Supplemental Data: Gorilla susceptibility to Ebola virus: the cost of sociality

Damien Caillaud, Florence Levréro, Romane Cristescu, Sylvain Gatti, Maeva Dewas, Mélanie Douadi, Annie Gautier-Hion, Michel Raymond and Nelly Ménard

Supplemental Background

Ebola virus has first been identified in wild apes in 1994, when 8 individuals of a 43 individual chimpanzee community were affected in Ivory Coast [S1]. Since then, wild ape mortality has been reported during most of human outbreaks in eastern Gabon and north-western Congo [S2]. An animal mortality monitoring network created in 2001 discovered in this area 98 wild animal carcasses from August 2001 to June 2003 [S2], 65 of which were great apes. Among all the carcasses, 21 were tested and 14 (of which 13 apes and 1 duiker) were found to be EBOV-positive. These results suggest that Ebola may severely affect great ape populations.

Estimating Ebola-induced mortality in wild apes is however a difficult task. Two types of methods have been used. The first consists in comparing results of pre-epidemic and post-epidemic nest counts along transects or 'paths of least resistance'. To our knowledge this methodology has been used twice, in Gabon and in Republic of the Congo [S3,S4]. Surveys conducted in 1998-2000 in Minkébé forest block (Gabon) after Mékouka (1994) and Mayibout 2 (1996) human Ebola outbreaks, revealed a decrease in the ape population of more than 90% since 1990 [S3]. Another study based on this methodology, led in Lossi Gorilla Sanctuary (Republic of the Congo) after 2001–2 outbreak, showed a lower decrease of 56% of a gorilla population [S4], which may be explained by the fact that this area was only partially affected. The second method involves the visual identification of individuals of a population before and after an epidemic. To our knowledge, this approach has been used only once before, in Lossi, where 8 groups of gorillas have been monitored before, during and after the outbreak cited above [S4,S5]. None of these groups has been seen since the end of the outbreak, revealing that group EBOV mortality was probably very high [S4]. However, the mortality rate of solitary males remains unknown. None of these studies has investigated the way EBOV spreads in a population of great apes.

Supplemental Experimental Procedures

Data collection

The Lokoué observation site is a 4 ha swampy clearing close to Lokoué river, in the east of Odzala-Kokoua National Park (Figure S1). No human activity, including hunting, has been reported in this site since the study began. Large mammals, including elephants (*Loxodonta africana cyclotis*), buffalos (*Syncerus cafer nanus*), antelopes (*Tragelaphus eurycerus* and *Tragelaphus spekei*) and gorillas (*Gorilla gorilla gorilla*) frequently visit the clearing to feed on herbaceous vegetation. Lokoué site is around 60 km from the site where the carcass that tested positive for EBOV in June 2003 was found (fig. S1).

From October 4 2001 to June 24 2005 (study period of 1,360 days), the clearing was watched for 699 days, 9 hours a day. Before the outbreak, gorillas were seen on most (94.5%) observation days. Visits lasted on average 2 h 28 min (st.dev.: 1 h 43 min), which in addition to the relative small surface of the clearing allowed a reliable morphology-based identification of social units.

Face shape, body shape and pelage coloration have proved to be useful individual identification criteria [S6,S7]. Powerful spotting scopes, binoculars and a video camera were

used to identify gorillas individually. When instantaneous identification was uncertain, gorillas were photographed using a numeric reflex camera equipped with a 1200 mm/f8 lens. Photos were subsequently compared with a complete numeric picture database. When necessary, sketches were also drawn. Less than 10% of groups or solitary males could not be identified, in general because they made very furtive visits.

All groups included a single mature male (the silverback), which was chosen as the reference for the identity of the group. This allowed a clear definition of the groups in situations where individuals transferred between groups. In 2001–2, 45 groups and 31 solitary males, that is 76 units, were thus identified [S6]. All but 8 of the 45 groups included adult females (these groups were therefore called breeding groups). Solitary males were either mature males or young adult males (blackbacks).

Data transformation

In the statistical analysis we performed (see below), the binary variable 'presence/absence of an individual on a particular day' was the response variable. Data on individuals belonging to a given group were then not independent. Therefore, all group-living individuals except silverbacks were excluded. The initial dataset was then a record of the presence/absence of 53 solitary males and 56 group silverbacks, that is 109 individuals, on each day of the 1,360 day long period of the study.

This corresponds to 109 social units, a higher number than that observed in 2001–2. This is simply due to the continuous turnover of the population. Lokoué population is open, with social units appearing and disappearing.

Statistical modelling based on a $1{,}360 \times 109$ matrix was computationally excessively time-expensive. Thus sighting occasions were pooled for each 10-day period, as described [S8], finally resulting in a 136×109 matrix dataset.

Statistical modelling: general framework

A capture–recapture (CR) modelling approach was adopted. CR models take into account both the variation of the sighting probability of living individuals (due to season, for example) and the variation of the survival rate of individuals, respectively noted p and φ, to explain total variation in the sighting probability of individuals. Here, two factors possibly influencing p and φ were considered: time (t) and class (g). Values of t ranged from 1 (second 10-day period) to 135 (last 10-day period). Parameter g distinguished two classes of individuals: solitary males and group silverbacks.

As there is an extensive literature dedicated to the mathematical design of CR models (e.g., [S9,S10]), we focus below on the particularities of the present model.

