Manuscript submitted for peer review and publication on 05.07.2017

Phenology supports the eco-environmental hypothesis for Ebola spillover events

Katharina C. Wollenberg Valero[1*], Raphael D. Isokpehi[2], Noah E. Douglas[2], Seenith Sivasundaram[3], Brianna Johnson[2], Kiara Wootson[3], and Ayana McGill[2]

[1] School of Environmental Sciences, University of Hull, Kingston-Upon-Hull, United Kingdom.
[2] Department of Natural Sciences, College of Science, Engineering and Mathematics, Bethune-Cookman University, Daytona Beach, Florida, United States of America.
[3] Department of Mathematics & Physics, College of Science, Engineering and Mathematics, Bethune-Cookman University, Daytona Beach, Florida, United States of America.
*Corresponding author: k.wollenberg-valero@hull.ac.uk
Authors are ordered by descending level of contribution to the manuscript

Abstract

Ebola virus disease outbreaks in mammals (including humans and great apes) start with sporadic host switches from unknown reservoir species. The factors leading to such spillover events are not clearly understood. Filoviridae have a wide range of natural hosts and are unstable once outside hosts. Spillover events, which involve the physical transfer of viral particles across species, could therefore be directly promoted by conditions of host ecology and environment. In this report we outline a proof of concept that temporal fluctuations of a set of eco-environmental variables describing the dynamics of the host ecosystem are able to predict such events of Ebola virus spillover to humans and animals. We newly compiled a dataset of climate and phenology variables and Ebola virus disease spillovers in humans and animals. We identified critical biotic and abiotic conditions for spillovers via multiple regression and neural networks based time series regression. Phenology variables proved to be overall better predictors than climate variables. African phenology variables are not yet available as a comprehensive online resource. Given the likely importance of phenology for forecasting the likelihood of future Ebola spillover events, our results highlight the need for more phenology monitoring to supply data for predictive modelling efforts.

Key Words: emerging infectious disease, phenology, normalized difference vegetation index, climate, climate change

Summary

Since the 1970s, the Ebola virus recurrently emerges from the Afrotropics. As viruses cannot survive outside their host species, the transfer from species that naturally harbor it to other susceptible species via so-called spillover events might involve the host's ecosystem. Consequently, properties of the ecosystem could be used to predict future spillovers. In this paper, we provide a proof of concept for the idea that plant characteristics such as flowering, fruiting and greening can be used for predicting spillover events. If confirmed, this might serve as a future cost-effective method to locally monitor conditions favorable for spillovers.

Introduction

Ebola Virus Disease (EVD) can emerge simultaneously at geographically distant locations throughout the West African tropical rainforest belt biome. For example, the 1996 wave of outbreaks in both animals and humans occurred within just a few months in remote forest locations in Gabon and the Democratic Republic of Congo (Georges et al. 1999; Milleliri et al. 2004; Yamagiwa and Basabose 2006; Lahm et al. 2007). This quality evoked migrating species such as bats as possible natural reservoir of the virus ((Leroy et al. 2004, 2009). However, while mammals undoubtedly represent an important link in the infection chain, the presence of Ebola virus antibodies in bats is neither particularly high in areas of EVD prevalence (1%-

4% antibody prevalence, (Pourrut et al. 2005), nor is it higher than in humans (1.4%, (Lahm et al. 2007)), nor is antibody prevalence in bats restricted to areas of EVD emergence, evidenced by Ebola and Marburg virus antibodies in bats native to Bangladesh (Olival et al. 2013). Active virus RNA in serum of bats has also not been found yet (Ogawa et al. 2015). One of the suspected reservoir bat species, *Eidolon helvum*, is a migrant, panmictic species that travels annually to and from the known zoonotic ecological niche of the Ebola virus (Richter and Cumming 2006; Pigott et al. 2014), but does not transmit Ebola virus disease (EVD) outside of that range. Known susceptible species that are consumed by humans and known to transmit the disease include, except bats, a variety of mammals, such as chimpanzee, gorilla, and duiker antelopes (Rouquet et al. 2005). Other species susceptible to the disease include elephants, porcupine, genets, mongoose, pangolin, *Tragelaphus* antelope, python, cane rat, hogs, and birds of prey (Leroy et al. 2004; Lahm et al. 2007).

Most emergences of Ebola virus disease have also not been linked to direct human exposure to bats, and the connection between bats and other susceptible species is unclear (Leendertz et al. 2016). Based on known viral load, global predictions do not identify the African tropical rainforest belt biome as hotspots for viral spillovers from bats, but they do identify great apes as likely to mediate viral spillovers (Olival et al. 2017). Since current evidence for migrating bats being the primary natural reservoir of Ebola virus is inconclusive, alternative explanations for the sporadic emergence of Ebola virus disease in different places at the same time should be explored. Phylogenetic distance between susceptible potential host species has recently been used to predict additional susceptible taxa (Weber et al. 2017; Olival et al. 2017). Mononegavirales, the supergroup containing Filoviridae such as the Ebola virus group, include a broad spectrum of natural reservoirs, including many mammals and other tetrapods (Taylor et al. 2010), so that no potential hosts can be *a priori* excluded.

Simultaneous infections of humans and animals in the Gabon-Congo border region between 2001 and 2005 originated from viral strains with different amino acid sequences of the viral GP protein, indicating that these originated from at least five independent, simultaneous spillover events (Leroy et al. 2004; Wittmann et al. 2007). Based on this evidence, and the fact that viruses cannot exist outside of their natural host organisms, previous studies have highlighted a possible eco-environmental dimension to Ebola virus spillover events. (Wittmann et al. 2007) hypothesized that “particular, but unknown environmental conditions” may have caused the multiple spillover events from the unknown host to the human and animal populations recorded to be affected (Peterson et al. 2004).

An environmental niche model has been generated for the Ebola virus based on environmental parameters in its localities of emergence, showing the African tropical rainforest biome as clearly defined region with highest probability of occurrence (Pigott et al. 2014). The 1990s Ebola outbreaks were related to short-term changes in regional precipitation and normalized difference vegetation index (NDVI)(Tucker et al. 2002). Evapotranspiration and enhanced vegetation index are predictive of the spatial environmental niche of Ebola virus (i.e., being located in the African tropical rainforest biome), and that monthly rainfall and rainfall anomaly index can predict Ebola virus (EBOV) spillovers (Pinzon et al. 2004; Schmidt et al. 2017). Besides this spatial dimension, an apparent seasonality of spillovers seems to be associated with wet-to-dry or dry-to-wet transition periods (Tucker et al. 2002; Alexander et al. 2015). Increased spillover events from the first emergence of the virus in the 1970s towards the present, might be linked to human intrusion into the tropical forest biome and land use change for agricultural purposes Gottdenker et al., 2014).

