sqldf(15) to extract habitat use data for a given host species from the IUCN Red List database using a personal access (API) token(11). For IUCN Red List-recognized host species, habitats were defined according to IUCN Red List definitions: Artificial – Aquatic, Artificial – Terrestrial, Caves and Subterranean Habitats, Desert, Forest, Grassland, Introduced Vegetation, Marine – Coastal, Marine – Intertidal, Marine – Neritic, Marine – Oceanic, Marine - Deep Ocean Floor, Rocky – Areas, Other Savanna, Shrubland, and Wetlands. For species not assessed by the IUCN Red List, habitat use was listed as 'Unknown'. *Risk Levels* were then calculated as the total number of unique habitat categories (excluding 'Unknown') each host species is known to use. The information in this study is valid as of October 2020. For future updates to the risk ranking, the SpillOver website has the functionality to update from the IUCN Red List database on a regular basis.

**Table S4:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Habitat breadth of the host'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| 1 habitat category | 0.29 |
| 2 habitat categories | 0.59 |
| 3 habitat categories | 0.88 |
| 4 habitat categories | 1.18 |
| 5 habitat categories | 1.47 |

| | |
|---|---|
| 6 habitat categories | 1.76 |
| 7 habitat categories | 2.06 |
| 8 habitat categories | 2.35 |
| 9 habitat categories | 2.65 |
| 10 habitat categories | 2.94 |
| 11 habitat categories | 3.23 |
| 12 habitat categories | 3.53 |
| 13 habitat categories | 3.82 |
| 14 habitat categories | 4.12 |
| 15 habitat categories | 4.41 |
| 16 habitat categories | 4.70 |
| 17 habitat categories | 5.00 |
| Unknown | 2.5 |

*NOTE: If multiple risk levels were present for a virus, the risk level with the highest score was used for the risk calculation. Risk Level* ‘Unknown’ with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

*Risk Factor Influence* (1.857) x *Risk Level Score*

### Diet breadth of the host species – NOT INCLUDED IN TOP 31 RISK FACTORS

Diet breadth for mammals and birds was calculated from the total number of dietary categories outlined in the reference dataset produced by Wilman*, et al.* (16). As per their definition, categories included: Invertebrates, Birds/Mammals, Reptiles/Snakes/Amphibians/Salamanders, Fish, Vertebrates – other, Scavenged objects, Fruit, Nectar/Pollen/Plant exudates/Gums, Seed/Maize/Nuts/Spores/Wheat/Grains, Plant – other. For the purposes of this analysis, a species consumes a food category if it’s estimated percent use was ≥ 1%. Scientific names were corrected to align with IUCN Redlist designations, and where duplicate records occurred, the record with the highest confidence is taken. Species not assessed by Wilman*, et al.* (16), are defaulted as 'Unknown' diet breadth. *Risk Levels* are calculated as the total number of unique diet categories each host species is known to use.

**Table S5:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Diet breadth of the host'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| 1 diet category | 0.5 |
| 2 diet categories | 1.0 |
| 3 diet categories | 1.5 |
| 4 diet categories | 2.0 |
| 5 diet categories | 2.5 |
| 6 diet categories | 3.0 |
| 7 diet categories | 3.5 |
| 8 diet categories | 4.0 |
| 9 diet categories | 4.5 |
| 10 diet categories | 5.0 |
| Unknown | 2.5 |

*NOTE: If multiple risk levels were present for a virus, the risk level with the highest score was used for the risk calculation. Risk Level* 'Unknown' with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

*Risk Factor Influence* (1.059) x *Risk Level Score*

## Geography of the host(s)

Species range files for *Mammallia*, *Amphibia* and *Reptilia* were downloaded from the IUCN Red List database(11) on October 25th 2017. Shapefiles for bird ranges (*Aves*) were provided by Birdlife International(17) (downloaded

October 27th 2017). Spatial polygon datasets were uploaded into QGIS (version 3.0.3-Girona)(18) for processing. The fix geometry function was conducted on all files to overcome geometry errors. IUCN range files were filtered to remove uncertain range files ('Not Mapped', 'Origin uncertain', 'Possibly Extant (resident)', 'Possibly Extinct', 'Presence Uncertain', 'Probably Extant (resident)'). Species range size was calculated using the lets.rangesize function in the R package letsR(13, 19). Bird range files were filtered to remove areas of passage and undetermined or sporadic occurrence, thus reducing uncertainty contained in the dataset. Range size attributes were extracted from spatial files for bird species range. The sum of all range areas for a given species was calculated in R software(13). Country land mass was collected from the World Bank 2016 records(20), unless unavailable, in which case, data were extracted from the United Nations 2007 land mass data(21) followed by the CIA World Factbook(22).

A custom function "HostGeog" was created within R software(13) in order to calculate the geographic range of the host species. Country of occurrence for each host species was extracted from the IUCN Red List database(11) using sqldf(15) and rredlist(14) with a personal API token.

Country names/designations are defined as per the IUCN Red List database(11) based on ISO 3166 list of countries. Global distribution of the host species is classified according the following descriptions of species occurrence; (1) Isolated endemic – one country with a range size <10,000 km$^2$, (2) National (small) - a single country with a land mass < 2.5 million km$^2$, (3) National (large) – a single country with a land mass > 2.5 million km$^2$, (4) Regional - more than 1 country in 1 WHO region, (5) Semi-global - more than 1 country and 2-3 WHO regions, (6) Global - more than 1 country and > 3 WHO regions(23), or (7) Unknown - data not available. In this study, the Arctic region and Antarctica are considered a country and their own respective WHO region.

**Table S6:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Geography of the host'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| Isolated | *1* |
| National - small | *2* |
| National - large | *3* |
| Regional | *3.5* |

| | |
|---|---|
| Semi-global | *4* |
| Global | *5* |
| Unknown | 2.5 |

*NOTE: If multiple risk levels were present for a virus, the risk level with the highest score was used for the risk calculation. Risk Level* ‘Unknown’ with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

*Risk Factor Influence* (2.418) x *Risk Level Score*

## Genetic relatedness between host species and humans

The timescale of evolutionary divergence is used as an indicator of genetic relatedness between host species and humans (*Homo sapiens*). To determine suitable risk level categories, a list of species in the IUCN Red List database was downloaded within R software(13) using rredlist(14) with a personal API token on the 31st October 2017. Up to 2 (if available) randomly chosen species per order in the IUCN list was used to evaluate the distribution of divergence times across species in the classes *Amphibia*, *Reptilia*, *Aves* and *Mammalia*. For each species, the estimated time of divergence (in Million Years Ago, MYA) from humans (*Homo sapiens*) was calculated using TimeTree(24) (http://www.timetree.org/), and based on the their distribution (Fig. S4), the following categories were created: Very Close (0-10 MYA), Very Recent (10-50 MYA), Recent (50-100 MYA), Far (100-200 MYA) and Very Far (>200 MYA).

For each host species in the database, the estimated time of divergence (in Million Years Ago, MYA) from humans (*Homo sapiens*) was calculated on the TimeTree website (24) and the value manually entered into a database. A function “MYAgroup” was created in R software(13) using sqldf(15) to group the estimated time of divergence for each host species accordingly into risk levels. Because this factor is not extractable from a public database without further calculation, submitters of viruses to *SpillOver* are provided with simple instructions for calculating it when they upload their new virus information.

**Table S7:** Risk Levels and associated Risk Level Scores assigned to the risk factor ‘Genetic relatedness between the host species and humans’

| *Risk Level* | *Risk Level Score* |
|---|---|
| Very Far (>200 MYA) | 1 |
| Far (100-200 MYA) | 2 |
| Recent (50-100 MYA) | 3 |
| Very Recent (10-50 MYA) | 4 |
| Very Close (0-10 MYA) | 5 |
| Unknown | 2.5 |

*NOTE: If multiple risk levels were present for a virus, the risk level with the highest score was used for the risk calculation. Risk Level* 'Unknown' with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

Formula: *Risk Factor Influence* (2.420) x *Risk Level Score*

*NOTE: Risk Level* 'Unknown' with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* available.

### *Environmental Risk Factors*

### Land use in the host's ecosystem

The location where the virus was detected in an animal was used to determine the type of land use in the host's ecosystem. A custom function "LandUse" was created to extract land use categories from the LADA land-use systems maps(25) using raster(26), plyr(27) and sqldf(15) packages in R(13). Land use was re-categorized to the broader LADA descriptors (Forest, Grasslands, Shrubs, Crops, Agriculture, Urban land, Wetlands, Sparse vegetation, Bare areas, Open water, and Unknown for undefined or data deficient areas).

**Table S8:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Land use in the host ecosystem'

| *Risk Level* | *Risk Level Score* |
|---|---|
| Forest | 5 |
| Grasslands | 3 |
| Shrubs | 2 |
| Crops | 4 |
| Agriculture | 4 |
| Urban land | 3.5 |
| Wetlands | 2 |
| Sparse vegetation | 2 |
| Bare areas | 1 |
| Open water | 1 |
| Unknown | 2.5 |

*NOTE: If multiple risk levels were present for a virus, the risk level with the highest score was used for the risk calculation. Risk Level* ‘Unknown’ with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

Formula: *Risk Factor Influence* (2.113) x *Risk Level Score*

## Livestock density the host’s ecosystem

The location where the virus was detected in an animal was used to determine the livestock density in the host’s ecosystem. A custom function “Livestock” was created to extract livestock density from LADA land-use systems maps(25) using raster(26), plyr(27) and sqldf(15) packages in R(13). Using LADA definitions, livestock density was assigned as either High, Medium, Low, None, or Unknown for undefined or data deficient areas.

