RESEARCH ARTICLE
The Effect of Life History on Retroviral Genome Invasions

Ravinder K. Kanda*, Tim Coulson
Department of Zoology, University of Oxford, South Parks Road, Oxford, United Kingdom
* ravinder.kanda@zoo.ox.ac.uk

Abstract
Endogenous retroviruses (ERV), or the remnants of past retroviral infections that are no longer active, are found in the genomes of most vertebrates, typically constituting approximately 10% of the genome. In some vertebrates, particularly in shorter-lived species like rodents, it is not unusual to find active endogenous retroviruses. In longer-lived species, including humans where substantial effort has been invested in searching for active ERVs, it is unusual to find them; to date none have been found in humans. Presumably the chance of detecting an active ERV infection is a function of the length of an ERV epidemic. Intuitively, given that ERVs or signatures of past ERV infections are passed from parents to offspring, we might expect to detect more active ERVs in species with longer generation times, as it should take more years for an infection to run its course in longer than in shorter-lived species. This means the observation of more active ERV infections in shorter compared to longer-lived species is paradoxical. We explore this paradox using a modeling approach to investigate factors that influence ERV epidemic length. Our simple epidemiological model may explain why we find evidence of active ERV infections in shorter rather than longer-lived species.

Introduction
A significant proportion of host genomes are littered with the remnants of past retroviral infections. Termed endogenous retroviruses (ERVs), in humans ERVs represent 8% of the genome and over 10% of the genome in mice [1,2]. Infection by a retrovirus requires integration into the cellular DNA as part of its replication cycle. Integration into the germline cells and subsequent vertical transmission provides us with a genomic fossil record of multiple, independent, ancient retroviral infections, described from a wide range of vertebrate genomes, including mammals, fish, birds, reptiles and amphibians [3,4]. Typically, an ERV consists of 3 genes (gag, pol and env) and two flanking non-coding long terminal repeats (LTRs), which are identical at the time of integration. Over time, these retroviral insertions accumulate mutations and deletions at the same rate as the mutation rate of the host genome [5], rendering them non-functional. ERVs may also be inactivated by recombinational deletion between the two flanking LTRs, which removes the internal coding region, leaving a solo LTR. Solo LTRs are 10–100
times more numerous than their full length counterparts [6], and many of these insertions are fixed in the host population. In various mammal species there are a few examples of intact, evolutionarily young ERVs that are polymorphic (present in some but not all individuals) in their host population [7–12]; whether this is persisting polymorphism maintained through various evolutionary forces or actual active ERVs remains to be established. However, some ERVs do appear to be intact, capable of expression and replication [11,13–17]—what we consider to be “active”.

To date, no active ERVs have been discovered in humans. Most ERV research in humans is computationally based, comprising of data mining of sequenced human genomes, which has revealed numerous polymorphic ERV insertions of the youngest known family of human ERVs, HERV-K(HML2) [7,18]. The most common insertions (those insertions present at high frequency in the population) are present in the reference genome; it is unclear whether the newer insertions more recently identified are evidence of activity of this particular retroviral family, or lingering polymorphism. However, until a particular insertion in the human genome reaches a high frequency, it is unlikely to be detected unless one is specifically looking for them [7]. By the same token, any new retroviral invasions into the human genome would also be unlikely to be detected until they reached an appreciable frequency in the population. In contrast, the other mammalian genome that has been extensively studied is that of the mouse. Their use as a model organism has led to a great deal of experimental research on mice, informing our understanding of the various roles that ERVs may play and revealing a number of active ERVs [2,12,16] (“active” in the sense that there are intact copies in the genomes which are capable of expression and replication—in some cases have been shown to produce infectious viral particles). This has led to the conclusion that mice have more frequent ERV invasions than humans [19], but is this true? To ascertain the activity of ERVs in different species would require the same amount of effort that has been invested in research on mice ERVs, which in most cases is unfeasible. A search through the literature shows that despite over 3500 endogenous retroviruses having been sequenced from 138 species of mammals, the majority of ERV research has focused on just a handful of species (Table 1—see Methods for further details). An alternative approach is to model ERV dynamics, and explore factors that would influence our chance of detecting active ERVs, such as epidemic length (i.e. how long does it take for an endogenous retroviral insertion to reach a high enough frequency that it would be detected in the few whole genomes from a species that are sequenced?).