Intraspecific or interspecific competition for resources such as food, or environmental contamination within the virus-harboring ecosystem can serve as a hypothetical functional links between eco-environmental factors and spillover events between the unknown natural reservoir and other species. Rouquet and co-authors (Rouquet et al. 2005) hypothesized that direct competition for

resources between susceptible species in periods of food scarcity leads to increased inter-species transmission (Rouquet et al. 2005).The annual fruit bat migration into the Democratic Republic of the Congo between April-May may be related to human disease emergence events in June (Leroy et al. 2009). A new mechanistic hypothesis invokes "viral rain", the shedding of infectious viral particles from the natural reservoir into the environment, such as bats spreading Hendraviruses around the trees they roost on (McFarlane et al. 2011).

Besides seasonal fluctuations, and despite annual bat migration and high rates of human consumption of bushmeat, Ebola virus disease does not emerge annually. Therefore, advising local populations against the consumption of bushmeat in general, are unrealistic – as bushmeat constitutes an important food source, and meat consumption outside of periods of Ebola virus disease outbreaks is common (Alexander et al. 2015). Still, efforts to predict Ebola outbreaks exclusively rely on monitoring on the level of animal species commonly hunted as bushmeat, and first incidences in humans (Rouquet et al. 2005) This type of monitoring practice is costly and associated with complicated logistics. Designing novel methods to forecasting the likelihood of future Ebola virus disease spillover events before the human-to-human transmission stage is reached, would be a cost-effective method of mitigation (Schar and Daszak 2014). For example, a recent report makes the case that any emerging infectious diseases can be predicted by a model including spatial information and potential mammalian host species (Olival et al. 2017).

In this article, we evaluate support for the eco-environmental or multi-emergence hypothesis from published ecological and environmental data as predictive for Ebola virus disease spillover events. We generate a novel, detailed dataset of both human and other mammal spillover events, and integrate it with local climate data gathered from the previously identified area of high occurrence probability, and with locally collected data on phenology. A comprehensive African phenology database does not yet exist (Adole et al. 2016), so that we computationally collected such relevant data from graphs published in ecological studies of phenology to make them available for analysis as an online resource. Our working hypothesis is, that locally collected phenology variables (greening, flowering, fruiting) will be better predictors for Ebola virus disease spillovers than remotely sensed climatic variables currently used for predictive modelling.

Materials and Methods

Generation of datasets. *Data on the occurrence of Ebola virus spillover events were obtained from published reports. Therefore, the true number of spillover events (especially in animals) might be higher, and we have to consider the dataset we compiled as a sample of this true number. Data for spillover events involving humans was obtained from the website of the United States Centers for Disease Control and Prevention (cdc.gov, accessed 2014-2017). As the CDC website only lists reports for human outbreaks, we screened literature sources for information on other species' Ebola infection records to also record outbreaks in non-human mammals (International Commission 1978; Heymann et al. 1980; Le Guenno et al. 1995; Muyembe and Kipasa 1995; Amblard et al. 1997; Formenty et al. 1999; Georges et al. 1999; Khan et al. 1999; Okware et al. 2002; Leroy et al. 2004, 2009; Lamunu et al. 2004; Milleliri et al. 2004; Rouquet et al. 2005; Pourrut et al. 2005; Lahm et al. 2007; Wittmann et al. 2007; Onyango et al. 2007; Towner et al. 2008; Wamala et al. 2010; MacNeil et al. 2010; Nkoghe et al. 2011; Shoemaker et al. 2012; Bausch and Schwarz 2014; Baize et al. 2014). (Wittmann et al. 2007; Leroy et al. 2009) reported on molecular screens for Ebola virus strains for the 2001-2005 outbreaks, we follow their argumentation and consider each confirmed independent epidemic chain as independent event. The data was recorded monthly, and we computed arithmetic means for each year, as well as for each month over the year. In total, our dataset includes 57 human and other mammal spill-over events between 1976 and 2017.*

Climate data was obtained from the Global Historical Climatology Network (GHCN, NOAA accessed 2015-2016). Data was obtained for monthly summaries between 1953-2015 for the weather station GHCND-GB000004556 in Makokou, Gabon (Lat 0.578004°, Long 12.889751°), and for the years 2016-2017 for the location FIPS:GB 9 (Libreville, Gabon, Lat 0.457088°, Long 9.409372°). We obtained data for the following variables: DT90, Number of days with maximum temperature ≥32.2°C; DP01, Number of days with ≥0.254 mm rainfall per month; DP05, Number of days with ≥12.7 mm rainfall per month; DP10, Number of days with ≥25.4 mm rainfall per month; CLDD, cooling degree days, computed when daily average temperature is more than 18.3°C; CDD, days per month with mean daily temperature - 18.3°C; EMNT, Lowest daily minimum temperature for the month in °C; EMXP, highest daily total of precipitation per month in mm; EMXT – Highest daily maximum temperature

per month in °C; MMNT, monthly mean minimum temperature in °C; MMXT, monthly mean maximum temperature in °C; MNTM, monthly mean temperature in °C; TPCP total precipitation, in mm/10, for the month. From this monthly dataset, we calculated arithmetic means for each year, as well for each month over the 1953-2017 period. We added additional climate data from other published sources. These datasets were computationally collected from published graphs using the WebPlotDigitizer software (V 3.12) (Rohatgi 2017). Average annual values and for average rainfall (1941-2000), and average monthly minimal and maximal temperatures (1976-2001) for Kibale National Park, Uganda (Lat 0.486621°, Long 30.390715°) were obtained from Chapman et al. (Chapman et al. 2005). From Polansky and Boesch (Polansky and Boesch 2013), we extracted monthly inter-annual linear trends of rainfall, average monthly rainfall in Central Africa (in mm), between 1979-2010 in Tai National Park, Ivory Coast (Lat 5.690020°, Long -6.939547°). We also extracted temperature for African humid tropics in °C (Fisher et al. 2013), as well as the averaged normalized departure (Apr-Oct rainfall departure) between 1979-2010 (Boyd et al. 2013).