**Table S9:** Risk Levels and associated Risk Level Scores assigned to the risk factor ‘Livestock density in the host ecosystem’

| *Risk Level* | *Risk Level Score* |
|---|---|
| High | *5* |
| Low | *1* |
| Medium | *3* |
| No Livestock | *0* |
| Unknown | 2.5 |

*NOTE: If multiple risk levels were present for a virus, the risk level with the highest score was used for the risk calculation. Risk Level* 'Unknown' with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

*Risk Factor Influence* (2.175) x *Risk Level Score*

### Human population density in the host's ecosystem

The location where the virus was detected in an animal was used to determine the human population density in the host's ecosystem. A custom function "PopDensity2015" was created that uses raster(26), rgdal(28), plyr(27) and sqldf(15) packages in R(13) to extract attributes from the NASA Sedac Gridded population of the World UN Adjusted population density v4 2015 map(29). Population density was grouped into the following categories: Unknown, 0-5, 5-25, 25-250, 250-1,000 and >1,000 people per km$^2$.

**Table S10:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Human population density in the host ecosystem'

| *Risk Level* | *Risk Level Score* |
|---|---|
| 0-5 persons per km$^2$ | *1* |
| 5-25 persons per km$^2$ | *2* |
| 25-250 persons per km$^2$ | *3* |
| 250-1,000 persons per km$^2$ | *4* |

|  |  |
|---|---|
| >1,000 persons per km$^2$ | *5* |
| Unknown | 2.5 |

*NOTE: If multiple risk levels were present for a virus, the risk level with the highest score was used for the risk calculation. Risk Level* ‘Unknown’ with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

*Risk Factor Influence* (2.547) x *Risk Level Score*

## Urbanization in the host’s ecosystem

Probability of urban expansion is used as an indicator of land conversion and urbanization processes. A custom function “Urbanization” was created to extract probability values from the Global Grid of Probabilities of Urban Expansion to 2030 using the location where the virus was detected in an animal, if available (30). To account for data source recommendations, a 20 km radius buffer was applied to longitude and latitude data points, and probability of urban extraction attributes were extracted from all 5 km$^2$ grids within the buffer zone using raster(26). Urbanization in the host’s ecosystem was defined as presence of a high probability (>75%) of urbanization within the 20 km radius buffer zone.

**Table S11:** Risk Levels and associated Risk Level Scores assigned to the risk factor ‘Urbanization in the host ecosystem’

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| No | 0 |
| Yes | 5 |
| Unknown | 2.5 |

*NOTE: If multiple risk levels were present for a virus, the risk level with the highest score was used for the risk calculation. Risk Level* ‘Unknown’ with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

*Risk Factor Influence* (2.246) x *Risk Level Score*

## Agricultural intensification in the host's ecosystem

Agricultural intensification was defined as presence of novel pasture or cropland within a 20 km radius buffer of the longitude and latitude of where the virus was detected in an animal, if available. Using previously described methods, global maps of cropland, pasture, and urban layers were adapted from the publicly available HYDE 2005 dataset (31, 32), which assigns a percentage score for each land use category (e.g. 14% Urban, 29% Crops, 41% Pasture) to each raster cell. Cells within the layers were re-categorized as only one land use type: urban cells were defined as those containing greater than 25% urban land, while cells with at least 50% human land-use were assigned either pasture or crops, depending on which is greater. Areas with less than 50% human land-use and less than 25% urban land were not assigned a land-use type. The cells in resulting rasters have one of three values: 1 = urban, 3 = crops, 5 = pasture. For the purposes of this analysis, a map of land use change over time was created by assigning human land-use categories as above for both HYDE datasets for 1970 and 2005, and then highlighting only cells with novel Urban, Pasture or Crop development since 1970. Using a custom R script, a 20 km radius buffer is applied to the longitude and latitude of where the virus was detected in an animal using the raster(26) package. Land use change attributes are then extracted from all cells within the buffer zone, and agricultural intensification is assigned if there is presence of novel pasture or cropland within any of the cells of the buffer zone.

**Table S12:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Agricultural intensification in the host ecosystem'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| Yes | 5 |
| No | 0 |
| Unknown | 2.5 |

*NOTE: If multiple risk levels were present for a virus, the risk level with the highest score was used for the risk calculation. Risk Level* 'Unknown' with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

*Risk Factor Influence* (2.262) x *Risk Level Score*

## Forest degradation / deforestation in the host's ecosystem

Loss in the forest canopy cover, defined as vegetation taller than 5 m in height, was used as an indicator of forest degradation/deforestation in the host's ecosystem. Using a custom function called "Deforestation" that calls packages gfcanalysis(33), sqldf(15), rgdal(28) and raster(26), annual forest cover and forest loss (in hectares) for each longitude and latitude of where the virus was detected in an animal was extracted from the Global Forest Change 2017 dataset(34) using a 20 km buffer zone around spatial points and a forest cover threshold set at 30%(35, 36). Percentage loss of tree cover was calculated on an annual time scale between 2000 and 2017 from the difference between annual area cover and loss. Finally, forest degradation/deforestation was determined to be present if there was ≥ 20% forest cover change between 2000 and 2017, as per the FAO 2015 Global Forest Resources Assessment definition forest cover change(37).

**Table S13:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Forest degradation / deforestation in the host ecosystem'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| Yes | 5 |
| No | 0 |
| Unknown | 2.5 |

*NOTE: If multiple risk levels were present for a virus, the risk level with the highest score was used for the risk calculation. Risk Level* 'Unknown' with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

*Risk Factor Influence* (2.271) x *Risk Level Score*

## Number of primary high-risk disease transmission interfaces

High-risk disease transmission interfaces were defined and adapted from categories previously described in Johnson, *et al.* (8). For this study, direct and indirect interface categories were combined into one group, and the category 'Restaurant' replaced 'Consumed' as the interface description. Four extra categories were added including: Extractive Industry, Livestock, Pristine Habitat and Wild Animal Farm. Therefore, categories were as follows:

Hunted, Wildlife Management, Human Dwellings, Market, Ecotourism, Veterinarian or Researcher, Crop or Agricultural, Pristine Habitat, Zoo or Sanctuary, Extractive Industry, Restaurant, Wild Animal Farm, Livestock, Pets - wild animal, Pets - domestic animal, Laboratory and Unknown. *Risk Levels* were calculated as the total number of unique interface categories in which animals have tested positive for the virus in the database.

**Table S14:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Number of primary disease transmission interfaces'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---:|
| 1 interface category | 0.3125 |
| 2 interface categories | 0.625 |
| 3 interface categories | 0.938 |
| 4 interface categories | 1.25 |
| 5 interface categories | 1.56 |
| 6 interface categories | 1.88 |
| 7 interface categories | 2.19 |
| 8 interface categories | 2.5 |
| 9 interface categories | 2.81 |
| 10 interface categories | 3.13 |
| 11 interface categories | 3.44 |
| 12 interface categories | 3.75 |
| 13 interface categories | 4.06 |
| 14 interface categories | 4.38 |
| 15 interface categories | 4.69 |
| 16 interface categories | 5.00 |

|  |  |
|---|---|
| Unknown | 2.5 |

*NOTE: If multiple risk levels were present for a virus, the risk level with the highest score was used for the risk calculation. Risk Level* 'Unknown' with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

*Risk Factor Influence* (2.658) x *Risk Level Score*

## Frequency and intimacy of contact at the transmission interface

The frequency and intimacy of contact between humans and wildlife and humans and domestic animals for each primary transmission interface were assessed based on expert opinion and risk characterization of primary interfaces during our virus discovery project activities. *Risk Levels* were categorized for each primary interface as follows: Unknown, Low, Medium or High. Levels were automatically assigned to the primary interface in which animals have tested positive for the virus. Unknown *Risk Levels* are assigned when no primary interface was available.

**Table S15:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Frequency and intimacy of contact at the transmission interface'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| Low | *1* |
| Medium | *3* |
| High | *5* |
| None | *0* |
| Unknown | 2.5 |

*NOTE: If multiple risk levels were present for a virus, the risk level with the highest score was used for the risk calculation. Risk Level* 'Unknown' with a set *Risk Level Score* of 2.5 is only used if no other *Risk Level* is available.

Virus Data Risk calculation:

Frequency of interaction: domestic animals – humans: *Risk Factor Influence* (2.540) x *Risk Level Score*

Intimacy of interaction: domestic animals – humans: *Risk Factor Influence* (2.514) x *Risk Level Score*

Frequency of interaction: wild animals – humans: *Risk Factor Influence* (2.709) x *Risk Level Score*

Intimacy of interaction: wild animals – humans: *Risk Factor Influence* (2.709) x *Risk Level Score*

### ***Virus Risk Factors***

## Characteristics of a virus

Three properties based on the nature of the viral genome were extracted from the Tenth Report of the International Committee on Taxonomy of Viruses (ICTV)(38). Descriptors include genome type (DNA or RNA), whether a virus has an envelope surrounding its genome, and if the viral genome is broken up into segments. Segmentation was defined as more than one linear or circular genomic segment. Properties were assigned to a virus based on the virus family to which they belong.

### Genome classification of the virus

**Table S16:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Genome classification of the virus'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| DNA | *3* |
| RNA | *5* |

Virus Data Risk calculation:

*Risk Factor Influence* (2.046) x *Risk Level Score*

### Envelope status of the virus

**Table S17:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Envelope status of the virus'

| *Risk Level* | *Risk Level Score* |
|---|---|
| No | *5* |
| Yes | *3* |
| Unknown | 2.5 |

Virus Data Risk calculation:

*Risk Factor Influence* (1.553) x *Risk Level Score*

*NOTE: Risk Factor Influence is below the threshold of 2, but evidence indicates this risk factor is important for human-to-human transmission(6)*

## Viral genome segmentation

**Table S18:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Virus genome segmentation'

| *Risk Level* | *Risk Level Score* |
|---|---|
| Yes | *3* |
| No | *5* |
| Unknown | 2.5 |

Virus Data Risk calculation:

*Risk Factor Influence* (1.703) x *Risk Level Score*

*NOTE: Risk Factor Influence is below the threshold of 2, but evidence indicates this risk factor is important for human-to-human transmission(6)*

## Ability of a virus species to infect humans, terrestrial mammals, birds and other animals

A virus species was classified as able to infect a taxonomic group (human, terrestrial mammals, birds, other animals) if identified within either the literature review, nucleotide virus records extracted from NCBI (both described above in ‘Reference Data Sources for Risk Factors and Viruses’), or wildlife detections contained in the database. A custom function called “speciesTaxonomy” was created to classify host species as either a Human (*Homo sapiens*), Birds (Order belongs to *Aves*), Terrestrial Mammals (mammals that inhabit the land) or Other Animals (all other animal that do not belong in previous classifications) (Table S33).