Intuitively, there are two factors that affect the likelihood of finding active ERVs: the rate of incorporation/loss—if it varies with life history, and how long they take to run their course (i.e. the time to fixation at a specific locus), which we call the length of the epidemic. Little is known about how the rate of incorporation/loss may vary across species with different life histories. Here we explore how we would expect the length of an epidemic to vary across different life histories: where should we be looking to identify active ERVs? Given that organisms with a faster life history are i) short lived (shorter generation times), ii) tend to have larger numbers of offspring, and iii) have a larger effective population size, we might at first expect that a beneficial insertion (those ERVs that are involved in endogenous viral element derived immunity (EDI), and would therefore be considered beneficial) would spread to fixation faster in these species, than those with a slower life history [20]. Hence, we would expect an ERV epidemic to take longer (in years) in species with slower life histories and be more easily detectable at a given time. We explore how a search across different life histories may affect what we find.

Susceptible-Infectious-Recovered (SIR) models are a useful tool for inferring the length of an epidemic [21]. They have been used extensively to describe the dynamics of various infectious disease epidemics, including foot and mouth disease [22–24] and measles [25–27]. We have previously described a SIR model to investigate the circumstances under which a disease-
Table 1. Focus of ERV research across species (105) with different life histories.

| Latin Name                      | Common Name         | Total no. of ERVs identified | GestationLength | Effort | Active ERVs described |
|--------------------------------|---------------------|------------------------------|-----------------|--------|-----------------------|
| Ornithorhynchus anatinus        | Platypus            | 1                            | 12.48           | 0      | 0                     |
| Monodelphis domestica          | Gray short-tailed opossum | 1                      | 14.55           | 1      | 0                     |
| Trichosurus vulpecula           | Common brushtail    | 3                            | 17.5            | 0      | 0                     |
| Mus cervicolor                  | Fawn-coloured mouse| 1                            | 17.6            | 3      | 0                     |
| Mus caroli                      | Ryukyu mouse       | 1                            | 18.1            | 4      | 0                     |
| Mus musculus                    | Mouse               | 234                          | 19.6            | 708    | 9                     |
| Myodes glareolus                | Bank vole           | 4                            | 17.94           | 0      | 0                     |
| Chionomys nivalis               | European snow vole  | 3                            | 20.92           | 0      | 0                     |
| Phodopus sungorus               | Siberian hamster   | 5                            | 21              | 1      | 0                     |
| Rattus norvegicus               | Norway rat          | 10                           | 21.74           | 2      | 1                     |
| Oryzomys palustris              | Marsh rice rat      | 16                           | 24.74           | 1      | 0                     |
| Peromyscus maniculatus          | North American deer mouse | 23                      | 26.68           | 1      | 0                     |
| Sigmodon hispidus               | Hispid cotton rat   | 11                           | 27              | 1      | 0                     |
| Macropus eugenii                | Tammar wallaby      | 4                            | 29.89           | 3      | 0                     |
| Macropus rufogriseus            | Red-necked wallaby | 6                            | 30              | 1      | 0                     |
| Suncus murinus                  | House shrew         | 4                            | 30.19           | 2      | 0                     |
| Oryctolagus cuniculus           | Rabbit              | 9                            | 30.45           | 43     | 0                     |
| Phascolarctos cinereus          | Koala               | 1                            | 34.29           | 12     | 1                     |
| Microcebus murinus              | Gray mouse lemur    | 4                            | 60.34           | 3      | 0                     |
| Cheiroleurus medius             | Fat-tailed dwarf lemur | 3                      | 61.79           | 0      | 0                     |
| Felis chaus                     | Jungle cat          | 3                            | 62.88           | 1      | 0                     |
| Felis catus                     | Domestic cat        | 38                           | 62.99           | 53     | 1                     |
| Canis lupus familiaris          | Dog                 | 5                            | 63.5            | 16     | 0                     |
| Felis silvestris                | Wildcat             | 4                            | 65.49           | 2      | 0                     |
| Cavia porcellus                 | Guinea pig          | 5                            | 66.99           | 10     | 0                     |
| Prionailurus bengalensis euptilurus | Leopard cat      | 1                            | 70.18           | 0      | 0                     |
| Puma concolor                   | Puma                | 2                            | 92.3            | 0      | 0                     |
| Dasyprocta leporina             | Brazilian agouti    | 1                            | 106.39          | 0      | 0                     |
| Chinchilla lanigera             | Long-tailed chinchilla | 1                        | 112.47          | 0      | 0                     |
| Sus barbatus barbatus           | Bearded pig         | 27                           | 114.63          | 0      | 0                     |
| Sus barbatus oi                 | Bearded pig         | 26                           | 114.63          | 0      | 0                     |
| Sus scrofa                      | Pig                 | 441                          | 115.2           | 281    | 1                     |
| Potamochoerus porcus            | Red river hog       | 32                           | 120             | 1      | 0                     |
| Eulemur fulvus                  | Brown lemur         | 1                            | 120.83          | 0      | 0                     |
| Sus verrucosus                  | Java warty pig      | 27                           | 121.66          | 0      | 0                     |
| Potamochoerus larvatus          | Bush pig            | 29                           | 121.73          | 1      | 0                     |
| Sus celebensis                  | Celebes wild boar   | 24                           | 126             | 0      | 0                     |
| Aotus trivirgatus               | Northern owl monkey | 11                           | 133.47          | 1      | 0                     |
| Lemur catta                     | Ring-tailed lemur   | 2                            | 134.74          | 1      | 0                     |
| Alluropoda melanoleuca          | Giant panda         | 1                            | 134.99          | 1      | 0                     |
| Saguinus midas                  | Red-handed tamarin  | 8                            | 138.24          | 0      | 0                     |
| Callithrix jacchus              | Marmoset            | 7                            | 144             | 6      | 0                     |
| Cercopithecus ascanius          | Red-tailed monkey   | 1                            | 148.5           | 0      | 0                     |
| Hydrochoerus hydrochaeris       | Capybara            | 1                            | 150.73          | 0      | 0                     |
| Hylodocherus meinertzhageni     | Giant forest hog    | 9                            | 151             | 0      | 0                     |
| Ovis aries                      | Sheep               | 18                           | 152.54          | 91     | 1                     |