Currently, no comprehensive database exists for African phenology and vegetation data (Adole et al. 2016). The African tropical rainforest belt, which has high occurrence probability for Ebola virus disease emergence (Pigott et al. 2014), is characterized by a regional, seasonal pattern of greening, fruiting and flowering linked to the cyclical West African Monsoon (Cornforth 2013). Such data sets are crucial to testing the eco-environmental hypothesis for Ebola virus disease spillover events, but not readily available online. We used WebPlotDigitizer (Rohatgi 2017) to collect quantitative phenology datasets in published reports from the area of high occurrence probability for the Ebola virus. From Plumptre et al., (Plumptre et al. 2012), we extracted data for the 1986-2012 flowering anomalies in Lope, Gabon (Lat -0.441156°, Long 11.524798°). Average monthly values were obtained from the same source for the proportion of trees flowering and the proportion of trees with ripe fruit in Goualougo, Nouabale Ndoki Park, Republic of the Congo (Lat 2.263195°, Long 16.591688°); the proportion of trees flowering; and the proportion of trees bearing ripe fruit in Lope, Gabon; the proportion of trees flowering and the proportion of trees bearing ripe fruit in Okapi National Park, Democratic Republic of the Congo (Lat 1.400931°, Long 28.577619°). We also obtained values for the annual proportion of trees fruiting in Kibale NP, Uganda 1970-2001 (Chapman et al. 2005). With an average monthly resolution, we extracted data for the percentage of trees with ripe fruit in Kibale National Park, Uganda between August 1990-September 2002 (Chapman et al. 2005). Arithmetic means were calculated for each year. From Philippon et al. (Philippon et al. 2007), we extracted data for the NDVI anomaly for West Africa including the Sahel zone (between Aug-Dec and between Jul-Dec 1982-2002). From Polansky and Boesch (Polansky and Boesch 2013), we also extracted the percentage of the forest community bearing fruit, and the count of number of species with peak fruiting times per month. Additional variables extracted (monthly, between 1994 and 2002, and annual averages) from graphs generated by Yamagiwa et al. (Yamagiwa et al. 2008), were a "fruit index" for fruit species consumed by Chimpanzees and Gorillas in both primary and secondary forest, in Kahuzi-Biega National Park, Democratic Republic of Congo (Lat. -1.963260°, Long 28.018609°). We further extracted data from the same publication, for the percentage of different plant species bearing fruit that contributed to the generation of the fruit indices (monthly data between 1998 and 2002, and annual arithmetic means). These include: Allophylus spp., Bridelia bridelifolia, Cassipourea spp., Diospyros honleana, Ekebergia capensis, Ficus oreodryadum, Ficus thonningii, Maesa lanceolata, Myrianthus holstii, Newtonia buchananii, Psychotria palustris, and Syzygium parvifolium.

Data Analysis. *Our comprehensive dataset includes three partitions with annual, seasonal, and monthly resolution, and*

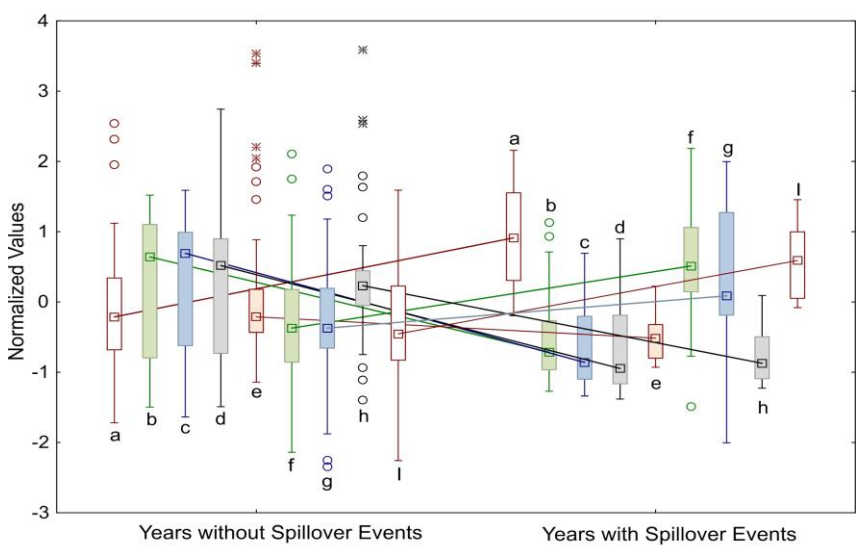

Figure 1. Box plots for climate and phenology variables with significant inter-annual variation (See Table 1 for statistical results) between years with recorded Ebola human + animal spillover events and years without known Ebola Virus spillover events spanning 1953-2016. Boxes represent 25-75% interval, whiskers represent non-outlier range, circles represent outliers, and stars represent extreme values. a) 10 Year average rainfall in Kibale National Park b) DP01, c) DP05, d) DP10, e) EMXP, f) EMNT, g) MMNT, h) TPCP, i) IC2 NDVI anomaly (Jul-Dec).

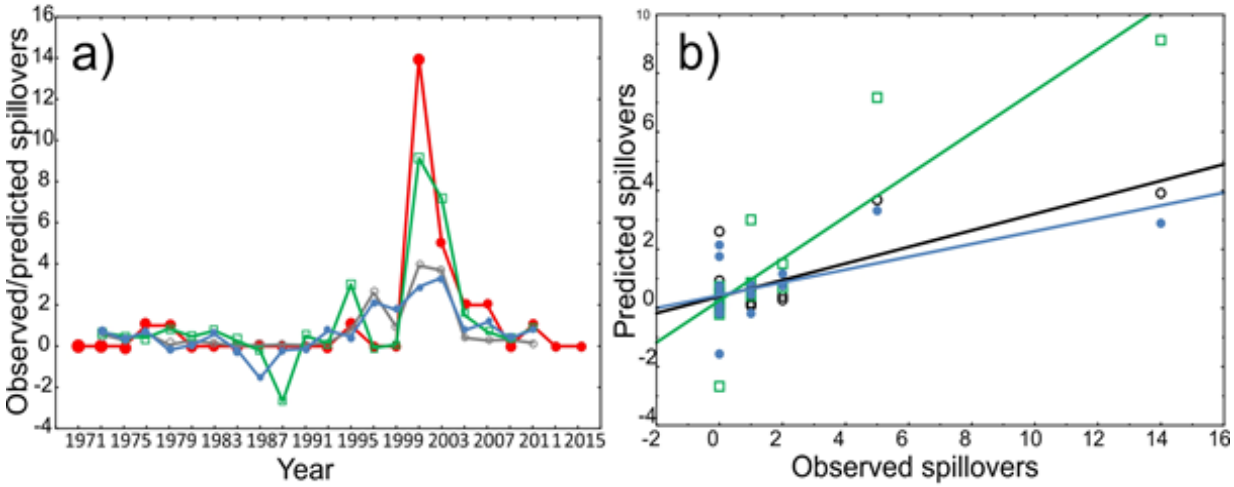

Figure 2. Performance for annual time series regression of animal/human spillover events from Neural Network modeling (SANN) using only climate PCs, only phenology PCs, and climate and phenology PCs combined as sets of input variables. Panel a) shows time series predictions for the observed number of human and animal spillover (closed large circles/red), vs. the predicted numbers from best retained models (500 network iterations) of climate variables only (small closed circles/blue), phenology variables only (open squares/green), and climate/phenology combined as input data sets (open circles/grey). Panel b) shows correlation scatter plots between observed spillovers and predicted spillovers using models with inputs: climate variables only (small closed circles/blue), phenology variables only (open squares/green), and climate/phenology combined as input data sets (open circles/grey). Correlation with Climate+Phenology model r=0.7663, p=0.00008; Correlation with only climate model r = 0.6254, p=0.0032; Correlation with only phenology r = 0.8860, p≤0.000001.

is made available as Supplemental data file (will be published upon manuscript acceptance and is, until then, available upon request). Because of the different resolutions, we analyzed each partition separately.