The first step was goodness-of-fit (GOF) testing of the Cormack–Jolly–Seber (CJS) model [S9]. This model (notation: φ_{t*g}, p_{t*g}), makes the assumption that within a given class, survival rate (φ) and probability of sighting (p) vary solely according to time. Moreover, all living individuals of a class are assumed to have equal capture probabilities and survival rates at a given sighting occasion t. U-CARE v2.0 software [S11] was used to perform GOF tests on contingency tables. GOF testing of the CJS model is decomposed into four components [S12–S14]. All those components except "2.Ct" component ('trap-happiness', $P < 10^{-4}$) were non-significant in our case. The significance of test '2.Ct' indicated that individuals sighted on a particular occasion tended to be sighted again on the following occasion more frequently than expected. This phenomenon is called trap-dependence [S13]. Basically, individuals visiting the clearing on a particular day were more likely to be sighted again during the following few days, which translated into an immediate trap-dependence effect after the

pooling transformation. We modelled this effect as suggested in ref. [S13], by inserting an additional parameter *dp* in the CJS model, such that :

$$\text{logit}(p_{trap_t}) = \text{logit}(p_t) + dp_t$$

$$\text{or } p_{trap_t} = \frac{p_t \, e^{dp_t}}{1 + p_t \, (e^{dp_t} - 1)}$$

where p_{trap_t} denotes the probability for an individual sighted at time $t-1$ and alive at time *t* to be sighted again at time *t*. p_t is the probability for an individual not sighted at time $t-1$ and alive at time *t* to be sighted at time *t*. dp_t represents trap-dependence at time *t*. In our case (trap-happiness), dp_t is expected to be positive. Computing the additive effect of trap-dependence on a logit scale is convenient since it constrains $p_{trap_t} < 1$ whatever the value of dp_t.

In the first model, *p*, φ and the trap-dependence were class-dependent and time-dependent. This model is denoted φ_{t*g}, p_{t*g*m} [S13], where *m* denotes the trap-dependence. For this model and following ones, a computer software called EpiFit was developed to obtain maximum likelihood estimates of parameters. The maximisation algorithm of EpiFit was adapted from the Metropolis algorithm used in ref. [S15]. It was extensively tested using simulated datasets. This software also provided profile likelihood estimates of confidence intervals of parameters.

AIC-based simplification of *p* parametrization led to retain the more parsimonious model p_{t+g+m}, φ_{t*g} (Table S1). Time is likely to have an effect on sighting probability either through environmental variation (*e.g.*, weather or season) or through variation of the presence of observers on the platform. For instance, if observers were absent two days during a 10-day period, the probability of sighting at that occasion was expected to be lower than if observers were present during all days of the 10-day periods. To control for this factor, p_t was corrected, such that:

$$p_{pres_t} = 1 - (1 - p_t)^{\frac{x_t}{10}},$$

with p_{pres_t} : corrected sighting probability at time *t*, p_t : sighting probability expected if presence were permanent and x_t : number of days of effective presence per pooling interval. The resulting model allowed us to reduce the number of *p* parameters by maintaining p_t constant over three consecutive occasions (Table S1). With 10-day pooling intervals, this means that sighting probability was allowed to vary every 30 days.

Statistical modelling: epidemiological models

In order to investigate the way Ebola virus spread in Lokoué population, φ was constrained by two epidemiological models. These models were designed following recent literature on Ebola epidemiology [S2,S4,S16–S20].

The first model, denoted model SEIR2, assumed that at the very beginning of the epidemic a small number of individuals got contaminated, for instance after coming into contact with the reservoir or with infectious apes from a neighbouring area. These index cases then contaminated other individuals, and so on (see Figure 1 in the main paper). This scenario has not received much attention for two main reasons. First, infectivity would be too short-lived and physical contacts between social units would be to rare to allow an ape-to-ape spread of the virus [S4]. Second, the detection of different strains of Ebola virus in gorilla

carcasses during one outbreak would also argue against this hypothesis [S4]. However, this last argument has recently been challenged [S19], and a quantitative model would have been required to sustain the first one.

The second model, named model Spillover2, assumed that during the epidemic, ape-to-ape transmission was by far less important than reservoir-to-ape transmission (Figure 1). To date this scenario has received the larger support [S2,S4,S17].

Model SEIR2

This model was derived from the classical SEIR model initially designed to describe epidemics in close populations (e.g., [S21]). The SEIR model categorizes individuals of a finite population according to infectious status as susceptible (S), exposed (E), infectious (I) or recovered (R). We adapted this model to the case of an open population divided into two classes of individuals. As the EBOV recovery rate is less than 20% in humans, even with medical cares, it is likely to be very low in gorillas. Thus, the recovered compartment (R) was suppressed and a 'dead, not infectious' compartment (M) was created (Figure 1). Note that the compartment I includes infectious dead individuals. In the following description, index j refers to individuals living in group ($j=1$) or to solitary males ($j=2$).

Let S_j be the number of susceptible individuals of class j per km² during the epidemic, E_j the number of exposed but not infectious individuals, I_j the number of infectious individuals and M_j the number of dead individuals. The following epidemiological parameters are defined:

t_0 : date corresponding to the beginning of the outbreak. The time unit is equal to 10 days (see Data transformation).

φ_{0_j} : natural, non epidemic, survival of class j individuals, per ten days. It is assumed to be constant over time.

r_j : susceptibility of class j individuals to the virus

x : relative infectivity of solitary males compared to group-living individuals ($x \in [0,1]$). Individuals living in group are assumed to be equally or more contagious than solitary males, because when such an individual is infectious, other individuals of its group are likely to be also infectious, hence increasing the contamination probability of a healthy individual meeting it.