(Continued)
Table 1. (Continued)

| Latin Name                  | Common Name          | Total no. of ERVs identified | GestationLength | Effort | Active ERVs described |
|-----------------------------|----------------------|-------------------------------|-----------------|--------|-----------------------|
| Babyrousa babyrussa         | Babirusa             | 11                            | 156.5           | 0      | 0                     |
| Saimiri boliviensis boliviensis | Bolivian squirrel monkey | 2                            | 157.79          | 0      | 0                     |
| Macaca cyclops              | Formosan rock macaque | 2                            | 161.06          | 0      | 0                     |
| Cebus capucinus             | White-headed capucin | 2                            | 161.06          | 0      | 0                     |
| Pithecia pithecia           | White-faced saki     | 2                            | 161.13          | 0      | 0                     |
| Macaca radiata              | Bonnet macaque       | 2                            | 161.56          | 2      | 0                     |
| Saimiri sciureus            | Squirrel monkey      | 9                            | 164.09          | 23     | 0                     |
| Miopithecus talapoin        | Angolan talapoin     | 3                            | 164.38          | 1      | 0                     |
| Macaca fascicularis         | Crab eating macaque  | 42                           | 164.69          | 17     | 0                     |
| Macaca sylvanus             | Barbary macaque      | 2                            | 164.84          | 0      | 0                     |
| Cercopithecus atys          | Sooty mangabey       | 1                            | 165.08          | 1      | 0                     |
| Phacochoerus aethiopicus    | Desert warthog       | 18                           | 165.4           | 0      | 0                     |
| Macaca mulatta              | Rhesus macaque/monkey| 27                           | 166.07          | 61     | 0                     |
| Saguinus oedipus            | Common-top tamarin   | 3                            | 166.49          | 1      | 0                     |
| Macaca maura                | Moor macaque         | 1                            | 167.19          | 0      | 0                     |
| Erythrocebus patas          | Patas monkey         | 1                            | 167.2           | 0      | 0                     |
| Macaca thibetana            | Tibetan macaque      | 2                            | 169.02          | 0      | 0                     |
| Cercopithecus pogonias      | Crested mona monkey  | 1                            | 169.51          | 0      | 0                     |
| Cercopithecus nictitans     | Greater spot-nosed monkey | 1                        | 169.51          | 0      | 0                     |
| Macaca nemestrina           | Pig-tailed macaque   | 7                            | 171             | 6      | 0                     |
| Macaca silenus              | Lion-tailed macaque  | 2                            | 172             | 1      | 0                     |
| Cercopithecus neglectus     | De Brazza's monkey   | 2                            | 172.07          | 0      | 0                     |
| Macaca nigra                | Celebes crested macaque | 1                         | 172.43          | 0      | 0                     |
| Phacochoerus africanus      | Warthog              | 33                           | 172.49          | 1      | 0                     |
| Macaca fuscata              | Japanese macaque     | 65                           | 172.99          | 3      | 0                     |
| Papio cynocephalus          | Yellow baboon        | 2                            | 172.99          | 5      | 0                     |
| Mandrillus sphinx           | Mandrill             | 2                            | 173.99          | 1      | 0                     |
| Cercopithecus galeritus     | Tana River mangabey  | 1                            | 174.43          | 0      | 0                     |
| Macaca arctoides            | Stump-tailed macaque | 4                            | 176.6           | 3      | 0                     |
| Theropithecus gelada        | Gelada baboon        | 2                            | 178.64          | 0      | 0                     |
| Papio anubis                | Olive baboon         | 8                            | 178.96          | 2      | 0                     |
| Papio hamadryas             | Hamadryas baboon     | 15                           | 180             | 2      | 0                     |
| Macaca sinica               | Toque macaque        | 2                            | 180.9           | 0      | 0                     |
| Papio papio                 | Guinea baboon        | 2                            | 184.42          | 1      | 0                     |
| Papio ursinus               | Chacma baboon        | 3                            | 185.92          | 0      | 0                     |
| Alouatta seniculus          | Venezuela red howler | 7                            | 189.9           | 0      | 0                     |
| Semnopithecus entellus      | Hanuman langur       | 2                            | 197.7           | 3      | 0                     |
| Hylobates pileatus          | Pileated gibbon      | 2                            | 200.16          | 0      | 0                     |
| Odocoileus hemionus         | Mule deer            | 1                            | 203.49          | 1      | 1                     |
| Hylobates lar               | Gibbon               | 4                            | 212.91          | 55     | 0                     |
| Symphalangus syndactylus    | Siamang              | 2                            | 230.66          | 1      | 0                     |
| Procavia capensis           | Cape rock hyrax      | 20                           | 231.24          | 0      | 0                     |
| Pan troglodytes             | Chimpanzee           | 93                           | 231.49          | 63     | 0                     |
| Pan paniscus                | Bonobo chimp         | 25                           | 235.24          | 2      | 0                     |
| Tragelaphus spekii          | Sitatunga            | 1                            | 241.15          | 0      | 0                     |
| Hylobates moloch            | Silvery gibbon       | 2                            | 241.2           | 0      | 0                     |
| Gorilla gorilla             | Gorilla              | 56                           | 257             | 32     | 0                     |