Analysis of average annual data partition. *Based on the availability of data for the respective variables, we were able to analyze differences in climate and phenology variables between years without any spillover events, and years with at least one spillover event. The climate variables included were 10 yr average rainfall Kibale, avg. monthly max. temp Kibale (annual average) DT90, DP01, DP05, DP10, CLDD, EMNT, EMXP, EMXT, MMNT, MMXT, MNTM, and TPCP. Phenology variables included were first and second independent components of NDVI (See (Philippon et al. 2007) for details). First we performed Kolmogorov-Smirnov tests on each quantitative variable to test for normality. For variables that were distributed normally, we subsequently performed univariate ANOVAs. For variables not normally distributed, we performed Kruskal-Wallis tests between years with and without spillover events. We subsequently box-cox transformed non-normally distributed variables. We then tested whether climate and phenology variables can predict the number of human and other mammal Ebola spillover events (according to the multi-emergence hypothesis). To this purpose, we performed a multiple regression in STATISTICA (Dell, Inc.) with number of spillover events as dependent variable, and climate and phenology variables as independent predictors. To further test whether phenology, or climate variables are better predictors for the recorded human + other animal spillover events, we generated and compared neural networks in Statistica (Statistica automated neural networks, SANN). First, principal components (PCs) were generated for 16 climate variables (YR_Clim PC1-5 with cumulative explained variance of 83.1%, Supplementary Table 1, 2), and for four phenology variables (YR_Pheno PC1-2 with cumulative explained variance of 66.3%, Supplementary Table 3,4, for the time frame between 1970-2002. Subsequently, the data set was partitioned into a modelling dataset (years with even numbers) and a cross-validation data set (years with odd numbers). On the modelling dataset (70% training, 15% test, 15% validation data), we generated three models based on neural network time series regression, with different input variables, and the number of spillover events as target variable. First, a model was generated with climate + phenology PCs. Second, we generated a model using only phenology PCs. Third, we generated a model using only climate PCs as input. For model generation, we used the MLP (Multilayer Perceptron, (Rumelhart et al. 1986)) method with 500 networks and the 5 best networks retained. We applied weight decay to prevent overfitting and used 1 step for forecasting. For each analysis, the best model was chosen and subsequently deployed to the cross-validation dataset. We then compared the model estimates for the number of spillover events with the recorded number of spillover events in the cross- validation data set. Best-fit to the recorded number of spillover events was determined with correlations and t-test for dependent samples.*

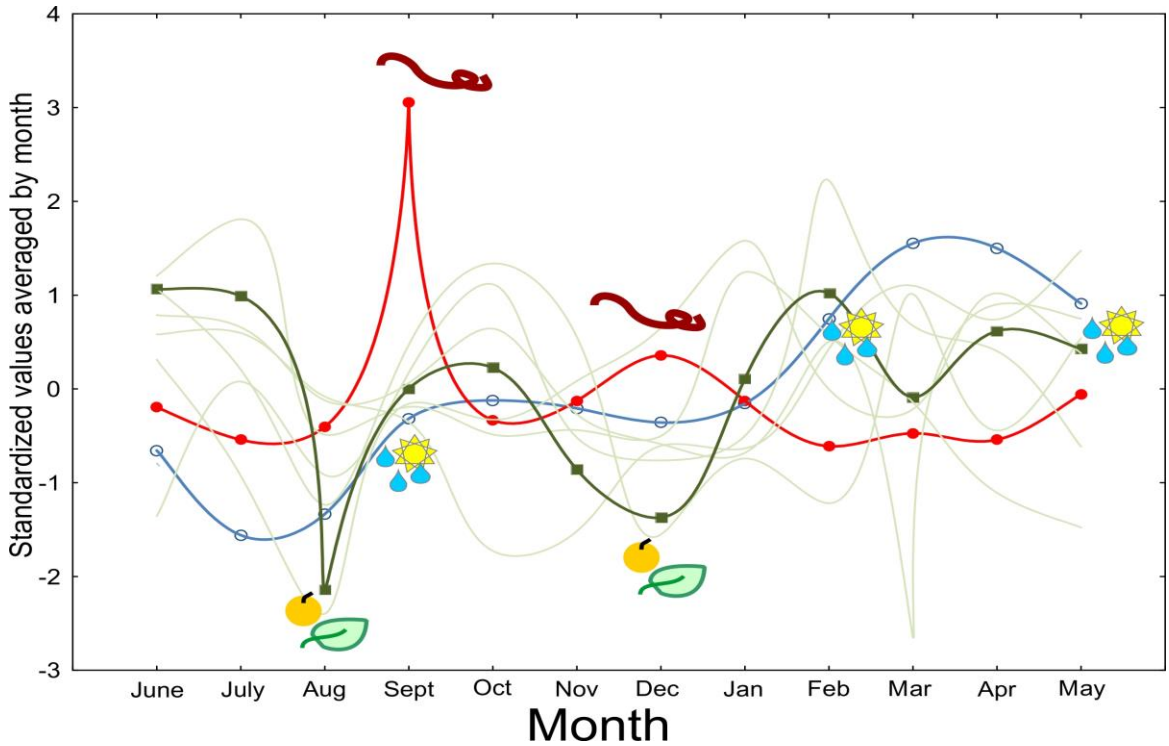

Figure 3. Seasonality of Ebola virus spillover events (humans + animals, standardized values, closed circles), the standardized average of climate PCs 1-3 (open circles) and the standardized average of phenology PCs 1-6 (closed squares). Standardized values for single phenology PCs 1-6 are shown as background lines (no symbols). Both September and December spikes in recorded human + animal Ebola virus spillover events (virus symbols) are associated with lower values of the average of phenology PCs 1-6 (fruit/leaf symbols), but only for the September spike with average climate variable changes (weather symbols). Curves are spline-fitted.

Analysis of seasonal data partition. *Monthly averages across years were available from a set of climate and phenology variables to determine seasonal patterns in the data. To reduce dimensionality in the data for analysis, we performed a principal component analysis on this data set. Thirteen climate variables were reduced to three factors (in following called S_Clim PC1-3). Cumulative variance explained by these variables was 87.6% (Supplementary Tables 5, 6). 25 phenology variables were reduced in dimensionality to six principal components (S_Pheno PC 1-6), together explaining 93.2% of overall variance of this data set (Supplementary Tables 7,8). To explore the seasonal progression of these variables relative to one another, we plotted the number of spillovers per month, together with the average of climate PC values, and the average of phenology PC values.*

Analysis of monthly data partition. *Between 1994 and 2002, monthly data was extracted for fruiting phenology, additionally to climate variables. For analysis, the monthly data partition was therefore cropped to only include monthly values from the years 1994-2002. To reduce dimensionality of climate and phenology variables, three principal*

Table 1. Significant differences in climate and phenology between years with no recorded spillover, and years with spillover of Ebola virus disease (to humans and animals). Shown are results of Kolmogorov-Smirnov tests (D-statistic, p) for variable normality, ANOVA (F-statistic, p, in the case of normality) or Kruskal-Wallis tests (H-Statistic, p in the case of non-normality). For variable abbreviations see Methods section.