**Table S19:** Risk Levels and associated Risk Level Scores assigned to the risk factor ‘Ability of the virus to infect humans, terrestrial mammals, birds and other animals’

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| Not known to infect | *1* |
| Yes | *5* |

Virus Data Risk calculation:

Viral infectivity - humans: *Risk Factor Influence* (2.858) x *Risk Level Score*

Viral infectivity - terrestrial animals: *Risk Factor Influence* (2.432) x *Risk Level Score*

Viral infectivity - birds: *Risk Factor Influence* (1.846) x *Risk Level Score* – NOT INCLUDED IN TOP 31 RISK FACTORS

Viral infectivity - other animals: *Risk Factor Influence* (1.460) x *Risk Level Score* – NOT INCLUDED IN TOP 31 RISK FACTORS

## Proportion of virus species known to infect animal taxa in a viral family

The proportion of virus species known to infect humans, terrestrial mammals, birds and other animals within each of the 26 viral families was calculated using the ICTV list of 1,126 viruses. First, a virus species was classified as able to infect a taxonomic group if identified as zoonotic from the literature review or if the nucleotide virus records extracted from NCBI included human samples (both described above in ‘Reference Data Sources for Risk Factors and Viruses’). The proportion of virus species known to infect humans, terrestrial mammals, birds and other animals within each virus family was then calculated and grouped into categories: Very Low (0-10%), Low (10-25%), Medium (25-50%), High (50-75%) and Very High (75-100%). The *Risk Level* was determined by the virus family to which the virus belongs.

**Table S20:** Risk Levels and associated Risk Level Scores assigned to the risk factor ‘Proportion of virus species known to infect animal taxa in a viral family’

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| Very Low (0-10%) | *1* |
| Low (10-25%) | *2* |
| Medium (25-50%) | *3* |
| High (50-75%) | *4* |
| Very High (75-100%) | *5* |

Virus Data Risk calculation:

Proportion viral family infect humans: *Risk Factor Influence* (2.432) x *Risk Level Score*

Proportion viral family infect terrestrial animals: *Risk Factor Influence* (1.989) x *Risk Level Score*

Proportion viral family infect birds: *Risk Factor Influence* (1.532) x *Risk Level Score* – NOT INCLUDED IN TOP 31 RISK FACTORS

Proportion viral family infect other animals: *Risk Factor Influence* (1.167) x *Risk Level Score* – NOT INCLUDED IN TOP 31 RISK FACTORS

## Duration of virus species infection in humans

Duration of infection in humans is defined in categories; less than 4 weeks and more than 4 weeks(6), Not known to infect humans or Unknown duration. The *Risk Level* was determined by the virus species to which the virus belongs. For virus species within the ICTV list, duration of infection in humans was assessed using the literature review described above. For other virus species, if the virus species was listed as known to infect humans in the risk factor

‘Ability of a virus to infect humans’, it was assigned a risk level of ‘Unknown duration’; if the virus species was not known to infect humans, it was assigned a risk level of ‘None’.

**Table S21:** Risk Levels and associated Risk Level Scores assigned to the risk factor ‘Duration of virus species infection in humans’

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| Not known to infect humans | 0 |
| Less than 4 weeks | *3* |
| More than 4 weeks | *5* |
| Unknown duration | *2.5* |

Virus Data Risk calculation:

*Risk Factor Influence* (2.021) x *Risk Level Score*

## Severity of the disease caused by the virus species in humans – NOT INCLUDED IN TOP 31 RISK FACTORS

The severity of illness in humans was classified according to known symptoms and disease caused by a virus species in people: Deadly (known cause of mortality), Severe Illness (unable to function in a normal manner), Mild Illness (limited disruption to function), Found in a sick human of unknown cause of illness or Not known to cause disease. The *Risk Level* was determined by the virus species to which the virus belongs. For virus species within the ICTV list, severity of disease was assessed using the literature review described above. Other viruses in the database were assigned the *Risk Level* of ‘Not known to cause disease’, unless stated overwise.

**Table S22:** Risk Levels and associated Risk Level Scores assigned to the risk factor ‘Severity of the disease caused by the virus species in humans’

| *Risk Level* | *Risk Level Score* |
|---|---|
| Not known to cause disease | *1* |
| Found in sick human of unknown cause of illness | *2* |
| Mild illness | *3* |
| Severe illness | *4* |
| Deadly | *5* |

Virus Data Risk calculation:

*Risk Factor Influence* (1.886) x *Risk Level Score*

## Proportion of virus species that are known human pathogens in the viral family

The proportion of virus species that are known human pathogens in a viral family was calculated from the number of virulent (known to cause deadly, severe or mild illness) versus avirulent virus species in the ICTV list grouped by virus family. Virulence was assessed using the literature review described above. Results were subsequently grouped into categories: Very Low (0-10%), Low (10-25%), Medium (25-50%), High (50-75%) and Very High (75-100%). The *Risk Level* was determined by the virus family to which the virus belongs.

**Table S23:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Proportion of virus species that are known human pathogens in the viral family'

| *Risk Level* | *Risk Level Score* |
|---|---|
| Very Low (0-10%) | *1* |
| Low (10-25%) | *2* |
| Medium (25-50%) | *3* |
| High (50-75%) | *4* |
| Very High (75-100%) | *5* |

Virus Data Risk calculation:

*Risk Factor Influence* (2.406) x *Risk Level Score*

### Transmission between humans

The ability of a virus to transmit between humans was assessed using the literature review described above in 'Reference Data Sources for Risk Factors and Viruses'. For newly-discovered viruses, a *Risk Level* of 'Not known to transmit' was assigned, unless stated overwise.

**Table S24:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Transmission between humans'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| Not known to transmit | *1* |
| Yes | *5* |

Virus Data Risk calculation:

*Risk Factor Influence* (2.782) x *Risk Level Score*

### Ability of a virus to transmit between animals and humans

A list of known zoonotic viruses was extracted from the database created by Johnson*, et al.* (8). Virus names were updated to ICTV standards. Vector-borne viruses and viruses of animals that have been domesticated to be kept as pets or used as food were removed, as well as Influenza viruses due to the overwhelming amount of data for this group and the complexity of their subunits, resulting in a list of 38 viruses included in the analysis. For newly-discovered viruses, a *Risk Level* of 'Not known to transmit' was assigned, unless stated overwise.

**Table S25:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Ability to transmit between animals and humans'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| Not known to transmit | *1* |
| Yes | *5* |

Virus Data Risk calculation:

Formula: *Risk Factor Influence* (2.921) x *Risk Level Score*

## Ability of the virus species to cause an epidemic

The epidemicity of a virus species was categorized into *Risk Levels* according to historic records of small outbreaks (a noticeable but small number of cases), large outbreak (large number of cases over a small area), epidemics (an outbreak over a larger geographic area) or pandemic (an epidemic that has spread to multiple countries or regions of the world, with sustained transmission) in humans, animals or both. The *Risk Level* was determined by the virus species to which the virus belongs. For virus species within the ICTV list, incidences of outbreaks or epidemics/pandemics were identified using the literature review described above in 'Reference Data Sources for Risk Factors and Viruses'. For newly-discovered viruses, a *Risk Level* of 'No' was allocated, unless stated overwise.

**Table S26:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Ability of the virus to cause an epidemic'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| No | *0* |
| Outbreak or epidemic - animals only | *3* |
| Small outbreak - humans (+/- animals) | *3* |
| Large outbreak or epidemic - humans (+/- animals) | *4* |
| Pandemic - humans (+/- animals) | *5* |

Virus Data Risk calculation:

*Risk Factor Influence* (2.689) x *Risk Level Score*

### Virus caused a pandemic in humans

The historic or current ability of the virus to cause a pandemic in humans was determined from if the World Health Organization declared the virus as the cause of a pandemic(39). For newly-discovered viruses, a *Risk Level* of 'No' was allocated, unless stated overwise.

**Table S27:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Virus caused a pandemic in humans'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---:|
| No | *0* |
| Yes | *5* |

Virus Data Risk calculation:

Formula: *Risk Factor Influence* (2.689) x *Risk Level Score*

### Virulence of a virus species in terrestrial mammals/birds/other animals – NOT INCLUDED IN TOP 31 RISK FACTORS

Virulence of a virus species within each of the taxa groups (terrestrial mammals, birds, other animals) as a result of natural infections was classified according to symptoms and degree of pathology caused by a virus species to the relevant host: Severe Illness (unable to function in a normal manner), Mild Illness (limited disruption to function) or Not known to cause disease. Evidence of virulence via animal models or cell lines were not included. The *Risk Level* was determined by the virus species to which the virus belongs. For virus species within the ICTV list, severity of disease was assessed using the literature review described above. For newly-discovered viruses, a *Risk Level* of 'Not known to cause disease' was assigned, unless stated otherwise.