(Continued)
causing retrovirus can become incorporated into the host genome and spread through the host population, if it confers an immunological advantage [28]. This use of compartmental models is now being used by others to investigate retroviral dynamics [29]. We use the model of Kanda et al [28] to explore which factors would influence the length of an epidemic of such an ERV (i.e. one involved in EDI), and how this varies with the life history of the host.

Methods

A keyword search on the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov/nuccore) for “endogenous retrovirus” AND “mammals” shows the number of endogenous retroviruses that have been described in mammals (3542 in 138 species—see S1 Table for accession numbers). Using PanTHERIA [30], we retrieved information on gestation length for 105 of these species. Species where information was not available in PanTHERIA, were excluded from this analysis. Gestation length has been shown to be suitable indicator of speed of life history [31]. To assess the amount of “effort” focused on the ERVs of these species, we conducted a PubMed search using the “latin name” OR “common name” AND “endogenous retrovirus” as search terms (see S2 Table for PubMed search results). The number in the column headed “effort” represents the number of papers returned by PubMed. From the literature, we identified those species described as having active ERVs. This data is collated in Table 1 where species are ordered from faster life history to slow. From this table we examined the association between life history, measured as gestation length, and the number of active endogenous retroviruses reported having corrected for effort (measured as the number of papers on ERVs in the host species), using a generalized linear model (GLM) with Poisson distribution. This analysis, although relatively crude, will highlight evidence for any association between host life history and the number of active ERVs.

The Model

Kanda et al [28] developed a set of epidemiological models to investigate the conditions under which incorporation of retroviruses into the host genome benefits the host. Here, we use one of

| Latin Name         | Common Name     | Total no. of ERVs identified | Gestation Length | Effort | Active ERVs described |
|--------------------|-----------------|-----------------------------|------------------|--------|-----------------------|
| Pongo pygmaeus     | Orangutan       | 46                          | 259.42           | 25     | 0                     |
| Homo sapiens       | Human           | 1580                        | 274.78           | 2082   | 0                     |
| Bos taurus         | Cattle          | 11                          | 280.48           | 44     | 0                     |
| Bos javanicus      | Banteng         | 1                           | 296.78           | 0      | 0                     |
| Bubalus bubalis    | Water buffalo   | 1                           | 320              | 0      | 0                     |
| Trichechus manatus | West Indian manatee | 20                      | 334.58           | 1      | 0                     |
| Equus caballus     | Horse           | 7                           | 338.97           | 7      | 0                     |
| Equus asinus asinus| Donkey          | 4                           | 367              | 0      | 0                     |
| Globicephala macrocephalus | Short-financed pilot whale | 1              | 452              | 0      | 0                     |
| Orcinus orca       | Killer whale    | 1                           | 456.25           | 2      | 0                     |
| Elephas maximus    | Asiatic elephant| 41                          | 634.49           | 0      | 0                     |
| Loxodonta africana | African elephant| 42                          | 660              | 0      | 0                     |