$1/k$: mean duration of incubation period

$1/l_2$: mean duration of infectivity of solitary males

$1/z$: relative duration of infectivity of group-living individuals compared to solitary males ($z \in [0,1]$). Individuals living in group are assumed to be infectious during a longer period than solitary males to take into account the possible infectivity of other individuals of their groups.

We derive from the model described in Figure 1:

$$\begin{cases} t \le t_0 \Rightarrow \begin{bmatrix} dS_1 \\ dS_2 \\ dE_1 \\ dE_2 \\ dI_1 \\ dI_2 \\ dM_1 \\ dM_2 \end{bmatrix} = \begin{bmatrix} -(1-\varphi_{0_1})S_1 \\ -(1-\varphi_{0_2})S_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ (1-\varphi_{0_1})S_1 \\ (1-\varphi_{0_2})S_2 \end{bmatrix}.dt \\ t > t_0 \Rightarrow \begin{bmatrix} dS_1 \\ dS_2 \\ dE_1 \\ dE_2 \\ dI_1 \\ dI_2 \\ dM_1 \\ dM_2 \end{bmatrix} = \begin{bmatrix} -r_1(I_1 + x\,I_2)S_1 - (1-\varphi_{0_1})S_1 \\ -r_2(I_1 + x\,I_2)S_2 - (1-\varphi_{0_2})S_2 \\ r_1(I_1 + x\,I_2)S_1 - k\,E_1 - (1-\varphi_{0_1})E_1 \\ r_2(I_1 + x\,I_2)S_2 - k\,E_2 - (1-\varphi_{0_2})E_2 \\ k\,E_1 - z\,l_2\,I_1 - (1-\varphi_{0_1})I_1 \\ k\,E_2 - l_2\,I_2 - (1-\varphi_{0_2})I_2 \\ z\,l_2\,I_1 + (1-\varphi_{0_1})(S_1 + E_1 + I_1) \\ l_2\,I_2 + (1-\varphi_{0_2})(S_2 + E_2 + I_2) \end{bmatrix}.dt \end{cases} \quad (1)$$

Let s_j, e_j, i_j and m_j be the proportions of susceptible, exposed, infectious and dead individuals in the population, respectively. Note that here, the term population refers only to solitary males and group silverbacks (see Data transformation). Let *N* be the density (no. of individuals per km²) of the susceptible individuals of this population at t_0. Equation (1) is equivalent to:

$$\begin{cases} t \le t_0 \Rightarrow \begin{bmatrix} ds_1 \\ ds_2 \\ de_1 \\ de_2 \\ di_1 \\ di_2 \\ dm_1 \\ dm_2 \end{bmatrix} = \begin{bmatrix} -(1-\varphi_{0_1})s_1 \\ -(1-\varphi_{0_2})s_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ (1-\varphi_{0_1})s_1 \\ (1-\varphi_{0_2})s_2 \end{bmatrix} .dt \\ t > t_0 \Rightarrow \begin{bmatrix} ds_1 \\ ds_2 \\ de_1 \\ de_2 \\ di_1 \\ di_2 \\ dm_1 \\ dm_2 \end{bmatrix} = \begin{bmatrix} -r_1 N(i_1 + x i_2)s_1 - (1-\varphi_{0_1})s_1 \\ -r_2 N(i_1 + x i_2)s_2 - (1-\varphi_{0_2})s_2 \\ r_1 N(i_1 + x i_2)s_1 - k e_1 - (1-\varphi_{0_1})e_1 \\ r_2 N(i_1 + x i_2)s_2 - k e_2 - (1-\varphi_{0_2})e_2 \\ k e_1 - z l_2 i_1 - (1-\varphi_{0_1})i_1 \\ k e_2 - l_2 i_2 - (1-\varphi_{0_2})i_2 \\ z l_2 i_1 + (1-\varphi_{0_1})(s_1 + e_1 + i_1) \\ l_2 i_2 + (1-\varphi_{0_2})(s_2 + e_2 + i_2) \end{bmatrix} . dt \end{cases} \quad (2)$$

The incubation period was fixed at 7 days (0.7 time unit), the usual value observed in human during Ebola-Zaire outbreaks [S22–S24]. Given a particular set of epidemiological parameters and the following initial values:

$$i_{1_0} = i_{2_0} = 0.001$$
$$s_{1_0} = 0.593 - i_{1_0} \qquad \text{(from [S6])}$$
$$s_{2_0} = 0.407 - i_{2_0} \qquad \text{(from [S6])}$$
$$e_{1_0} = e_{2_0} = m_{2_0} = m_{2_0} = 0,$$

it was possible to evaluate s_1, s_2, e_1, e_2, i_1, i_2, m_1 and m_2 for any value of *t* using Euler's method. Considering that only susceptible and exposed individuals frequent the clearing, the apparent survival of class *j* individuals between occasions *t* and $t+1$could then be written as:

$$\varphi_{j,t\to t+1} = \frac{s_{j,t+1} + e_{j,t+1}}{s_{j,t} + e_{j,t}}$$

This model was implemented in EpiFit to obtain maximum likelihood estimates of t_0, Nr_j, x, l_2, φ_{0_i}, and z. An estimate of the global survival of class *j* individuals was also computed as:

$$\varphi_{glob_j} = \prod_{t=1}^{q} \varphi_{j,t\to t+1} \ ,$$

with *q*: number of resighting occasions. From this expression we derived the estimate of Ebola-induced mortality for class *j* individuals:

$$m_{tot\ j} = 1 - \frac{\varphi_{glob_j}}{\varphi_{0_j}^{\ q}} \ .$$

Model Spillover2

This model assumed that between social units ape-to-ape transmission of Ebola virus was negligible compared to reservoir-to-ape transmission (Figure 1). The outbreak would then have resulted from a massive spillover from the reservoir [S2,S4,S17]. Two modifications of the equations of the former model sufficed to write model Spillover2:

- Replacement of terms $r_j N(i_1 + x i_2)s_j$, which correspond to the contamination of susceptible individuals following contacts with infectious gorillas, by a new term: $r_j f(t) s_j$. The function $f(t)$ reflects the infectivity of the reservoir along with time.
- Replacement of compartments I and M by a single new compartment. In model SEIR2, the time spent by gorillas in compartment I influences the probability of susceptible gorillas to be infected and hence the survival of gorillas. So, this duration can be estimated from the data, and has to be taken into account. Conversely, model Spillover2 assumes that no virus transmission occurred between social units. The duration of infectivity is then assumed to have no effect on gorilla susceptibility, and hence on gorilla survival. So this parameter cannot be estimated. This is the reason why the model reduces to a three compartment model.

According to [S4], gorilla outbreaks would be favoured during the dry season. A recent paper proposing several species of fruit bats as a reservoir for EBOV hypothesises that fruit scarcity during this period could promote contacts between apes and bats, or that pregnancy of bats during this period could alter their immune functions and increase viral excretion [S20]. So, it seemed realistic to model reservoir infectivity with a bell-shaped f function.

We chose to define f as follows:

$$\begin{cases} t < c \quad \Rightarrow \quad f(t) = \dfrac{4e^{2(t-c)\, l^{-1} \ln(\frac{2+2\sqrt{1-\varepsilon}-\varepsilon}{\varepsilon})}}{\left(1+e^{2(t-c)\, l^{-1} \ln(\frac{2+2\sqrt{1-\varepsilon}-\varepsilon}{\varepsilon})}\right)^2} \\ c \le t < c+a \quad \Rightarrow \quad f(t) = 1 \\ t \ge c+a \quad \Rightarrow \quad f(t) = \dfrac{4e^{2(t-a-c)\, l^{-1} \ln(\frac{2+2\sqrt{1-\varepsilon}-\varepsilon}{\varepsilon})}}{\left(1+e^{2(t-a-c)\, l^{-1} \ln(\frac{2+2\sqrt{1-\varepsilon}-\varepsilon}{\varepsilon})}\right)^2} \end{cases}$$

The meaning of the four parameters of this function can be easily understood (Figure S2). Note that ε was chosen to be equal to 0.05.

Finally, model Spillover2 could be written as follows:

$$\begin{Bmatrix} ds_1 \\ ds_2 \\ de_1 \\ de_2 \\ d(i+m)_1 \\ d(i+m)_2 \end{Bmatrix} = \begin{bmatrix} -r_1\, f(t)\, s_1 - (1-\varphi_{0_1}) s_1 \\ -r_2\, f(t)\, s_2 - (1-\varphi_{0_2}) s_2 \\ r_1\, f(t)\, s_1 - k\, e_1 - (1-\varphi_{0_1})\, e_1 \\ r_2\, f(t)\, s_2 - k\, e_2 - (1-\varphi_{0_2})\, e_2 \\ k\, e_1 + (1-\varphi_{0_1})(e_1 + s_1) \\ k\, e_2 + (1-\varphi_{0_2})(e_2 + s_2) \end{bmatrix} . dt$$

As for model SEIR2, this model was implemented in EpiFit to obtain both its deviance and maximum likelihood estimates of r_j, φ_{0_j}, a, l and c.

Once estimations were performed, we checked the two following important hypotheses of the statistical model:

1. The survival of all the individuals living in group was assumed to be identical to that of group silverbacks. A simple, ad hoc, way to check this was to compare two samples of immatures and adult females of Lokoué population collected with the same effort before and after the epidemic. Thus, during 150 days of sighting before the epidemic, 299 females and immatures were identified. After the epidemic, during an equivalent period of 150 days of sighting, 13 females and immatures were identified. This led, roughly, to the estimate of the survival: $13/299 = 0.043$, a value that is inside the 95% confidence interval of the estimate based on group silverbacks.
2. Data pooling is assumed not to bias survival estimates. However, this is theoretically challenged [S8]. The bias is function of the survival, sighting probability and degree of pooling [S8]. Here, estimates of daily sighting probability ranged from 0 to 0.056. Estimates of daily survival ranged from 0.98 to 1. With such values, 10-day pooling intervals and for a stationary population, the bias, computed as proposed in [S8], was inferior to 1%. This showed that the effect of data transformation on survival estimates was probably negligible.

Model simplification and comparison

Models SEIR2 and Spillover2 were progressively simplified using an AIC based step-by-step procedure (table S1). The most parsimonious models obtained were:

- model SEIR2 with $\varphi_{0_1} = \varphi_{0_2}$
- model Spillover2 with $\varphi_{0_1} = \varphi_{0_2}$ and $a = 0$

The deviances of these models are very close (model SEIR2: deviance = 3384.23; model Spillover2: deviance = 3384.86). However, modelling survival according to model SEIR2 required 7 parameters, whereas only 5 parameters were required by model Spillover2. As $AIC = deviance + 2 \times nb.\ of\ parameters$, the AIC of model Spillover2 is 4 points lower than that of model SEIR2. So, an AIC-based comparison of the models would lead to retain model Spillover2.