**Table S28:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Virulence of a virus species in animal taxa'

| *Risk Level* | *Risk Level Score* |
|---|---|
| Not known to cause disease | *1* |
| Mild Illness | *3* |
| Severe Illness | *4* |
| Deadly | *5* |

Virus Data Risk calculation:

Virulence - terrestrial mammals: *Risk Factor Influence* (1.602) x *Risk Level Score*

Virulence – birds: *Risk Factor Influence* (1.323) x *Risk Level Score*

Virulence - other animals: *Risk Factor Influence* (1.040) x *Risk Level Score*

## Proportion of virus species within a viral family that are known to infect more than 1 host species

The proportion of virus species that are known to infect more than 1 host species within a viral family was then calculated from the total number of distinct host species per virus species in the NCBI nucleotide records for ICTV list (described above in ‘Reference Data Sources for Risk Factors and Viruses’). Unidentified host species were not included in calculations. *Risk Levels* were categorized as follows: Very Low (0-10%), Low (10-25%), Medium (25-50%), High (50-75%) or Very High (75–100%). The *Risk Level* for a virus was determined by the virus family to which the virus belongs.

**Table S29:** Risk Levels and associated Risk Level Scores assigned to the risk factor ‘Proportion of virus species within a viral family that are known to infect more than 1 species’

| *Risk Level* | *Risk Level Score* |
|---|---|
| Very Low (1-10%) | *1* |
| Low (10-25%) | *2* |
| Medium (25-50%) | *3* |

| | |
|---|---|
| High (50-75%) | *4* |
| Very High (75-100%) | *5* |

Virus Data Risk calculation:

*Risk Factor Influence* (1.950) x *Risk Level Score*

## Transmission mode of the viral genus

Due to limitations in knowledge of how individual viruses are transmitted, the transmission mode of the viral genus was used as a proxy for all viruses within that genus. Transmission modes were extracted from the ViralZone database(7) and grouped into relevant categories as per described in Antonovics*, et al.* (40), resulting in the following categories; Vertical/Maternal, Sexual, Direct, Airborne, Indirect-Environment, Indirect-Fomite and Indirect-Vector. *Risk Levels* were then calculated as the total number transmission group categories within the virus genus to which the virus belongs. Unknown transmission is only counted if no other transmission category is present.

**Table S30:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Transmission mode of the virus genus'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| 1 transmission group | *0.714* |
| 2 transmission groups | *1.428* |
| 3 transmission groups | *2.142* |
| 4 transmission groups | *2.856* |
| 5 transmission groups | *3.570* |
| 6 transmission groups | *4.284* |
| 7 transmission groups | *4.998* |

| | |
|---|---|
| Unknown | 2.5 |

Virus Data Risk calculation:

*Risk Factor Influence* (2.597) x *Risk Level Score*

## Geography of the virus in animals

A custom function called "GeogVirus" was created in R software(13) using sqldf(15) to determine the virus distribution from the number of countries and WHO regions(23) in which the virus has been detected within the database. Using definitions of international spread adapted from Johnson*, et al.* (8), virus distribution was determined as: (1) National (small) – a single country with a land mass < 2.5 million km$^2$ (2) National (large) – a single country with a land mass > 2.5 million km$^2$, (3) Regional – more than 1 country in 1 WHO region, (4) Semi-global – more than 1 country and 2-3 WHO regions, or (5) Global – more than 1 country and > 3 WHO regions. The Arctic region and Antarctica were each considered a country and their own respective WHO region. Country names or designations were defined as per the IUCN Red List database(11) based on ISO 3166 list of countries.

**Table S31:** Risk Levels and associated Risk Level Scores assigned to the risk factor 'Geography of the virus in animals'

| ***Risk Level*** | ***Risk Level Score*** |
|---|---|
| National - small | *1* |
| National - large | *2* |
| Regional | *3* |
| Semi-global | *4* |
| Global | *5* |

Virus Data Risk calculation:

*Risk Factor Influence* (2.168) x *Risk Level Score*

## Creating a Database of Viruses of Animal Origin

### *PREDICT Dataset*

Collection and Testing Animal Samples: Between 2009 and 2019, 74,635 individual animals were sampled from eleven mammalian and avian orders (*Anseriformes, Carnivora, Cetartiodactyla, Chiroptera, Eulipotyphla, Galliformes, Lagomorpha, Passeriformes, Primates, Rodentia, Scandentia*) at high-risk pathogen transmission interfaces in 28 countries as part of PREDICT surveillance. The research protocol was approved by the University of California Animal Care and Use Committee (IACUC) (permit 17504, 19300).

In the process of sample collection, geographic and demographic data about the host and its environment were recorded. Animals were given a unique identifier and species identified. In cases of uncertainty, animals were identified to the genus level. Fecal samples were either collected as a freshly voided specimen, or via rectal or anal swabs. Respiratory samples were obtained via an oral, nasal or oropharyngeal swab during handling. Additional samples included urogenital swabs and urine. Blood was collected via needle or syringe or a needle prick and blood spot on a collection card. When possible, duplicate samples were stored in up to 1 mL of Viral Transport Medium (VTM) and 1 mL of NucliSens® Lysis Buffer (bioMérieux, Inc., Marcy-I'Étoile, France). In cases where the animal was either previously euthanized (for example in restaurants) or an animal died during capture or handling (rare), tissue biopsies were collected and placed into VTM and NucliSens® Lysis Buffer.

A total of 509,721 samples were tested for DNA and RNA viral families and genera using PREDICT standardized conventional PCR protocols, cDNA extraction and sequencing as per previously described(41-53). Target viral families included members of the *Adenoviridae, Astroviridae, Picobirnaviridae, Retroviridae, Parvoviridae, Polyomaviridae, Picornaviridae, Coronaviridae, Reoviridae, Papillomaviridae, Paramyxoviridae, Arenaviridae, Bunyaviridae, Poxviridae* and *Filoviridae*

Eight hundred and fifty-five viruses were detected in samples from 7,726 individual animals belonging to 251 mammalian species (plus 30 identified only to genus) within 26 countries. Viruses were classified by first comparing the distribution of pairwise sequence identities to define a % cut-off between taxa, and the resulting monophyletic groups were used as our operating taxonomic units. Groups that shared less nucleotide identity than the cut-off to a known sequence were labelled sequentially e.g., PREDICT_CoV-1, -2, -3 etc.; while groups sharing more identity than the cut-off to a sequence already in GenBank were considered to be strains of a known virus and assigned the same name as the matching sequence (e.g. Kenya_bat_CoV_KY33 or HKU-9). Virus sequences and associated metadata are available in the USAID Data Development Library (https://data.usaid.gov/)

Additionally, for non-ICTV classified viruses that had previously been identified prior to PREDICT, Blastn searches were run on the virus sequences to identify closely related viruses with ≥ 99% identity, ≥ 97% query cover and a low E-value. The Pub Med Central (PMC) database was used to conduct a literature search of all publications citing the matching NCBI accession numbers. When available, host and sample information was extracted from nucleotide records and corresponding publications. Detection records were included in the dataset only if confirmed by PCR and sequencing.