Variable	K-S Test D	K-S Test p	ANOVA F	ANOVA p	KW-H	KW p	N Years
10 yr avg Rainfall Kibale	0.1024	n.s.	10.0263	0.0025			60
DP01	0.1595	<0.1			5.4921	0.0191	59
DP05	0.1405	<0.2			8.8842	0.0029	59
DP10	0.1335	n.s.	10.54	0.0020			59
EMNT	0.0693	n.s.	10.158	0.0024			55
EMXP	0.3016	<0.01			10.3339	0.0013	55
MMNT	0.125	n.s.	6.1472	0.0163			57
TPCP	0.1972	<0.05			19.911	0.00001	55
IC2 NDVI anomaly(Jul-Dec)	0.173	n.s.	5.3154	0.0326			21

component analyses were performed. Firstly, the eight climate variables were reduced to three principal components (M_Climate PC1-3), that together explained 81.3% of the overall variance in this data set (See Supplementary Tables 9 and 10). From the set of four fruit index variables of Yamagiwa et al. (Yamagiwa et al. 2008), two principal components were extracted (M_Pheno PC1-2, see Supplementary Tables 11 and 12), explaining 64.4% of overall variance. A principal component analysis for 12 single species variables recovered 5 principal components (M_Plant PC1-5) explaining 72.7% of overall variance (Supplementary Tables 13 and 14) To test whether phenology and climate variables are predictive for Ebola spillover events in the monthly data partition, we performed a multiple regression.

Results

Interannual patterns. Results of tests for normality for annual climate and phenology variables are shown in Table 1. Univariate ANOVAs (for normally distributed variables) and Kruskal-Wallis tests (for non-normally distributed variables, respectively, revealed significant differences in the climate variables 10 year avg. rainfall Kibale, DP01, DP05, DP10, EMNT, EMXP, MMNT, and TPCP between years with versus years without spillover events. Of the two phenology variables, IC2 NDVI significantly differed between the years with vs. without spillover events. Figure 1 shows box plots for these variables. Regression of climate and phenology variables against the number of human + other animal spillover events was overall significant (ANOVA results: R = 0.6020; R^2 = 0.36241524, Adjusted R^2=0.2309; F(13,63) = 2.7546, p < 0.00374, Table 2). The residuals were distributed normally (not shown). Significant univariate predictors for this model were EMXT (p=0.049), and IC2 NDVI (p=0.012). To further determine the effectiveness of phenology versus climate variables in predicting spillover events, we generated and compared neural networks on principal components for the two types of predictor variables separately (phenology only, climate only), and combined (phenology and climate). Performance of each of the best models is shown in Supplementary Table 15. We do not aim for predictive forecasting with these models here - instead, we aimed to test for best fit between modeled data using different sets of input variables and observed data. For this purpose, we employed a paired t-test. When deploying each best model for the three sets of input variables, no model significantly differed from the observed values of Ebola spillover events, indicating an overall good fit (Figure 2a, results of t-test shown in Supplementary Table 16). The model with only phenology PCs as input variables performed best when deployed to the cross-validation data set, with the best correlation coefficient between predicted and observed spillover events (Figure 2b, r = 0.8860, p≤0.000001).

Seasonal patterns. Using average values of phenology and of climate principal components,

Table 2. Regression of climate and phenology variables against annual spillover events in humans and animals. Variables were normalized with Box-cox transformation. The overall model is significant. ANOVA results: R= 0.6020 R^2= 0.3624 Adjusted R^2= 0.2309 F(13,63)=2.7546 p<.00374 Std.Error of estimate: 1.8584. Significant univariate predictors are shown in bold.

N=77	b*	Std.Err. of b*	b	Std.Err. of b	t(63)	p-value
Intercept			-13.9136	19.5252	-0.7126	0.4787
CLDD	-0.2978	0.2060	-0.0009	0.0006	-1.4456	0.1533
EMNT	0.2754	0.2463	0.0805	0.1675	1.1183	0.2677
EMXP	0.0706	0.2916	1.6249	6.7085	0.2422	0.8094
EMXT	**0.3969**	**0.1979**	**0.1459**	**0.0728**	**2.0053**	**0.0492**
MMNT	-0.0754	0.3221	-0.0294	0.1256	-0.2340	0.8157
MMXT	-0.7001	0.4910	-0.2098	0.1472	-1.4258	0.1589
MNTM	0.1837	0.4901	0.0702	0.1874	0.3749	0.7090
TPCP	-0.1259	0.3330	-0.0557	0.1473	-0.3781	0.7066
%of trees with ripe fruit Kibale	0.1500	0.1265	0.4504	0.3796	1.1865	0.2399
Proportion of population fruiting in Kibale	-0.0439	0.1445	-0.0169	0.0557	-0.3039	0.7622
NDVI anomaly (Aug-Nov)	-0.1416	0.1257	-0.5641	0.5007	-1.1265	0.2642
NDVI anomaly (Jul-Dec)	**0.3807**	**0.1468**	**1.4756**	**0.5691**	**2.5931**	**0.0118**
Flowering Anomalies Lope Gabon	0.1917	0.1161	0.1351	0.0818	1.6509	0.1037

line plots for seasonality (Figure 3) show a (previously reported) pattern of seasonality of Ebola spillover events. Our strategy to record both human and other animal spillover events, as well as recording independent epidemic chains as separate spillover events, reveals a clear peak of recorded spillovers in September and a smaller peak in December (Figure 3). As previously reported, the September peak is associated with a change in average climate values, but the second peak in December is not. In contrast, both peaks of Ebola spillover events are clearly associated with in average low phenology values. This clear association between Ebola spillover events and phenology was not previously reported and corroborates the newly found close association between inter-annual Ebola spillover events and interannual variation in vegetation and phenology variables.

Monthly patterns. The Yamagiwa et al. (Yamagiwa et al. 2008) data set allowed for a detailed investigation into which kinds of plants are causing the inter-annual and seasonal variation in phenology that we found to be associated with Ebola spillover events. Multiple regression revealed that for the investigated time period (which includes a total of 56 spillover events), climate variation does not significantly predict the number of monthly spillovers. While the "fruit index" variables also were not significant predictors, three of the five plant species PCs (M_Plant PC 2-4) significantly predicted spillover events (Table 3). The overall regression model was significant (ANOVA for overall goodness of fit: R= 0.4867, R^2= 0.2368 Adjusted R^2= 0.1466, F(11,93)=2.6238 p<0.00583, Std.Error of estimate, 0.9740). Variables with significant factor loadings for these three PCs were the proportion of trees bearing fruit be: *Ekebergia capensis, Ficus oreodryadum, Ficus thonningii*, and *Newtonia buchananii* (Supplementary Table 13). A comparison of means and errors for the proportion of these plants fruiting for the analyzed time series 1993-2002 is shown in Figure 4.