However this comparison is irrelevant. Model Spillover2 cannot be rejected in the presence of model SEIR2 as an alternative hypothesis. Let's imagine the virus spread among gorillas through ape-to-ape transmission according to model SEIR2. The proportion of infectious individuals along with time would then be bell-shaped. In this model, this proportion constrains the exposure of susceptible individuals exactly the same way as the reservoir does in model Spillover2. Consequently, the model Spillover2 can also fit the data

perfectly, provided that the shape of its function f is close to that of the proportion of infectious individuals predicted by model Spillover2. So, even if the model SEIR2 is the right one, it is statistically impossible to prove.

Let's have look to the estimates obtained thanks to both models. Given estimates provides in Table S2, model SEIR2 predicts that at the epidemic peak, $i_1 + x i_2$ was equal to 0.21. This means that the maximum probabilities for a solitary male and for a group silverback to become exposed during a 10-day period were, respectively, $r_2 N(i_1 + x i_2) = 0.45 \times 0.21 = 0.095$ and $r_1 N(i_1 + x i_2) = 1.03 \times 0.21 = 0.216$. These values are low enough to be plausible. Exposition of susceptible individuals to EBOV could have occurred for instance during inter-unit encounters. As gorillas home-ranges partially overlap, such events are usual. Encounters have been described for instance in the vicinity of fruit trees [S25] and in forest clearings [S26]. In addition to inter-unit encounters, susceptible individuals could have come into contact with fresh, infectious carcasses. At last, as mentioned in ref. [S2], groups in which the silverback died probably disbanded, facilitating inter-unit transfers of exposed individuals. So our analysis reveals that ape-to-ape transmission of EBOV is a plausible scenario.

The analyses based on model Spillover2 show that Lokoué outbreak could also have originated from a massive spillover from the reservoir. The necessary conditions for that scenario to have occurred include a delimited period of virus excretion from the reservoir, which duration is estimated to 322 days (95% IC: 130 days-539 days, see Table S3). As model Spillover2 made the assumption that there has been a single spillover at Lokoué, this long estimated duration could also have been obtained if two consecutive spillovers from the reservoir occurred. We tested this by fitting a model slightly different from model Spillover2. The function f of this model had two peaks, such that:

$$f(t) = \frac{4e^{2(t-c_1)\, l_2^{-1} \ln\left(\frac{2+2\sqrt{1-\varepsilon}-\varepsilon}{\varepsilon}\right)}}{\left(1+e^{2(t-c_1)\, l_2^{-1} \ln\left(\frac{2+2\sqrt{1-\varepsilon}-\varepsilon}{\varepsilon}\right)}\right)^2} + \frac{4\, h_2\, e^{2(t-c_2)\, l_2^{-1} \ln\left(\frac{2+2\sqrt{1-\varepsilon}-\varepsilon}{\varepsilon}\right)}}{\left(1+e^{2(t-c_2)\, l_2^{-1} \ln\left(\frac{2+2\sqrt{1-\varepsilon}-\varepsilon}{\varepsilon}\right)}\right)^2}$$

with c_i : abscissa of peak i ; l_i : width of peak i and h_2 : height of peak 2.

The deviance of this last model is 3381.7. Its AIC is 3467.7, a value higher than that of model Spillover2. Consequently, the former model assuming a single peak is better than this one.

Interestingly, the estimated duration of the spillover exceeds the duration of the dry seasons, which thus cannot be ecologically related to the occurrence of the spillover from the reservoir.

Anyway, we would like to emphasise that these two mechanisms do not exclude each other. The ape-to-ape transmission hypothesis itself assumes an initial introduction of the virus from the reservoir. Even if we show that at a local geographic scale ape-to-ape transmission is likely to be important, things may be different at a larger geographic scale where natural barriers probably limit the efficiency of this way of spreading.

Supplemental References

S1. Formenty, P., Boesch, C., Wyers, M., Steiner, C., Donati, F., Dind, F., Walker, F., and Le Guenno, B. (1999). Ebola virus outbreak among wild chimpanzees living in a rain forest of Cote d'Ivoire. J. Infect. Dis. *179*, S120–S126.

S2. Rouquet, P., Froment, J.M., Bermejo, M., Kilbourn, A., Karesh, W., Reed, P., Kumulungui, B., Yaba, P., Delicat, A., Rollin, P.E., and Leroy, E.M. (2005). Wild

animal mortality monitoring and human Ebola outbreaks, Gabon and Republic of Congo, 2001-2003. Emerg. Infect. Dis. *11*, 283–290.

S3. Huijbregts, B., De Wachter, P., Obiang, L.S.N., and Akou, M.E. (2003). Ebola and the decline of gorilla Gorilla gorilla and chimpanzee Pan troglodytes populations in Minkebe Forest, north-eastern Gabon. Oryx *37*, 437–443.

S4. Leroy, E.M. (2004). Multiple Ebola virus transmission events and rapid decline of central African wildlife (vol 303, pg 387, 2004). Science *303*, 628–628.

S5. Walsh, P.D., Abernethy, K.A., Bermejo, M., Beyersk, R., De Wachter, P., Akou, M.E., Huljbregis, B., Mambounga, D.I., Toham, A.K., Kilbourn, A.M., *et al.* (2003). Catastrophic ape decline in western equatorial Africa. Nature *422*, 611–614.