**Figures S1 to S6**

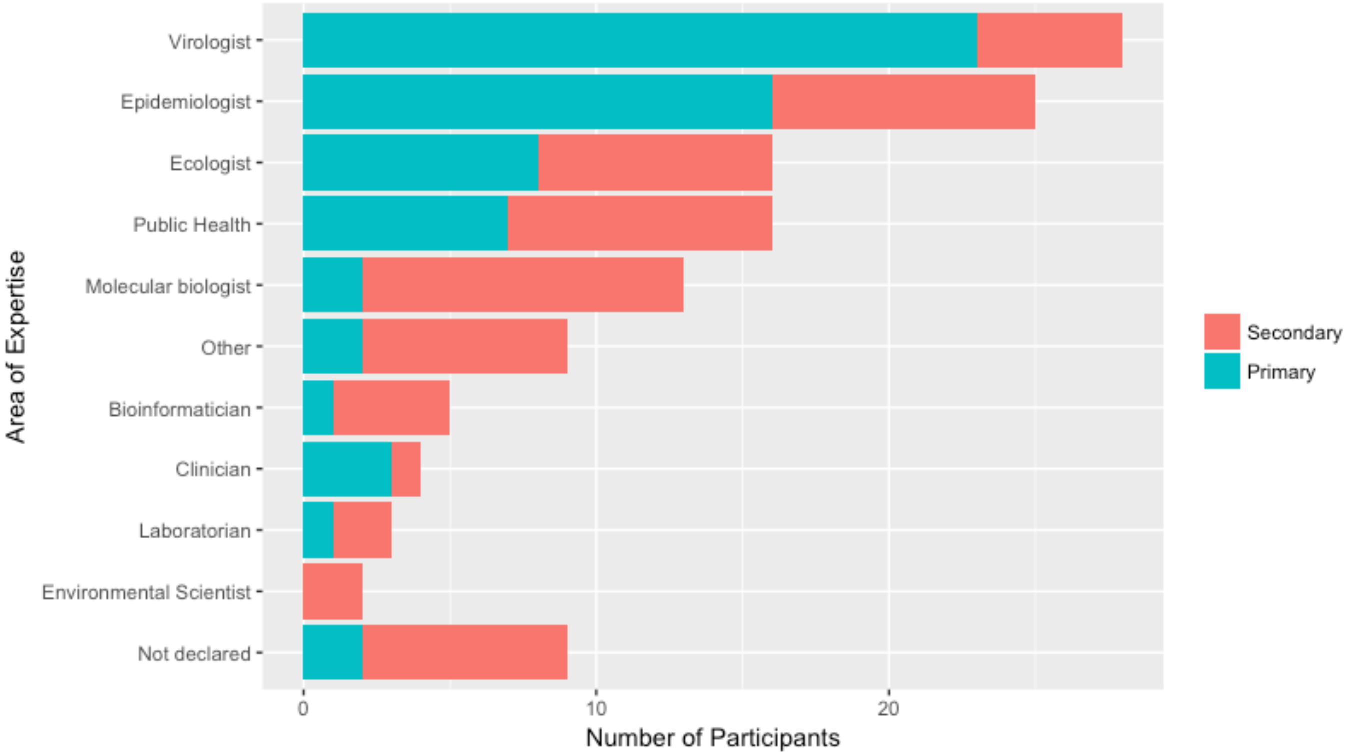

**Fig. S1** Participant (n = 65) reported areas of primary (blue) and secondary (red) expertise.

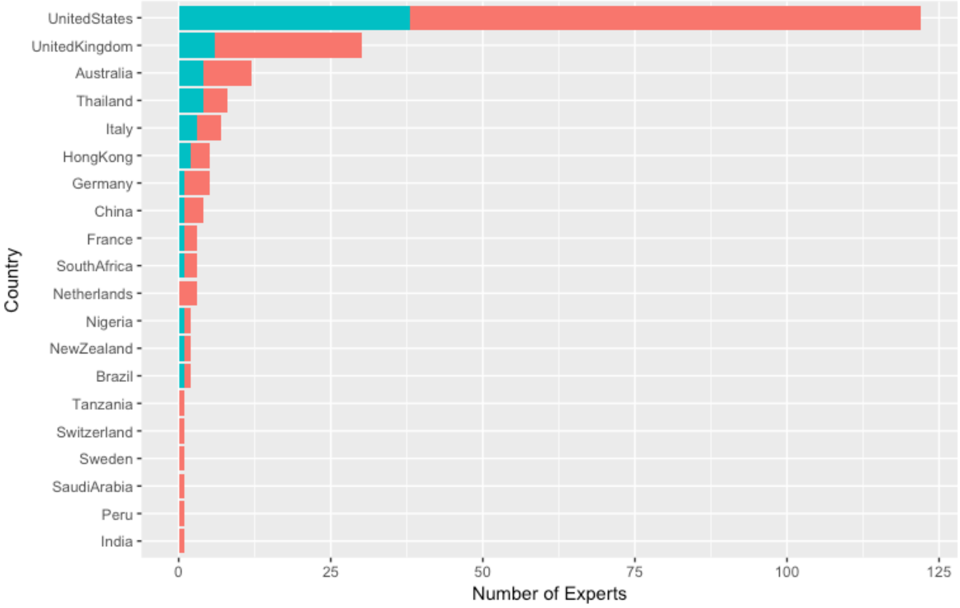

**Fig. S2** Number of experts (n = 150) recruited (blue) versus declined (red) to participate in the risk ranking survey by country of residence.

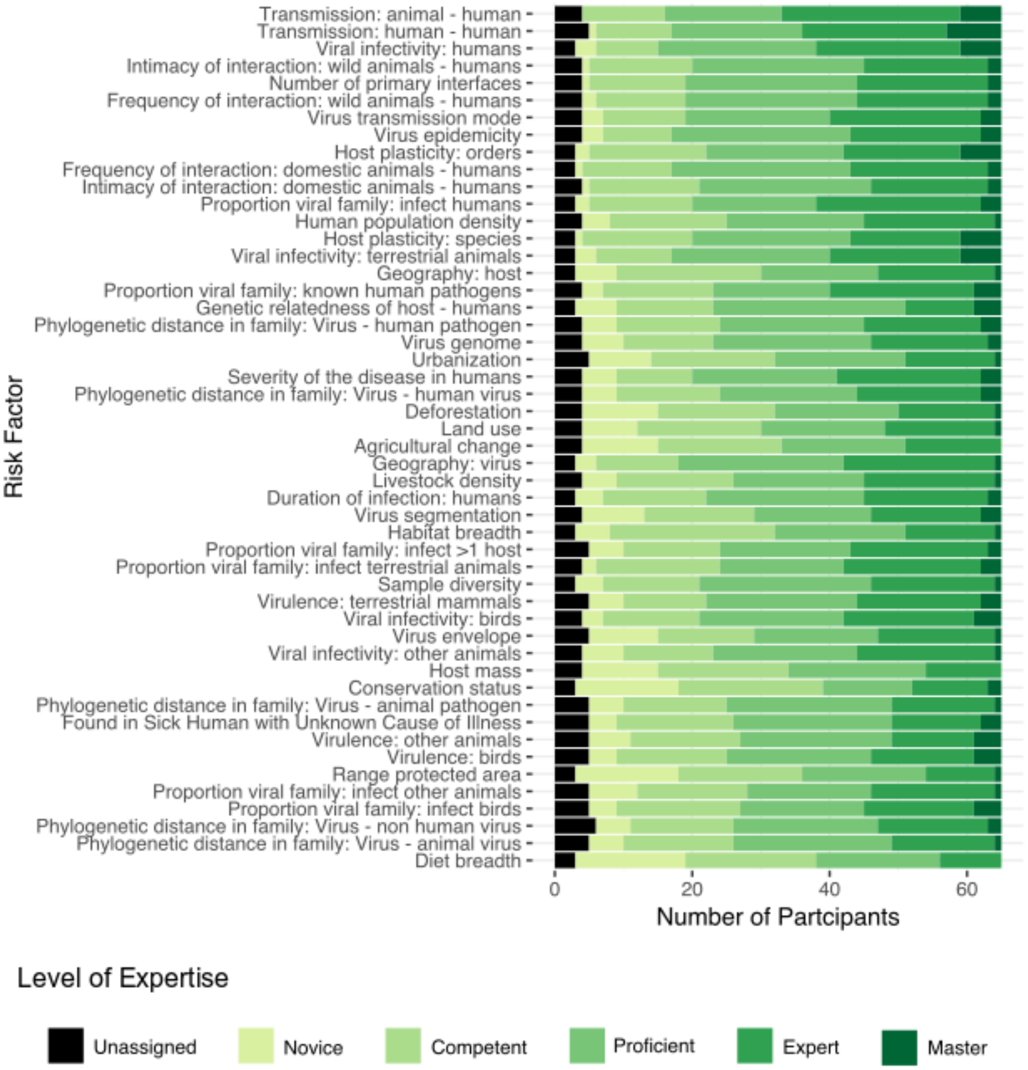

**Fig. S3** Participant (n = 65) self-assessment of their level of expertise for each factor in the risk ranking assessment. Order of competence is represented by strength of hue from Novice (light green) to Master (dark green). If the participant did not provide an answer, it was categorized as Unassigned (black).

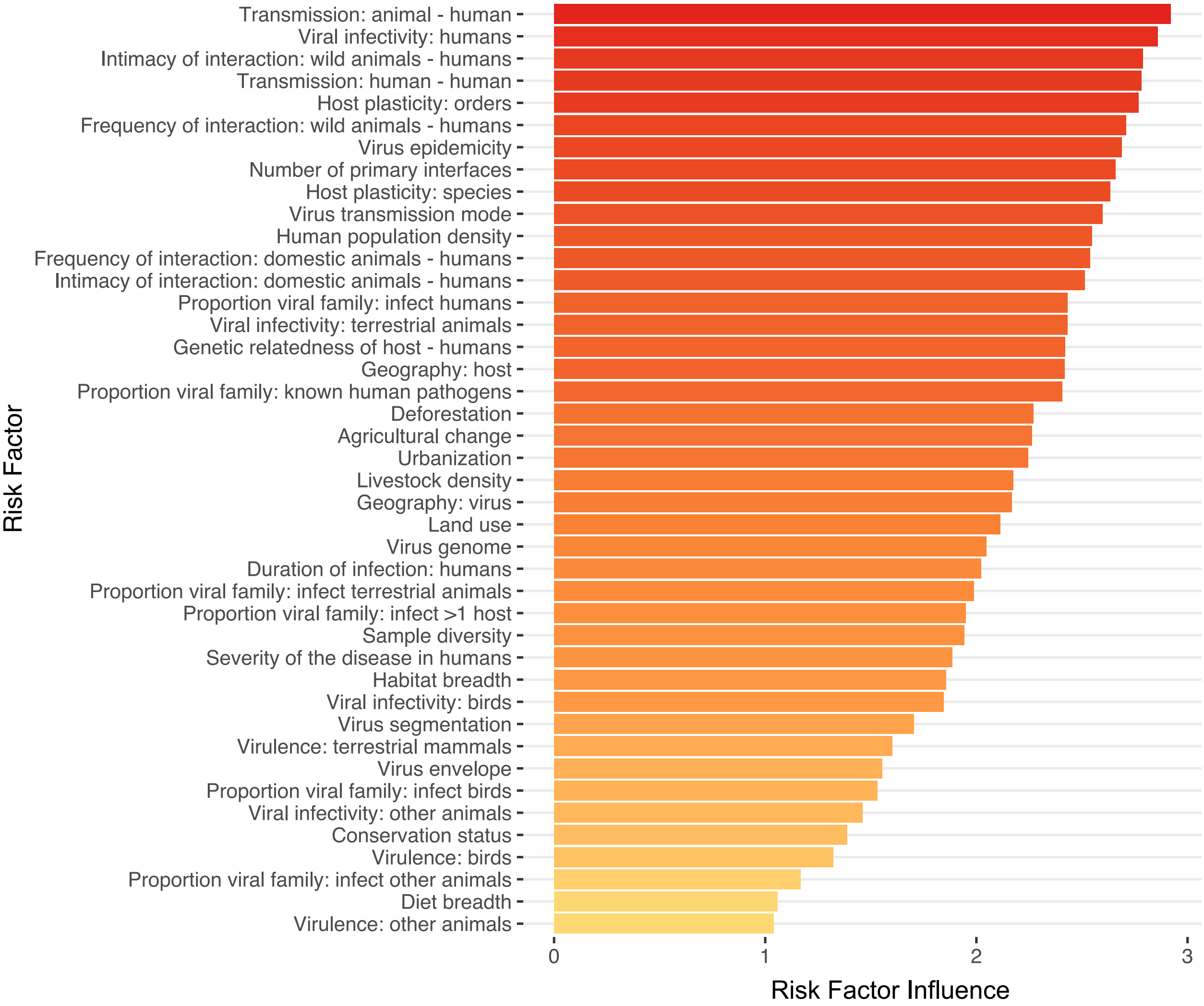

**Fig. S4** Expert opinion (n=65) of the contribution each factor explored in the risk ranking framework (n=42) plays in the risk of a new virus spillover from animals to humans. *Risk Factor Influence* was calculated as a weighted average risk score from the panelists on the assessed *Level of Risk* and their self-reported *Level of Expertise*.