Discussion

We found significant differences in climate and vegetation parameters between years with and years without Ebola virus spillover events. Years with spillover events were characterized by generally drier conditions (DP01, DP05, EMXP, TPCP), as well as higher minimum temperatures (EMNT, MMNT). In contrast, the anomaly of the vegetation index in the months from July - December was higher, combined with a higher average of rainfall (measured in Kibale, Uganda). An overall model of climate and phenology variables as predictors for the number of human + other animal spillovers in a given year was significant. EMXT, the highest monthly daily maximum temperature, was a significant predictor for the number of Ebola spillover events in humans and animals. These results align to previous studies that have found a climatic dimension for Ebola spillover events (Tucker et al. 2002; Pigott et al. 2014; Schmidt et al. 2017). In our study, however, we could additionally show that phenology variables (including the anomaly of NDVI between July and December, the proportion of population fruiting in Kibale NP, Uganda, and the flowering anomalies in Lope, Gabon), informed

Table 3. Multiple regression analysis for human + animal spillover events against climate, fruit, and plant variables (Monthly data partition). The overall model was significant (ANOVA for overall goodness of fit: R= .48666117 R^2= .23683909 Adjusted R^2= .14657275, F(11,93)=2.6238 p<.00583 Std.Error of estimate: 0.97403).

	b*	Std.Err. of b*	b	Std.Err. of b	t(93)	p-value	Valid N
Intercept			-0.2323	0.4198	-0.5532	0.5814	
M_Climate PC1	-0.0458	0.1003	-0.0483	0.1058	-0.4562	0.6493	108
M_Climate PC2	-0.1105	0.0922	-0.1165	0.0972	-1.1983	0.2338	108
M_Climate PC3	0.0727	0.0957	0.0767	0.1009	0.7595	0.4495	108
M_Fruit PC1	-0.0667	0.1161	-0.0703	0.1224	-0.5747	0.5669	108
M_Fruit PC2	0.1316	0.2026	0.1387	0.2136	0.6495	0.5176	108
M_Plant PC1	-0.1538	0.0972	-0.1622	0.1025	-1.5822	0.1170	108
M_Plant PC2	**0.2669**	**0.1049**	**0.2813**	**0.1106**	**2.5428**	**0.0126**	108
M_Plant PC3	**-0.2591**	**0.0958**	**-0.2731**	**0.1010**	**-2.7050**	**0.0081**	108
M_Plant PC4	**-0.2618**	**0.1054**	**-0.2760**	**0.1111**	**-2.4834**	**0.0148**	108
M_Plant PC5	-0.0200	0.0930	-0.0211	0.0981	-0.2154	0.8299	108

neural network models with a superior fit to the data than when climate or climate in conjunction with phenology variables were used as inputs. While some of our phenology variables were correlated with climate variables across all partitions of data (inter-annual, seasonal, monthly), indicating the existence of phenological seasons, previous studies using NDVI or Enhanced Vegetation Index have not yet considered their predictive power for Ebola spillover events independently from climate variables (see for example (Schmidt et al. 2017)).

Aligned to previous reports (Alexander et al. 2015), we found a pronounced seasonality in recorded human and other mammal spillover events. This seasonal association between Ebola spillover events and a multitude of phenology variables describing seasonality of flowering and fruiting was not previously reported, and corroborates the newly found close association between inter-annual Ebola spillover events and interannual variation in vegetation and phenology variables. The seasonality in spillover events is thought to be associated with transitions between wet and dry or dry and wet conditions (Tucker et al. 2002; Pinzon et al. 2004; Groseth et al. 2007; Altizer et al. 2013; Schmidt et al. 2017), and thus may be associated with the African Monsoon (Cornforth 2013). While the September spike in human + other animal spillover events was associated with such a transition of climatic variables, the December spike was not. Other time points with shifts in climatic conditions were vice versa not associated with increases in Ebola spillover events. In contrast, both spikes in Ebola spillovers are associated with lower values in phenology variables, lending additional support to the eco-environmental hypothesis. Previous studies have found seasonal and yearly oscillations of fruiting status within specific localities, with a trend that is generated by differences in peak fruiting times among individual species (Anderson et al. 2005; Polansky and Boesch 2013). Most species fruit only for a month time per year (Anderson et al. 2005). Fruit abundance of species within the Tai forest (Ivory Coast) plant community has been found to increase over the past decade, with no apparent correlation with local climatic patterns

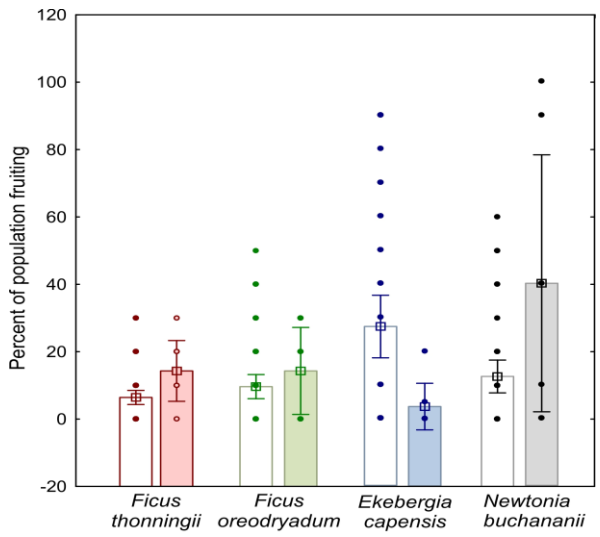

Figure 4. Changes in the proportion of plants bearing fruits between months with no recorded spillovers (open bars) and months with recorded spillovers (closed bars). Dots represent raw data points, column height means, bars represent errors. Plant species shown here were significant predictors in the regression model for number of spillovers per month for the period 1993-2002 in Kahuzi-Biega National Park.

(Polansky and Boesch 2013). Basabose (Basabose 2002) also reported significant differences in the mean percentage of trees fruiting at Kahuzi-Biega National Park between 1994 and 2000, both between the dry and the wet season, as well as between years. Data from both these localities were part of our analysis. Local flowering and fruiting patterns consequently might reflect components of the local environment of the Ebola virus reservoir that differ from climatic variables.