Total number of ERVs identified represents the number of entries in the NCBI nucleotide database for each species. Species where information was not available in PanTHERIA were excluded for the purposes of this table. Gestation length is given in days. Effort represents the number of entries found in PubMed. Search criteria are described in the methods. Searches are correct as of 30/09/13.

doi:10.1371/journal.pone.0117442.t001
these models to examine the effects of life history on epidemic length. This is a type of compartmental model, standardly used to study the dynamics of a range of diseases [32,33].

We focus on the final model presented by Kanda et al [28] that realistically represented the dynamics of an EDI type ERV across a range of life histories. This model differs from a standard SIR model, in that it includes 2 infected compartments, $I_X$ and $I_E$ (Fig. 1). The $I_X$ compartment refers to individuals who are infected with the exogenous retrovirus; upon successful incorporation of the retrovirus into the germline of an individual, the offspring of these individuals enter the $I_E$ compartment—these individuals have an endogenous copy of the retrovirus. The model also has 2 recovered compartments, $R_N$ and $R_{LTR}$. Individuals in the $R_N$ compartment have successfully dealt with the exogenous retroviral infection without incorporation of the retrovirus. The $R_{LTR}$ compartment consists of individuals who have recovered from the retroviral infection with an endogenous copy of the retrovirus in the genome. The model is described mathematically by the equations below and represented graphically in Fig. 2. The equations describe the flow of individuals between compartments:

$$S(t+1) = (\phi_s + \tau)S(t) - \frac{\beta_X S(t)I_X(t)}{N(t)} - \frac{\beta_E S(t)I_E(t)}{N(t)} + \tau R_N(t)$$

$$I_X(t+1) = (\phi_E \phi_s + \tau)I_X(t) + \frac{\beta_X S(t)I_X(t)}{N(t)} +$$

$$+ \frac{\beta_E S(t)I_E(t)}{N(t)} - \theta I_X(t) - \gamma \tau I_X(t) + \alpha I_E(t)$$

$$I_E(t+1) = (\phi_E \phi_s + \tau)I_E(t) + \gamma \tau I_X(t) - \alpha I_E(t) - \alpha \theta I_E(t)$$

$$R_N(t+1) = \phi_s R_N(t) + \theta I_X(t)$$

$$R_{LTR}(t+1) = (\phi_{LTR} + \tau)R_{LTR}(t) + \theta \alpha I_E(t)$$

(1)

where $t$ represents time, $\tau$ is the birth rate (which does not differ between compartments) and $\phi_s$ is the survival rate of susceptible individuals. $\gamma$ is the rate of incorporation of the retrovirus, $\alpha$ is the rate of loss of the endogenous retrovirus (mutation or recombinational deletion), $\beta_X$ and $\beta_E$ are the infection rates of the exogenous and endogenous virus respectively, $\theta$ and $\theta'$ are the rates at which immunity is acquired to the exogenous and endogenous virus respectively, $\phi_X \phi_s$ and $\phi_E \phi_s$ are the survival rates of the individuals infected by the exogenous and endogenous virus respectively. $N(t) = S(t) + I_X(t) + I_E(t) + R_N(t) + R_{LTR}(t)$, is the total population size. Baseline values and a description of the parameters are summarised in Table 2.

Parameter values for the model

The survival ($\phi_s$) and fertility ($\tau$) rate parameters define the life history of a species in this model [34]. Large $\phi_s$ and small $\tau$ correspond to a species with a slow life history, while large $\tau$ and small $\phi_s$ correspond to a species with a faster life history. We constrain $\phi_s + \tau = 1.016$, as smaller values increase the probability of extinction, but above this value the extinction risk is virtually zero. We alter the life history of the species by changing the values of $\phi_s$ and $\tau$ such that their sum always equals 1.016. As $\tau$ gets larger the life history speeds up, and as $\phi_s$ increases the life history slows down. We vary $\phi_s$ and $\tau$ in increments of 0.01. For each life history we then independently vary values of all other parameters as described below.
We model the invasion of an immunologically beneficial ERV, by imposing a conservative mortality increase of $1 - \phi_X = 0.03$ on survival (new survival rate = $\phi_X\phi_S$) attributable to the exogenous virus. For the endogenous virus, we impose a mortality increase on survival of $1 - \phi_E = 0.015$ (new survival rate = $\phi_E\phi_S$). $\phi_X$ and $\phi_E$ are reductions in $\phi_S$ (survival). We assume that the mortality rate of the $I_E$ group is less than that of the $I_X$ compartment. Similarly, we assume the infection rates of the exogenous virus ($\beta_X$) and the endogenous virus ($\beta_E$) are also likely to differ, with the endogenous virus being less infectious than the exogenous. Baseline values are set at 0.5 and 0.4 respectively. We explore how varying these values ($\beta_X$ and $\beta_E$) in increments of 0.05 affect the epidemic length across different life histories.