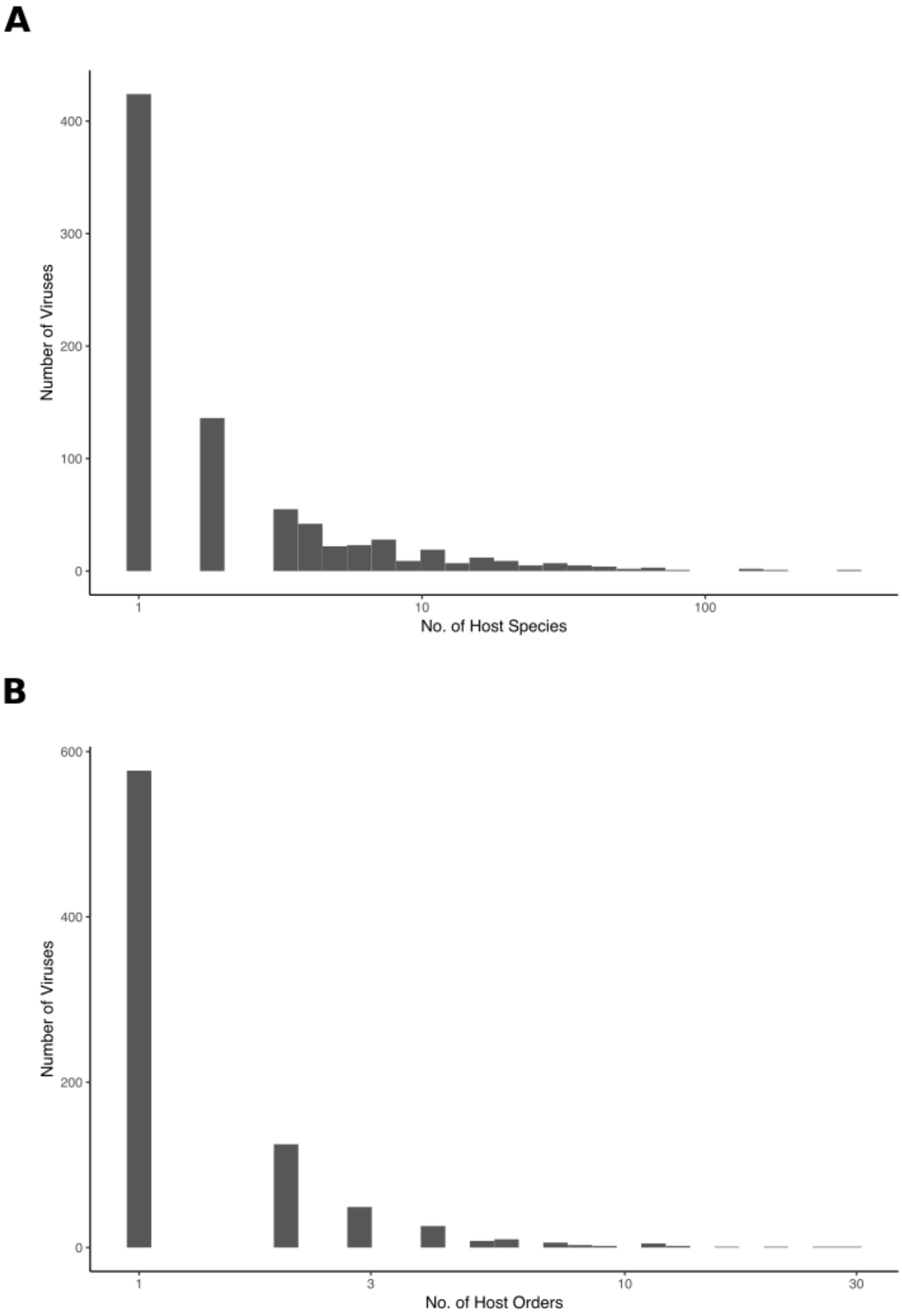

**Fig. S5** Number of host species and orders per virus from NCBI virus nucleotide records for 1,126 viruses from the ICTV list.

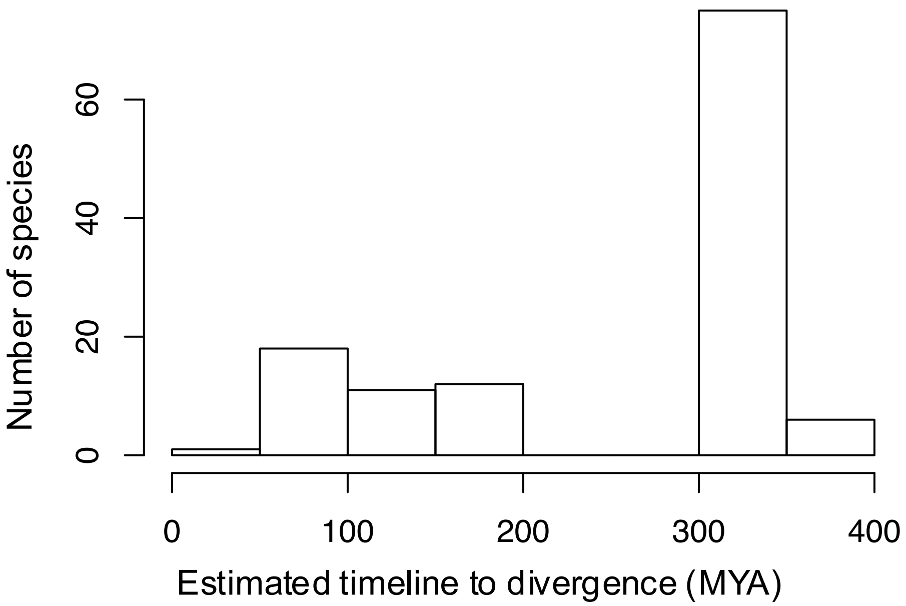

**Fig. S6** Estimated timeline of divergence between *Homo sapiens* and up to 2 (if available) randomly chosen species per order in the classes *Amphibia*, *Reptilia*, *Aves* and *Mammalia.*

**Tables S32 to S33**

**Table S32** Affiliation of experts invited (n = 150) to participate in the risk ranking assessment.

| **Affiliation** | **Number of Invitations** |
|---|---|
| Government Department | 2 |
| Government Agency | 19 |
| Government Research Institute | 4 |
| International Organization | 9 |
| Private Foundation | 13 |
| Private Organization | 3 |
| University | 100 |

**Table S33** Options provided to the expert opinion panel in a worksheet for the *Level of Risk* each risk factor contributes towards spillover risk, and self-assessed *Level of Expertise* in the factor subject. The scores were allocated to options for use in *Risk Factor Influence* calculations.

| | **Options** | **Score** |
|---|---|---|
| *Level of Risk* | High | 3 |
| | Medium | 2 |
| | Low | 1 |
| | Not Relevant | 0 |
| | Unassigned | 0 |
| *Level of Expertise* | Master | 16 |
| | Expert | 8 |
| | Proficient | 4 |
| | Competent | 2 |
| | Novice | 1 |
| | Unassigned | 8 |

## SI References

1. D. Carroll *et al.*, The Global Virome Project. *Science* **359**, 872-874 (2018).
2. E. J. Dubovi, N. J. Maclachlan, Eds., *Fenner's Veterinary Virology* (Academic Press, Boston, 2016), 5 Ed.
3. C. Tidona, G. Darai, *The Springer Index of Viruses* (Springer, New York, USA, ed. 2, 2011), 10.1007/978-0-387-95919-1.
4. C. J. Burrell, C. R. Howard, F. A. Murphy, *Fenner and White's Medical Virology* (Academic Press, ed. Fifth Edition, 2017), pp. 604.

5. M. E. J. Woolhouse, L. Brierley, Epidemiological characteristics of human-infective RNA viruses. *Sci Data* **5**, 180017 (2018).

6. J. L. Geoghegan, A. M. Senior, F. Di Giallonardo, E. C. Holmes, Virological factors that increase the transmissibility of emerging human viruses. *Proc Natl Acad Sci U S A* **113**, 4170-4175 (2016).

7. C. Hulo *et al.*, ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res* **39**, D576-582 (2011).

8. C. K. Johnson *et al.*, Spillover and pandemic properties of zoonotic viruses with high host plasticity. *Sci Rep-Uk* **5** (2015).

9. Anaconda Software Distribution (2016) Anaconda.

10. P. J. Cock *et al.*, Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423 (2009).

11. IUCN (2017) The IUCN Red List of Threatened Species.

12. Y. Roskov *et al.* (2018) Species 2000 & ITIS Catalogue of Life, 2018 Annual Checklist. (Species 2000, Naturalis, Leiden, the Netherlands).

13. R Core Team (2013) R: a language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria).

14. S. Chamberlain (2016) rredlist: 'IUCN' Red List Client.

15. G. Grothendieck (2017) sqldf: Manipulate R Data Frames Using SQL.

16. H. Wilman *et al.*, EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology* **95**, 2027-2027 (2014).

17. Anonymous, Bird species distribution maps of the world. http://datazone.birdlife.org/species/requestdis.

18. QGIS Development Team (2017) QGIS Geographic Information System. (Open Source Geospatial Foundation Project).

19. B. Vilela, F. Villalobos, letsR: a new R package for data handling and analysis in macroecology. *Methods in Ecology and Evolution* **6**, 1229-1234 (2015).

20. T. W. Bank. (2016) Land area (sq. km).

21. UNData (2007) Land Area.

22. C. I. Agency (2017) The World Factbook 2017. (Washington, DC).

23. World Health Organization (2017) Definition of region groupings.

24. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* **34**, 1812-1819 (2017).