Without in-depth knowledge of the proximate mechanism of Ebola spillovers, we could here show a covariation between ecological variables related to food resources, and Ebola spillovers to mammalian hosts. For example, “Virus shedding” (McFarlane et al. 2011; Plowright et al. 2015), the process of bats transmitting virus particles to plant matter which is then uptaken by other animals, provides a possible functional link between phenology and virus spillover. Phenology variables that significantly predicted monthly spill-over events included *Ficus spp.* fruit, leaves, and stems which are consumed by Gorilla and Chimpanzee. *Ekebergia capensis* fruit and *Newtonia buchananii* seed are likewise consumed by Gorilla and Chimpanzee in Kahuzi-Biega National Park (Democratic Republic of Congo, (Yamagiwa and

Basabose 2006)). *Ficus thonningii* constitutes an important food source for African frugivores (Bleher et al. 2003; Kirika et al. 2008). Seasonally dry climate has shown to cause synchronicity for flowering, but not fruiting in this species (Damstra et al. 1996). It has been proposed as a keystone species for frugivores, as it bears fruit during periods of fruit scarcity (Bleher et al. 2003). We could here show an association between Ebola spillovers and phenology of plants that constitute a seasonally varying food source for susceptible species. Plant species like *F. thonningii* could serve as preliminary focal species to test the hypothesis that competition for fruit or other plant matter in dry periods increase the likelihood of spillover events.

Our results on the link between Ebola spillovers and forest phenology is mirrored by results from another viral hemorrhagic fever in Belgium; There, dynamics of vegetation phenology and their alteration under climate change have been found to influence the dynamics of Puumala virus (causing the disease Nephropathia endemica, (Barrios et al. 2010)). In conclusion, both (Barrios et al. 2010) and our present study support the eco-environmental hypothesis for virus spillover events. The phenology source data analyzed in this study, however, were far from comprehensive. Clearly, to improve the understanding of the importance of phenology for predicting disease emergence, more data are needed. Phenology can vary locally, and seasonally (Bleher et al. 2003), so that this information cannot be replaced by remote-sensing data. We therefore highlight the importance of ecological field work additional to remote sensing in providing important data on phenology variables related to the ecosystem ecology of the unknown natural reservoir of Ebola virus. Local fruiting patterns that may constitute a useful indicator for the likelihood of impending Ebola spillover events and may readily be classified through local and cost-effective transect surveys. The most recent Ebola outbreak in May 2017 in Likati (Bas-Uele Province, Democratic Republic of Congo) suggests that the emergence of Ebola virus is a recurring event that necessitates the generation of upstream models for disease management sensitive to economic cost and cost of human lives (Schar and Daszak 2014). We therefore encourage other researchers to make unpublished African phenology data available to the public domain to aid in understanding this dimension of eco-environmental variables related to emerging infectious disease.

Acknowledgements

A 2015 Mathematics-Biology Data Analytics Workshop at Bethune-Cookman University provided the venue for initial discussion, planning and data generation for this project. We thank all student and faculty participants for valuable discussions.

Funding Statement

The Mathematics-Biology Data Analytics Workshop was funded by National Science Foundation award HRD-1435186.

Conflicts Statement

The authors declare no conflict of interest.

References

Adole T, Dash J, Atkinson PM (2016) A systematic review of vegetation phenology in Africa. Ecol Inform 34:117–128.

Alexander KA, Sanderson CE, Marathe M, et al (2015) What factors might have led to the emergence of Ebola in West Africa? PLoS Negl Trop Dis 9:e0003652.

Altizer S, Ostfeld RS, Johnson PTJ, et al (2013) Climate change and infectious diseases: from evidence to a predictive framework. Science 341:514–519.

Amblard J, Obiang P, Edzang S, et al (1997) Identification of the Ebola virus in Gabon in 1994. Lancet 349:181–182.

Anderson DP, Nordheim E, Moermond TC, et al (2005) Factors influencing tree phenology in taï national park, Côte d'Ivoire. Biotropica 37:631–640.

Baize S, Pannetier D, Oestereich L, et al (2014) Emergence of Zaire Ebola virus disease in Guinea. N Engl J Med 371:1418–1425.

Bärbel Bleher, Potgieter CJ, C. J. Potgeiter, et al (2003) The Importance of Figs for Frugivores in a South African Coastal Forest. J Trop Ecol 19:375–386.

Barrios JM, Verstraeten WW, Maes P, et al (2010) Satellite derived forest phenology and its relation with nephropathia epidemica in Belgium. Int J Environ Res Public Health 7:2486–2500.

Basabose AK (2002) Diet composition of chimpanzees inhabiting the montane forest of Kahuzi, Democratic Republic of Congo. Am J Primatol 58:1–21.

Bausch DG, Schwarz L (2014) Outbreak of ebola virus disease in Guinea: where ecology meets economy. PLoS Negl Trop Dis 8:e3056.

Boyd E, Cornforth RJ, Lamb PJ, et al (2013) Building resilience to face recurring environmental crisis in African Sahel. Nat Clim Chang 3:631–637.

Chapman CA, Chapman LJ, Struhsaker TT, et al (2005) A Long-Term Evaluation of Fruiting Phenology: Importance of Climate Change. J Trop Ecol 21:31–45.