With the exception of a few HERVs, there is little information available regarding the rate of incorporation of the retrovirus ($\gamma$) and the rate of loss ($\alpha$) of the virus (mutation or recombinational deletion). The estimates that are available for humans suggest that $\alpha < \gamma$ (see [28] for further details). Our baseline values are set at $\alpha = 0.00001$ and $\gamma = 0.0001$. We explore how varying these two parameters affects the length of the epidemic (from 0.0001 to 0.001).

Fig 1. Overview of the different compartments in the SIR model. Susceptible individuals (S), have not encountered the retroviral infection. Those individuals in the $I_X$ compartment have encountered exogenous viral particles and are infected, but the virus has not integrated into their genome. Integration of the retrovirus into the germline of these $I_X$ individuals results in the offspring of these individuals entering the $I_E$ compartment—these individuals are infected and infectious, and have an endogenous copy of the retrovirus in their genome. Individuals in the $R_X$ compartment have successfully dealt with the infection, without incorporating the virus into their genome—they are the same as the S individuals. Individuals in the $R_{LTR}$ compartment have also successfully dealt with the infection, but they are left with a copy of the integration in their genome (illustrated here as the LTR (green), but may also be a full length provirus (red and green as in the virus)).

doi:10.1371/journal.pone.0117442.g001
The rates at which immunity would arise to the exogenous virus (θ) and the endogenous virus (θ') are unknown and so we set the baseline values for these parameters at θ = 0.05 and θ' = 0.01; we have previously shown the absolute parameter values of θ, θ' and α are relatively small.

Table 2. Parameters and definitions in the SIR model.

| Parameter | Baseline Values | Definition |
|-----------|-----------------|------------|
| ϕ_s       | 1.016-τ         | Survival rate per time step of an individual uninfected with a virus |
| ϕ_X       | 0.97            | The proportional reduction in ϕ_s caused by the exogenous retrovirus |
| ϕ_E       | 0.985           | The proportional reduction in ϕ_s caused by the endogenous retrovirus |
| τ_1        | 1.016-ϕ_s      | The birth rate per time step |
| β_X        | 0.5             | The infection rate of an individual with the exogenous virus |
| β_E        | 0.4             | The infection rate of an individual with the endogenous virus |
| γ          | 0.0001          | Rate of incorporation |
| α          | 0.00001         | Rate of loss of the incorporated virus |
| θ          | 0.05            | The rate at which immunity is acquired to the exogenous retrovirus |
| θ'         | 0.01            | The rate at which immunity is acquired to the endogenous retrovirus |

Summary of the parameters in the model and their baseline values.

doi:10.1371/journal.pone.0117442.t002
unimportant, as it is their relative values that determine the dynamics [28]. As the values of $\theta$, $\theta'$ and $\alpha$ approach zero, the longer the epidemic lasts and the longer simulations need to be run until the asymptotic equilibrium is reached. We vary $\theta$ and $\theta'$ from 0.0001 to 0.1 in 10 increments to determine the influence of immunity on the length of an epidemic.

Conducting the simulation

We ran simulations of the model until the proportion of the population in each class was stable to a tolerance of 0.000001, and recorded the length of time until equilibrium. We start with one $I_X$ individual and the rest of the population in the $S$ compartment. The simulation was run for a maximum of 50000 iterations to allow the population to reach a stable equilibrium, and we calculated the number of generations this had taken. Generation length ($T_C$) was calculated using the equation below [35]:

$$T_C = \frac{\sum a \phi_s a \tau}{\sum \phi_s a \tau}$$  \hspace{1cm} (2)

where $a$ represents age, $l_a$ (or $\phi_s a$) is the survivorship from birth to age $a$, and $m_a$ (or $\tau$) is the fertility rate at age $a$. Baseline values for the various parameters are described above. All simulations were conducted in R version 2.15.2 [36].

Sensitivity of epidemic length to transition rates

Because we are interested in the effect of the range of parameter values on the length of the epidemic across life histories, we wish to examine how varying the model parameters impacts the time taken for the population to reach equilibrium. We do this by systematically altering the values of each model parameter in 10 increments (between the ranges described for each parameter above), and re-running the simulation for different life histories (differing values of $\phi_s$ and $\tau$).