25. F. Nachtergaele, M. Pertri (2011) Mapping land use systems at global and regional scales for land degradation assessment analysis. in *Land degradation assessment in drylands*, ed A. Woodfine (FAO, Rome, Italy), pp 1–19.

26. R. J. Hijmans (2016) raster: Geographic Data Analysis and Modeling.

27. H. Wickham, The Split-Apply-Combine Strategy for Data Analysis. *2011* **40**, 29 (2011).

28. R. Bivand, T. Keitt, B. Rowlingson (2017) rgdal: Bindings for the Geospatial Data Abstraction Library.

29. Center for International Earth Science Information Network - CIESIN - Columbia University, Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals. NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4HX19NJ. Deposited 20170810.

30. K. Seto, B. Güneralp, L. R. Hutyra, Global Grid of Probabilities of Urban Expansion to 2030. NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4Z899CG.

31. K. G. Kees, B. Arthur, v. D. Gerard, d. V. Martine, The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography* **20**, 73-86 (2011).

32. K. Klein Goldewijk, A. Beusen, P. Janssen, Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. *The Holocene* **20**, 565-573 (2010).

33. A. Zvoleff (2015) gfcanalysis: Tools for Working with Hansen et al. Global Forest Change Dataset.

34. M. C. Hansen *et al.*, High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**, 850-853 (2013).

35. M. C. Hansen, S. V. Stehman, P. V. Potapov, Quantification of global gross forest cover loss. *Proc Natl Acad Sci U S A* **107**, 8650-8655 (2010).

36. D.-H. Kim *et al.*, Global, Landsat-based forest-cover change from 1990 to 2000. *Remote Sensing of Environment* **155**, 178-193 (2014).

37. FAO (2015) FRA 2015: Terms and Definitions. in *Forest Resources Assessment Working Paper 180* (Food and Agriculture Organisation of the United Nations).

38. E. J. Lefkowitz, M. J. Adams, A. J. Davison, S. G. Siddell, P. Simmonds (2017) Virus Taxonomy: The Classification and Nomenclature of Viruses.

39. World Health Organization (2020) WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. pp https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

40. J. Antonovics *et al.*, The evolution of transmission mode. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20160083 (2017).

41. S. J. Anthony *et al.*, Non-random patterns in viral diversity. *Nat Commun* **6** (2015).

42. M. Aitichou, S. S. Saleh, A. K. McElroy, C. Schmaljohn, M. S. Ibrahim, Identification of Dobrava, Hantaan, Seoul, and Puumala viruses by one-step real-time RT-PCR. *J Virol Methods* **124**, 21-26 (2005).

43. S. J. Anthony *et al.* (2011) Laboratory Protocols for PREDICT Surveillance. (USAID PREDICT), pp 1-24.

44. A. Kapoor *et al.*, Identification and characterization of a new bocavirus species in gorillas. *PLoS One* **5**, e11948 (2010).

45. M. E. Lozano *et al.*, Characterization of arenaviruses using a family-specific primer set for RT-PCR amplification and RFLP analysis: Its potential use for detection of uncharacterized arenaviruses. *Virus Research* **49**, 79-89 (1997).

46. G. Moureau *et al.*, A real-time RT-PCR method for the universal detection and identification of flaviviruses. *Vector Borne Zoonotic Dis* **7**, 467-477 (2007).

47. P.-L. Quan *et al.*, Identification of a Severe Acute Respiratory Syndrome Coronavirus-Like Virus in a Leaf-Nosed Bat in Nigeria. *mBio* **1**, e00208-00210 (2010).

48. S. M. Raboni *et al.*, Clinical survey of hantavirus in southern Brazil and the development of specific molecular diagnosis tools. *Am J Trop Med Hyg* **72**, 800-804 (2005).

49. M. P. Sánchez-Seco, D. Rosario, E. Quiroz, G. Guzmán, A. Tenorio, A generic nested-RT-PCR followed by sequencing for detection and identification of members of the alphavirus genus. *J Virol Methods* **95**, 153-161 (2001).

50. S. Tong, S.-W. W. Chern, Y. Li, M. A. Pallansch, L. J. Anderson, Sensitive and Broadly Reactive Reverse Transcription-PCR Assays To Detect Novel Paramyxoviruses. *Journal of Clinical Microbiology* **46**, 2652-2658 (2008).

51. S. Wacharapluesadee, T. Hemachudha, Duplex nested RT-PCR for detection of Nipah virus RNA from urine specimens of bats. *J Virol Methods* **141**, 97-101 (2007).

52. S. Watanabe *et al.*, Bat Coronaviruses and Experimental Infection of Bats, the Philippines. *Emerg Infect Dis* **16**, 1217-1223 (2010).

53. J. Zhai *et al.*, Rapid molecular strategy for filovirus detection and characterization. *Journal of clinical microbiology* **45**, 224-226 (2007).

**List of group authors**

[8]Expert panel group members: Kevin Anderson, Prasert Auewarakul, Lark Coffey, Ronald Corley, Gwenaelle Dauphin, Jonathan Epstein, Keiji Fukuda, Simon Goodman, Barbara Han, James Hughes, Martin Jeggo, William Karesh, Rudovick Kazwala, Terra Kelly, Gerald Keusch, Michael Kurilla, John Mackenzie, Wanda Markotter, Corina Monagin, David Morens, Vincent Munster, Elke Muhlberger, Pranav Pandit, Alison Peel, Dirk Pfeiffer, Olivier Restif, Oyewale Tomori, Jonathan Towner, Sylvie Van Der Werf, Sophie VonDobschetz, Supaporn Wacharapluesadee, Michael Ward, Lidewij Weirsma, Mary Wilson, David Wolking, Kachen Wongsathapornchai,

Liam Brierley, University of Edinburgh Epigroup members, Carlos Tambrana Torellio, and those who wish to remain anonymous.

[9]PREDICT Consortium group members: A. Marm Kilpatrick, Abd. Aziz Jamaluddin, Abdoulaye Ousmane Koutate, Abdul Hai, Abdul Kadir Abu Hashim, Abdulhameed Kataregga, Abdullah Al Mamun, Abdullah Al Shakil, Abdullah Alshammari, Abel Ekiri, Aboubacar Bamba, Abraham Commey, Aftab Uddin Rumi, Ageng Wiyatno, Aghnianditya Kresno Dewantari, Ahmed Kandeil, Ahmed Mostafa, Ahmed N. El Taweel, Ahmed S. Kayed, Aiah Gbakima, Ajay Narayan Sharma, Alaa Fahmawi, Alain Ondzie, Albert Ross, Alberto Perez, Aleksei A. Chmura, Alexandra Gibson, Alexandre Tremeau-Bravard, Alexter Japrin, Alice Latinne, Alice Mathew, Alisa Pereira, Allison White, Alonso Aguirre, Alpha Oumar Camara, Alphonce Msigwa, Alvis A. Sartee, Amadou Ndiaye, Amanda Fine, Amara Jambai, Amara Leno, Amethyst Gillis, Aminata Ba, Amira S. El Rifay, Ammar Rafidah Saptu, Amos G. Kollie, Anatole N'télo, Andrew Chow, Andrew Clements, Andrew Ginsos, Angela Yang, Angelica de Almeida Campos, Anna Willoughby, Anne Laudisoit, Antoine Mudakikwa, Ardjouma Dembele, Ariful Islam, Aristide Dionkounda, Ariunbaatar Barkhasbaatar, Arsene Mossoun Mossoun, Asha Makweta, Ashley Lucas, Ashok Chaudhary, Aung Myo Chit, Aung Than Toe, Ava Sullivan, Awa Deme, Bach Duc Luu, Basil H. Amarneh, Batchuluun Damdinjav, Batsikhan Sodnom, Baudelaire Zorine Nkouantsi, Ben Hu, Benard Ssebide, Benoit Goossens, Bilal Al Omari, Bishwo Shrestha, Bouaphanh Khamphaphongphane, Brad Schneider, Brett Smith, Brian Bird, Bridgette Smith, Brooke Genovese, Brooke Watson, Cale Basaraba, Carlos Zambrana-Torrelio, Carlos Zariquiey, Carolina Churchill, Catherine Doyle-Capitman, Catherine Machalaba, Catia de Paula, Celina Roy, Chandrawathani Panchadcharam, Changwen Ke, Charles Chiu, Charles Kumakamba, Chong Chee Kheong, Chris Whittier, Christian Lange, Christina Rundi, Christopher Kilonzo, Citra Livi Kowel, Corina Monagin, Cristin Young, Cynthia Nkoua, Daba Zoumarou, Damien Joly, Dan Marcelin Haba, Dan Rejmanek, Dang Duc Anh, Daniel Chai, Daniel K. Harris, Daniel N'guessan Saraka, Daniel O'Rourke, Daniel Simon, Dao Le Anh, Daouda Ndiaye, David Bunn, David McIver, David J. Wolking, Dawn Zimmerman, Debapriyo Chakraborty, Debbie Mollard, Denis Byaruba, Denise Greig, Dennis Carroll, Desalegn Belay, Dhiraj Puri, Diah Iskandriati, Diana Ramirez, Dibesh Karmacharya, Diego Montecino-Latorre, Djeneba Bamba, Dodi Safari, Dorothy Lim, Douokoro Kalivogui, Dulam Purevtseren, Eddy Kambale Syaluha, Eddy Rubin, Edison Durigon, Ehab Abu-Basha, Eliza Liang, Elizabeth Leasure, Elizabeth Loh, Elizabeth VanWormer, Emilly Sion, Emily Hagan, Emma Mendelsohn, Emmanuel Couacy-Hymann, Emmanuel Larmouth, Emmanuel Mbuba, Emmanuel Tetteh, Enkhtuvshin Shiilegdamba, Eric Delwart,