Cornforth RJ (2013) West African Monsoon 2012. Weather

68:256–263.
Fisher JB, Sikka M, Sitch S, et al (2013) African tropical rainforest net carbon dioxide fluxes in the twentieth century. Philos Trans R Soc Lond B Biol Sci 368:20120376.
Formenty P, Hatz C, Le Guenno B, Stoll A (1999) Human infection due to Ebola virus, subtype Cote d'Ivoire: clinical and biologic presentation.
Georges AJ, Leroy EM, Renaut AA, et al (1999) Ebola hemorrhagic fever outbreaks in Gabon, 1994-1997: epidemiologic and health control issues. J Infect Dis 179 Suppl 1:S65–75.
Groseth A, Feldmann H, Strong JE (2007) The ecology of Ebola virus. Trends Microbiol 15:408–416.
Heymann DL, Weisfeld JS, Webb PA, et al (1980) Ebola hemorrhagic fever: Tandala, Zaire, 1977-1978. J Infect Dis 142:372–376.
International Commission (1978) Ebola haemorrhagic fever in Zaire, 1976. Bull World Health Organ 56:271–293.
Khan AS, Tshioko FK, Heymann DL, et al (1999) The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995. Commission de Lutte contre les Epidémies à Kikwit. J Infect Dis 179 Suppl 1:S76–86.
Kim St J. Damstra, Richardson S, Reeler B (1996) Synchronized Fruiting Between Trees of Ficus thonningii in Seasonally Dry Habitats. J Biogeogr 23:495–500.
Kirika JM, Bleher B, Böhning-Gaese K, et al (2008) Fragmentation and local disturbance of forests reduce frugivore diversity and fruit removal in Ficus thonningii trees. Basic Appl Ecol 9:663–672.
Lahm SA, Kombila M, Swanepoel R, Barnes RFW (2007) Morbidity and mortality of wild animals in relation to outbreaks of Ebola haemorrhagic fever in Gabon, 1994-2003. Trans R Soc Trop Med Hyg 101:64–78.
Lamunu M, Lutwama JJ, Kamugisha J, et al (2004) Containing a haemorrhagic fever epidemic: the Ebola experience in Uganda (October 2000-January 2001). Int J Infect Dis 8:27–37.
Leendertz SAJ, Gogarten JF, Düx A, et al (2016) Assessing the Evidence Supporting Fruit Bats as the Primary Reservoirs for Ebola Viruses. Ecohealth 13:18–25.
Le Guenno B, Formenty P, Formentry P, et al (1995) Isolation and partial characterisation of a new strain of Ebola virus. Lancet 345:1271–1274.
Leroy EM, Epelboin A, Mondonge V, et al (2009) Human Ebola outbreak resulting from direct exposure to fruit bats in Luebo, Democratic Republic of Congo, 2007. Vector Borne Zoonotic Dis 9:723–728.
Leroy EM, Rouquet P, Souquie`re S, et al (2004) Multiple Ebola Virus Transmission Eventsand Rapid Decline of Central African Wildlife.
MacNeil A, Farnon EC, Wamala J, et al (2010) Proportion of deaths and clinical features in Bundibugyo Ebola virus infection, Uganda. Emerg Infect Dis 16:1969–1972.
McFarlane R, Becker N, Field H (2011) Investigation of the climatic and environmental context of Hendra virus spillover events 1994-2010. PLoS One 6:e28374.
Milleliri JM, Tévi-Benissan C, Baize S, et al (2004) [Epidemics of Ebola haemorrhagic fever in Gabon (1994-2002). Epidemiologic aspects and considerations on control measures]. Bull Soc Pathol Exot 97:199–205.
Muyembe T, Kipasa M (1995) Ebola haemorrhagic fever in Kikwit, Zaire. International Scientific and Technical Committee and WHO Collaborating Centre for Haemorrhagic Fevers. Lancet 345:1448.
Nkoghe D, Kone ML, Yada A, Leroy E (2011) A limited outbreak of Ebola haemorrhagic fever in Etoumbi, Republic of Congo, 2005. Trans R Soc Trop Med Hyg 105:466–472.
Ogawa H, Miyamoto H, Nakayama E, et al (2015) Seroepidemiological Prevalence of Multiple Species of Filoviruses in Fruit Bats (Eidolon helvum) Migrating in Africa. J Infect Dis 212 Suppl 2:S101–8.
Okware SI, Omaswa FG, Zaramba S, et al (2002) An outbreak of Ebola in Uganda. Trop Med Int Health 12:1068–1075.
Olival KJ, Hosseini PR, Zambrana-Torrelio C, et al (2017) Host and viral traits predict zoonotic spillover from mammals. Nature 546:646–650.
Olival KJ, Islam A, Yu M, et al (2013) Ebola virus antibodies in fruit bats, bangladesh. Emerg Infect Dis 19:270–273.
Onyango CO, Opoka ML, Ksiazek TG, et al (2007) Laboratory diagnosis of Ebola hemorrhagic fever during an outbreak in Yambio, Sudan, 2004. J Infect Dis 196 Suppl 2:S193–8.
Peterson AT, Bauer JT, Mills JN (2004) Ecologic and geographic distribution of filovirus disease. Emerg Infect Dis 10:40–47.
Philippon N, Jarlan L, Martiny N, et al (2007) Characterization of the Interannual and Intraseasonal Variability of West African Vegetation between 1982 and 2002 by Means of NOAA AVHRR NDVI Data. J Clim 20:1202–1218.
Pigott DM, Golding N, Mylne A, et al (2014) Mapping the zoonotic niche of Ebola virus disease in Africa. Elife 3:e04395.
Pinzon JE, Wilson JM, Tucker CJ, et al (2004) Trigger events: enviroclimatic coupling of Ebola hemorrhagic fever outbreaks. Am J Trop Med Hyg 71:664–674.
Plowright RK, Eby P, Hudson PJ, et al (2015) Ecological dynamics of emerging bat virus spillover. Proc Biol Sci 282:20142124.
Plumptre AJ, Shoo L, Polansky L, et al (2012) Changes in Tree Phenology across Africa: A comparison across 17 sites. WCS Albertine Rift program
Polansky L, Boesch C (2013) Long-term Changes in Fruit Phenology in a West African Lowland Tropical Rain Forest are Not Explained by Rainfall. Biotropica 45:434–440.
Pourrut X, Kumulungui B, Wittmann T, et al (2005) The natural history of Ebola virus in Africa. Microbes Infect 7:1005–1014.
Richter HV, Cumming GS (2006) Food availability and annual migration of the straw-colored fruit bat (Eidolon helvum). J Zool 268:35–44.
Rohatgi A (2017) WebPlotDigitizer V 3.12. In: WebPlotDigitizer V 3.12. http://arohatgi.info/WebPlotDigitizer. Accessed 17 May 2017
Rouquet P, Froment J-M, Bermejo M, et al (2005) Wild animal mortality monitoring and human Ebola outbreaks, Gabon and Republic of Congo, 2001-2003. Emerg Infect Dis 11:283–290.
Rumelhart DE, Hinton GE, Williams RJ (1986) Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1. In: Rumelhart DE, McClelland JL, PDP Research Group C (eds). MIT Press, Cambridge, MA, USA, pp 318–362
Schar D, Daszak P (2014) Ebola economics: the case for an upstream approach to disease emergence. Ecohealth 11:451–452.
Schmidt JP, Park AW, Kramer AM, et al (2017) Spatiotemporal Fluctuations and Triggers of Ebola Virus Spillover. Emerg Infect Dis 23:415–422.
Shoemaker T, MacNeil A, Balinandi S, et al (2012) Reemerging Sudan Ebola virus disease in Uganda, 2011. Emerg Infect Dis 18:1480–1483.
Taylor DJ, Leach RW, Bruenn J (2010) Filoviruses are ancient and integrated into mammalian genomes. BMC Evol Biol 10:193.
Towner JS, Sealy TK, Khristova ML, et al (2008) Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda. PLoS Pathog 4:e1000212.
Tucker CJ, Wilson JM, Mahoney R, et al (2002) Climatic and ecological context of the 1994-1996 Ebola outbreaks. Photogramm Eng Remote Sens 68:147–152.
Wamala JF, Lukwago L, Malimbo M, et al (2010) Ebola hemorrhagic fever associated with novel virus strain, Uganda,

2007-2008. Emerg Infect Dis 16:1087–1092.
Weber DS, Alroy KA, Scheiner SM (2017) Phylogenetic Insight into Zika and Emerging Viruses for a Perspective on Potential Hosts. Ecohealth 14:214–218.
Wittmann TJ, Biek R, Hassanin A, et al (2007) Isolates of Zaire ebolavirus from wild apes reveal genetic lineage and recombinants. Proc Natl Acad Sci U S A 104:17123–17127.
Yamagiwa J, Basabose AK (2006) Diet and seasonal changes in sympatric gorillas and chimpanzees at Kahuzi–Biega National Park. Primates 47:74–90.
Yamagiwa J, Basabose AK, Kaleme KP, Yumoto T (2008) Phenology of fruits consumed by a sympatric population of gorillas and chimpanzees in Kahuzibiega National Park, Democratic Republic of Congo. African Study Monographs 3–22.