Results

Analysis of data in Table 1 revealed a negative association between host gestation length and the number of active ERVs identified in the host species, corrected for effort (slope = -0.0264, s.d. = 0.0005, $p < 0.001$). Fig. 3 illustrates the effect of life history, as predicted by the GLM (line) compared to the actual data (crosses), on the number of active ERVs we see. The number of active ERVs significantly increases with the speed of host life history (as approximated by gestation length). In species where data is available, active ERVs are significantly more likely to be found in species with faster life history. This contrasts with our original intuitive expectation that ERV epidemics should run their course faster in fast-lived species, and that we should see less active ERVs in fast-lived species than slow-lived species. The relationship between the survival rate ($\phi_s$) and fertility rate ($\tau$) is illustrated in Fig. 4. Given that mean survival and mean fertility are constrained, generation length increases as mean survival increases. Generation length, by its definition, is a function of $\phi_s$ and $\tau$, and is unaffected by the other parameters (equation 2).

The effects on the epidemic length, of varying the other transition parameters are illustrated in Fig. 5. The first striking observation overall, is that the length of the epidemic is more strongly correlated with generation length than anything else—something that is also suggested by data available from the literature (Fig. 3). The length of the epidemic is greater in species with a fast life history (short lived) than slow life histories, regardless of the parameters that describe infection rates, incorporation rate or rate of loss. For an epidemic to last, there must to be more
individuals in the susceptible compartment, and in fast-lived species we have faster generation of susceptible individuals. Loss and integration can only occur during reproduction. As $\tau$ (birth rate) increases, the rate of flow of individuals into the $I_X$ compartment increases much more than the rate of flow of individuals to the $R_{UTR}$ compartment i.e. $\alpha > \theta'\alpha$ and it increases at a rate of $\frac{1}{h}$ – hence the epidemic takes longer to run its course in species with a fast life history.

Secondly, we observe that when it is easy to acquire immunity to the exogenous retrovirus (higher values of $\theta$), individuals spend more time moving around the left hand side of Fig. 2 (between the $S$, $I_X$ and $R_N$ compartments), and the epidemic lasts longer (Fig. 5 I and J). If immunity to the exogenous virus is easy to acquire, individuals never progress to the $I_E$ compartment and the epidemic never progresses. This effect is less pronounced in species with longer life histories. At $\theta > 0.07$, the population fails to reach a stable equilibrium in 50000 years [28].

Fig 3. Predicted & observed number of active ERVs as a function of life history. Life history is approximated by gestation length (see Table 1). The black line represents the GLM predictions, whereas the crosses (blue and red) represent the actual data (from Table 1). Crosses in red indicate a high amount of “effort” (>10).

doi:10.1371/journal.pone.0117442.g003
The effect of most of the other parameters is negligible, with the exception of the rate of loss, $\alpha$ (Fig. 5 A and B), for species with a faster life history. As $\alpha$ increases, more individuals transition from the $I_E$ compartment to the $I_X$ compartment. This flow of individuals towards the left hand side of Fig. 2 results in the epidemic taking longer.

**Discussion**

Our approach of using SIR models to study the multigenerational dynamics of an endogenous retroviral infection is a novel approach, combining methods from demography [35] with epidemiological methods to address the question of ERV epidemic length. Because our model is density independent, the population size will increase exponentially, but population growth rate and structure will converge on a stable equilibrium.
Why are we interested in finding active ERVs?

There is a growing body of evidence to suggest that for some ERVs, incorporation into the genome provides the host with some immunity from related exogenous retroviruses (described as endogenous viral element derived immunity, EDI [37]), through a variety of different mechanisms [10,38–41]. Under this scenario, there is clearly an advantage to incorporation of an exogenous virus into the genome. However, despite the advantages of incorporation of retroviruses into the genome, there are also downsides to consider.