Eric LeRoy, Eric Mbunwe, Erica Johnson, Erika Alandia Robles, Eugene Kouassi Koffi, Eunah Cho Preston, Evangeline Obodai, Fabien Niama, Fabiola Suárez, Fakhrul Hatta Musa, Fatima Hussein, Felix Nkom, Fernandes Opook, Fernando Nassar, Flavia Miranda, François Lamah, Frankie Thomas Sitam, Frantz Jean Louis, Gafur Sheikh, Ganzorig Baasan, Gaye Laye Diop, Gerard Bounga, Gerardo Suzan, Ghadeer Zghoul, Ghazi Kayali, Ghislain Dzeret Indolo, Gilda Grard, Glenda Ayala Aguilar, Grace Mwangoka, Guangjian Zhu, Hani Talafha, Hannah Chale, Happy Mkali, Harjeet Mann, Heather Wells, Helal Uddin, Helen Lee, Helena Henry, Henri-Joseph Parra, Hermann Assemien Krou, Herminio Ticona, Hoang Bich Thuy, Hongying Li, Huda Dursman, Ian Trupin, Ipos Ngay, Isabel Galarza, Isabel Moya, Isamara Navarrete-Macias, Jaber Belkhiria, Jackson Y. Poultolnor, Jacques Iyanya, Jallah Arku, James Ayukebong, James Bangura, James Brandful, James Desmond, Janusz Paweska, Jasjeet Dhanota, Jason Euren, Jean Claude Tumushime, Jean Michel Takuo, Jean Paul Lukusa, Jean Vivien Mombouli, Jean-Jacques Muyembe Tamfum, Jeffrine Rovie Ryan Japning, Jennifer Lane, Jessica Schwind, Jimmy Lee, Jinnat Ferdous, Jocelyn Perez, Joël Akpaki, Joel Garcia, Joel Judson Jaimin, John Brownstein, Joko Pamungkas, Jonathan H. Epstein, Jordan Levinson, José Luis Mollericona, José R. Ferrer-Paris, Josefina Abedin, Joseph Awuni, Joseph Brown, Joseph Diffo Le Doux, Joseph Fair, Joseph Kamau, Joseph Malakalinga, Joseph Rosario, Joseph Turay, Jules Gomis, Julien Kalpy Coulibaly, Julius Lutwama, Julius Nwobegahay, Julius Nziza, Jum Rafiah Abd Sukor, Jun Liu, June Barrera, Junior Beal Akoundze, Jusuf Kalengkongan, Jyotsna Joshi, Kaisar Rahman, Kalil Doumbouya, Kandeh Kargbo, Karen Saylors, Kate Jones, Kate Thomas, Katherine Leasure, Ken Cameron, Kevin Gonzalez, Kevin Tolovou, Khebir Verasahib, Kidan Araya, Kimashalen Chor, Kimberly Dodd, Komal Jain, Kongsy Khammavong, Kortu M. Ndebe, Kpon Kakeuma Romeo Gueu, Kris Murray, Kristine Smith, Kyaw Yan Naing Tun, Lanash Thanda, Laura Benedict Mugok, Laurentius Ambu, Laurie Harris, Le Tin Vinh Quang, Leanne Wicker, Lee Heng Gee, Leilani Francisco, Lem Fui Fui, Leonoris Lojivis, Leticia Gutierrez, Libiao Zhang, Lim Ming Yao, Linda Boatemaa, Linda J. Lowenstine, Lucy O. Keatts, Luis Aguirre, Luis Barcena, Luz Dary Acevedo, Magassouba N'faly, Mahmoud M. Shehata, Mamadi Camara, Mame Cheikh Seck, Manisha Bista, Manjunatha Belaganahalli, Manuel Cespedes, Manzan Jean N'Guettia, Marc Levy, Marc Valitutto, Marcel Sidiki Kondiano, Marcela Uhart, Margret Bonason, Maria Makuwa, Maria Suleiman, Mariam Kourouma, Marie Pelagie Atrou Gbamele, Mark-Joel Akongo, Martin Gilbert, Ma-Sue Koffa, Mat LeBreton, Mathias Sango, Mathieu Pruvot, Matthew Blake, Maureen Miller, Md. Enayet Hossain, Md. Tarikul Islam, Meerjady Sabrina Flora, Megan Doyle, Megan Mitchell, Megan Vodzak, Mei Ho Lee, Melinda Rostal, Melissa Manhas, Melkor Sackie, Micaela De La Puente, Michael Cranfield, Michael Garbo, Michael Grodus, Michael Westfall, Michel Koropo, Millawati Gani, Min Thein Maw, Mireille Dosso, Misliah Mohamad Basir, Mizanur Rahman, Moctar Mouiche, Modou Moustafa Lo, Moh Moh Htun, Mohamed Ahmed Ali, Mohamed Idriss Doumbouya, Mohammad Borhan Al-Zghoul, Mohammad Saifullah Mohd Azian, Mohammad Yuery Wazlan Abdul Wahad, Mohammed Sidibey, Mohammed Ziaur Rahman, Mohd Isnaim Ismail, Mohd Lutfi Abdullah, Moise Bendoua Tolno, Mokhtar R. Gomaa, Monica Romero, Mouhamed Sy, Moukala Ndolo Hilarion, Moushumi Hassan, Mustafa Ababneh, Mustafizur Rahman, Mwokozi Mwanzalila, Nancy Cavero, Nathan Wolfe, Neal Liang, Nenneh Kamara-Chieyoe, Ngo Thanh Long, Nguyen Duc Thinh, Nguyen Thanh Phuong, Nguyen Thi Diep, Nguyen Thi Hoa, Nguyen Thi Lan, Nguyen Thi Thanh Nga, Nguyen Thu Thuy, Nguyen Van Long, Nicole Gardner, Nicole Ureda, Nigatu Kebede, Nistara Randhawa, Noam Ross, Noorliza Noordin, Nor Adilah Aziz, Norhidayah Danial, Norsharina Arshat, Novie Kasenda, Novie Rachmitasari, Nur Amirah Md Sungif, Ohnmar Aung, Ola Ababneh, Ola Bagato, Olivier Nsengimana, Omnia Kutkat, Onkirotin Dionne Olva, Oscar Rico, Otilia Dogbey, Pacifique Musabimana, Pamela Roualdes, Parviez Hosseini, Patarapol Maneeorn, Patricia Mendoza, Patricia Reed, Patrick Dawson, Paul Horwood, Pe Bhele Maomy, Peta Hitchens, Pete Marra, Pham Thi Bich Ngoc, Phan Quang Minh, Philippe Buchy, Philippe Dussart, Pierre Formenty, Pitu Biswas, Placide Mbala, Prajwol Manandhar, Pranav Pandit, Prateep Duengkae, Prime Mulembakani, Priscilla Joyner, Privat Godji Gnabro, Quaza Nizamuddin Hassan, Qun He, Rabeh El Shesheny, Rachael Mbabazi, Rachmitasari Noviana, Rafael Ojeda-Flores, Rahmat Topani, Rajindra Napit, Ramlan Mohamed, Rebecca Hill, Richard Suu-Ire, Ricky Okwir Okello, Ridzki M.F. Binol, Rima Shrestha, Rock Aimé Nina, Rodrigo Medellín, Rolando Limachi, Romain Bagamboula Mpassi, Ronald Herbert, Rosario Barradas, Rosario Rivera, Rudovick Kazwala, Runie David, Ruth Maganga, Saddam Hossain, Salif Camara, Saman Pradhan, Samath In, Samita Raut, Samsir Laimun, Samson Mutura, Samuel Bel-Nono, Sandra G. Samuels, Sandra Villar, Sara H. Mahmoud, Sayon Yombouno, Schadrack Niyonzima, Sen Nathan, Senthilvel K.S.S. Nathan, Serge Kaba, Shahanaj Shano, Shariful Islam, Shongo Lushima, Shusmita Dutta, Shu-Yi Zhang, Sia Alida Sayandouno, Sief Addeen Al Hanandeh, Simeon Fahn, Simon Anthony, Simone Dramou, Simorn Phon, Sinpakhome Singhalath, Sireeda Miller, Sokha Chea, Soubanh Silithammavong, Stephanie Martinez, Stephen Morse, Stephenie Ann Albart, Supaporn Wacharapluesadee, Suraya Hamid, Suryo Saputro, Sylivia Nakimera, Symphorosa Sipangkui, Tan Jun Ee, Tapan Dhole, Taylor Gabourie, Terra Kelly, Theingi Win Myat, Theodore Asigbee, Thiravat Hemachudha, Tierra Smiley Evans, Tiffany Bogich, Tin Tin Myaing, Tina Kusumaningrum, Titus Joe, Tom Hughes, Tracy Drazenovich, Tran Minh Phuc, Tran Nhu Duong, Tumendemberel Dorjnyam, Tunu

Mwamlima, Ubald Tamoufe, Ulaankhuu Ankhanbaatar, Ungke Antonjaya, Uus Saepuloh, Valchy Bel-Bebi Miegakanda, Valere Kouamé Kouakou, Veasna Duong, Velsri Sharminie, Vibol Hul, Victor Lungay, Victoria Andrew, Victoria Ontiveros, Victoria Pereira, Victoria Sedor, Victorine Maptue, Vo Van Hung, Volga Iñíguez, Vu Sinh Nam, Vu Trong Duoc, W. Ian Lipkin, Wai Zin Thein, Walter Simon Magesa, Watthana Theppangna, Wendy Weisman, Wenjun Ma, William Ampofo, William Karesh, Wirda Damanik, Wivine Mouellet, Woutrina Smith, Xavier de Lamballerie, Xiaoyu Che, Xinglou Yang, Xingyi Ge, Yagouba Kane, Yanne Vanessa Mavoungou, Yasha Feferholtz, Yassmin Moatasim, Yohannes Negash, Yovanna Murillo, Yunzhi Zhang, Zahidah Izzati Zeid, Zaidoun Hijazeen, Zainal Zainuddin, Zena Babu Mchomvu, Zenaa Emmanuel, Zhengli Shi, Zikankuba Sijali, Zoe Greatorex, Zuhair Bani Ismail