S6. Gatti, S., Levrero, F., Menard, N., and Gautier-Hion, A. (2004). Population and group structure of western lowland gorillas (Gorilla gorilla gorilla) at Lokoue, Republic of Congo. Am. J. Primatol. *63*, 111–123.

S7. Parnell, R.J. (2002). Group size and structure in western lowland gorillas (Gorilla gorilla gorilla) at Mbeli Bai, Republic of Congo. Am. J. Primatol. *56*, 193–206.

S8. Hargrove, J.W., and Borland, C.H. (1994). Pooled population parameter estimates from mark-recapture data. Biometrics *50*, 1129–1141.

S9. Lebreton, J.D., Burnham, K.P., Clobert, J., and Anderson, D.R. (1992). Modeling survival and testing biological hypotheses using marked animals - a unified approach with case-studies. Ecol. Monog. *62*, 67–118.

S10. Pollock, K.H., Nichols, J.D., Brownie, C., and Hines, J.E. (1990). Statistical inference for capture-recapture experiments. Wildlife Monog. *107*, 1:97.

S11. Choquet, R., Reboulet, A.M., Pradel, R., Gimenez, O., and Lebreton, J.D. (2002). User's manual for U-care, CEFE, CNRS.

S12. Burnham, K.P., Anderson, D.R., White, G.C., Brownie, C., and Pollock, K.H. (1987). Design and analysis methods for fish survival experiments based on release-recapture. American Fisheries Society Monographs.

S13. Pradel, R. (1993). Flexibility in survival analysis from recapture data : handling trap-dependence. In Marked Individuals in the Study of Bird Population, J.D. Lebreton and P.M. North, eds. (Basel: Birkhaüser verlag), pp. 29–37.

S14. Pradel, R., Hines, J.E., Lebreton, J.D., and Nichols, J.D. (1997). Capture-recapture survival models taking account of transients. Biometrics *53*, 60–72.

S15. Szymura, J.M., and Barton, N.H. (1986). Genetic analysis of a hybrid zone between the fire-bellied toads, Bombina-Bombina and Bombina-Variegata, near Cracow in Southern Poland. Evolution *40*, 1141–1159.

S16. Leroy, E.M., Telfer, P., Kumulungui, B., Yaba, P., Rouquet, P., Roques, P., Gonzalez, J.P., Ksiazek, T.G., Rollin, P.E., and Nerrienet, E. (2004). A serological survey of Ebola virus infection in central African nonhuman primates. J. Infect. Dis. *190*, 1895–1899.

S17. Pourrut, X., Kumulungui, B., Wittmann, T., Moussavou, G., Delicat, A., Yaba, P., Nkoghe, D., Gonzalez, J.P., and Leroy, E.M. (2005). The natural history of Ebola virus in Africa. Microb. Infect. *7*, 1005–1014.

S18. Allela, L., Bourry, O., Pouillot, R., Delicat, A., Yaba, P., Kumulungui, B., Rouquet, P., Gonzalez, J.P., and Leroy, E.M. (2005). Ebola virus antibody prevalence in dogs and human risk. Emerg. Infect. Dis. *11*, 385–390.

S19. Walsh, P.D., Biek, R., and Real, L.A. (2005). Wave-like spread of Ebola Zaire. PLoS Biol.*3*, 1946-1953.

S20. Leroy, E.M., Kumulungui, B., Pourrut, X., Rouquet, P., Hassanin, A., Yaba, P., Delicat, A., Paweska, J.T., Gonzalez, J.P., and Swanepoel, R. (2005). Fruit bats as reservoirs of Ebola virus. Nature *438*, 575–576.

S21. Wearing, H.J., Rohani, P., and Keeling, M.J. (2005). Appropriate models for the management of infectious diseases. PLoS Med. *2*, e174.

S22. Breman, J.G., Piot, P., Johnson, K.M., White, M.K., Mbuyi, M., Sureau, P., Heymann, D.L., Van Nieuwenhove, S., McCormick, J.B., Ruppol, J.P., *et al.* (1978). The epidemiology of Ebola hemorrhagic fever in Zaire, 1976. In Ebola Virus Haemorrhagic Fever, S.R. Pattyn, ed. (New York: Elsevier/North-Holland Biomedical Press), pp. 103–121.

S23. Bwaka, M.A., Bonnet, M.J., Calain, P., Colebunders, R., De Roo, A., Guimard, Y., Katwiki, K.R., Kibadi, K., Kipasa, M.A., Kuvula, K.J., *et al.* (1999). Ebola hemorrhagic fever in Kikwit, Democratic Republic of the Congo: Clinical observations in 103 patients. J. Infect. Dis. *179*, S1–S7.

S24. Ndambi, R., Akamituna, P., Bonnet, M.J., Tukadila, A.M., Muyembe-Tamfum, J.J., and Colebunders, R. (1999). Epidemiologic and clinical aspects of the Ebola virus epidemic in Mosango, Democratic Republic of the Congo, 1995. J. Infect. Dis. *179 Suppl 1*, S8–10.

S25. Bermejo, M. (2004). Home-range use and intergroup encounters in western gorillas (Gorilla g. gorilla) at Lossi forest, North Congo. Am. J.l Primatol. *64*, 223–232.

S26. Levrero, F., Gatti, S., Menard, N., and Gautier-Hion, A. (2002). Inter-unit encounters in western lowland gorillas (Gorilla g. gorilla). Folia Primatol. *73*, 157.

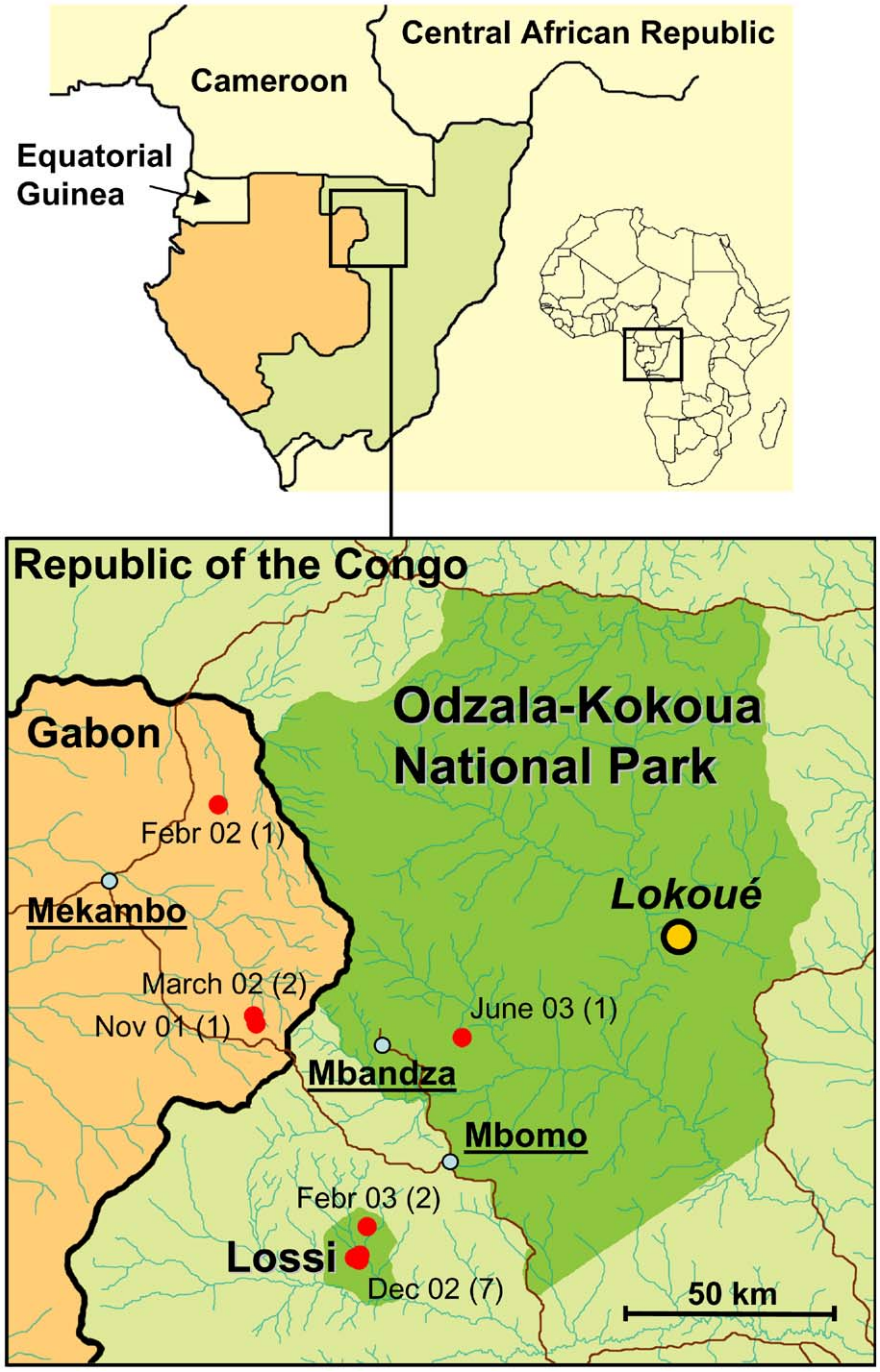

Figure S1. Ebola-positive carcasses around Lokoué site.
Red dots show the location and date of sampling of carcasses found before the Lokoué outbreak (from [S2]) and the number of carcasses appears in parentheses.

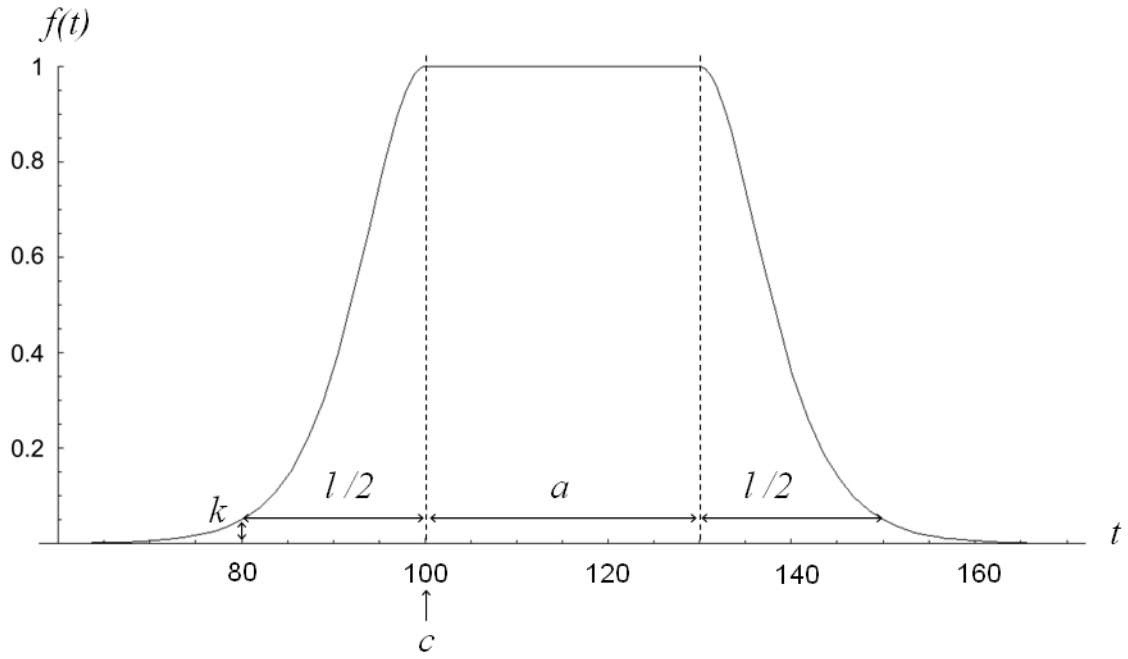

Figure S2. Graphic representation of the function describing reservoir infectivity in model Spillover2. The value of parameter *k* was chosen to be equal to 0.05.

Table S1. Progressive AIC-based simplification of statistical models. In blue and orange: minimal models.

***p* parametrization**	***φ* parametrization**	**nb. of parameters**	**deviance**	**AIC**
t*g*m	t*g	578	2980.81	4136.81
(t*g)+m	t*g	387	3137.47	3911.47
t*g	t*g	386	3175.17	3947.17
(t*g)+m, 30 days constancy	t*g	261	3302.51	3824.51
t+g+m, 30 days constancy	t*g	229	3332.59	3790.59
t+m, 30 days constancy	t*g	228	3334.96	3790.96
t+g+m, 30 days constancy	SEIR2	43	3384.02	3470.02
t+g+m, 30 days constancy	SEIR2, $\varphi_{0_1} = \varphi_{0_2}$	42	3384.23	3468.23
t+g+m, 30 days constancy	SEIR2, $\varphi_{0_1} = \varphi_{0_2}$, $r_1 = r_2$	41	3391.82	3473.82
t+g+m, 30 days constancy	SPILLOVER2	42	3384.22	3468.22
t+g+m, 30 days constancy	SPILLOVER2, $\varphi_{0_1} = \varphi_{0_2}$	41	3384.49	3466.49
t+g+m, 30 days constancy	SPILLOVER2, $\varphi_{0_1} = \varphi_{0_2}$, $a = 0$	40	3384.93	3464.93
t+g+m, 30 days constancy	SPILLOVER2, $\varphi_{0_1} = \varphi_{0_2}$, $a = 0$, $r_1 = r_2$	39	3392.34	3470.34

Notes :

1. Variables on each side of "*" affect independently *p* or φ. The variable to the right of "+" has an additive effect on *p* (on a logit scale). For example, in model $p_{(t*g)+m}$, the sighting probabilities are computed independently for each sighting occasion and statistical group; for individuals sighted twice consecutively, the sighting probability corresponding to the second sighting occasion is increased by a constant dp, on a logit scale, irrespective of the values of *t* or *g*.

2. As there are 135 sighting-sighting intervals, model $p_{(t*g)}, \varphi_{(t*g)}$, for example, is expected to have $(135+134)\times 2 = 538$ parameters. This is higher than the value given in the table, because for occasions when observers were absent, p and φ were not allowed to vary.

Table S2. Maximum likelihood estimates and profile likelihood confidence intervals of parameters obtained with EpiFit software from the final model SEIR2.

parameter	maximum likelihood estimate	profile likelihood estimate of 95% confidence interval
t_0	73.50	61.4 - 82.9
Nr_1	1.03	0.64 - 2.98
Nr_2	0.45	0.15 - 2.35
r_1 / r_2	2.28	1.27 - 4.35
$\varphi_{0_1}, \varphi_{0_2}$	0.996	0.994 - 0.998
m_{tot_1}	0.97	0.95 - 0.98
m_{tot_2}	0.77	0.69 - 0.87
l_2	0.75	0.20 - 10.00
z	0.25	0.00 - 1.00
x	1.00	0.00 - 1.00

Notes:

1. t_0 was arbitrarily set as time at which 1 ‰ of gorilla units of each class were infectious. So, the estimated value 73.5, corresponding to 0ctober 9, 2003, does not signify that the epidemic started exactly at this date. What is important to determine when the epidemic began is the variation of survival rate along with time, which does not depend on this arbitrary value (fig. 2B).
2. Confidence intervals of l_2, z and x correspond to definition domain of these parameters. Basically, this is due to the too slight effect of the variation of these parameters on the survival curves. Data on individuals of categories I (infectious) and M (dead) would have been required to obtain better estimations.

Table S3. Maximum likelihood estimates and profile likelihood confidence intervals of epidemiological parameters obtained with EpiFit software from the final model Spillover2.

parameter	maximum likelihood estimate	profile likelihood estimate of 95% confidence interval
r_2	0.22	0.12 - 0.55
r_1 / r_2	0.45	0.23 - 0.81
$\varphi_{0_1}, \varphi_{0_2}$	0.996	0.994 - 0.998
m_{tot_1}	0.97	0.92 - 0.98
m_{tot_2}	0.77	0.62 - 0.87
c	93.4	90.3 - 97.9
l	32.2	13.0 - 53.9