In many species there are numerous examples of ERVs causing disease [10,42,43]. In humans alone, HERVs have been associated with cancer [44–46], multiple sclerosis [47–49], and a whole host of other diseases (see [50] for review). Additionally, we also have to consider the possibility that the reverse of endogenisation may also occur—exogenous viruses emerging from active ERV lineages. The reservoir of viruses present in host genomes may also be able to recombine with exogenous retroviruses, resulting in novel recombinant viruses that maybe pathogenic—a phenomenon that has already been observed in cats [51]. It has been shown that the standard mechanism of ERV replication within a genome involves reinfection of germ-line cells, and hence possibly movement between host individuals [52,53]. Subsequently, cross-species transmission of these viruses is a very real concern—there are numerous examples identified to date, where cross-species transmission of retroviruses can lead to emergent disease, e.g. HIV was certainly acquired from the non-human primate version of the virus, SIV, which has crossed the species barrier on multiple occasions from chimpanzees and sooty mangabeys [54]—SIV has also been found to be endogenous in a species of lemur [39]. Other recent cross-species transmission event include the introduction of the koala retrovirus (KoRV),
which is suspected to have originated from the exogenous Gibbon ape Leukemia Virus (GaLV) [55]. There are also several cases of close evolutionary relationships between exogenous retroviruses and ERVs in the same host species, e.g., in the sheep, cat, chicken and mouse [13,56]. Interpreting ERV diversity remains challenging and a better understanding of where we are likely to find active ERVs, and consequently possible threats of emerging disease, is clearly important in informing the direction of research in this area.

Where and when would we expect to find active ERVs?

Our model suggests that the reason why more active ERVs are discovered in species with a fast life history (such as mice and koalas), than in those with a slow life history, is not necessarily that these species have more frequent ERV epidemics, but that those epidemics last longer and are therefore more likely to be detected. For the subset of ERVs involved in EDI, the life history of the host has the greatest bearing on the length of an epidemic. The next most influential factor is the rate at which immunity arises to the exogenous retroviral infection (θ). There is a greater rate of generation of susceptible individuals in faster life histories than slower, resulting in a longer time taken for the majority of the population to reach the R_{LTR} compartment. Not all ERVs will provide an immunity advantage. Our model applies to that subset of ERVs that are involved in EDI. The majority of active ERVs that have been described, are described in species that do have a faster life history, such as mice, koala and sheep [10,16,42], which would be in line with our predictions that these epidemics last longer in species with a fast life history, and are therefore more easily detectable. The unusually high number of active ERVs identified in mice may indeed be an anomaly; the numerous studies of ERVs in mice have focused on laboratory strains and it is possible that inbreeding and selection of certain characteristics of this model organism may have unintentionally contributed to the high levels of ERV activity observed in this particular species. For example, the first inbred mouse strain, DBA, was bred for its coat colour which has been shown to be the result of an ERV insertion [57]. However, until an equal amount of “effort” is invested into other species with similar life histories (or wild mouse populations), it is difficult to ascertain whether mice are simply more susceptible to ERV infections. Nonetheless, more active ERVs are described in species with a fast life history. Table 1 illustrates the number of active ERVs that have been identified in all mammals in which ERVs have been described. There is a strong correlation between the number of active ERVs and the life history of the host (as estimated from gestation length), when weighted for effort, supporting our finding that we are more likely to find active ERVs in shorter lived species than in longer lived species (Fig. 3). Previously, these observations have been attributed to a higher level of activity of ERVs in these species (particularly in mice) [58–60]. In our model, this would correspond to higher values of infection rates (β_X and β_E), and incorporation (γ), which interestingly, do not appear to have a considerable effect on the length of an epidemic. In light of these results, it is also worth noting that the proportion of the genome derived from ERVs in mice (short life history) and humans (long life history) is fairly similar—8% and 10% respectively [1,2]. If ERV activity were greater in species with short life histories, then we should expect more of their genomes to originate from ERV insertions. The implications of these results are that in species with a slow life history (such as humans), we should not expect to easily find active ERVs, as the epidemic occurs quickly.

Existing data on species where active ERVs have been discovered are consistent with the results of our model. However, we acknowledge that this could also be a bias in the available data, as more research in this respect has been conducted on species with faster life histories, as demonstrated in Table 1. The model we have described does not account for specific mechanisms of EDI; a better understanding regarding the mechanisms behind how retroviral
immunity (EDI) is gained, which may vary with life history, would be valuable in refining this model and allowing us to better target the search for active ERVs. Further studies of a range of species, with more active ERVs and contrasting life histories, will enable a better estimation of the parameters. However, our model suggests efforts to identify active ERVs should be focused on species with faster life histories, as this is where we stand a better chance of discovering active ERVs and potential threats of new emerging infections.

Supporting Information
S1 Table. Accession numbers for the 3542 endogenous retroviruses identified in the NCBI nucleotide database.

(SDOCX)

S2 Table. PubMed search results to assess the effort focused on the ERVs of a particular species.

(XXXS)

Acknowledgments
We thank Louise Johnson and Susanne Schindler for useful comments on the manuscript.

Author Contributions
Conceived and designed the experiments: RKK TC. Performed the experiments: RKK TC. Analyzed the data: RKK TC. Contributed reagents/materials/analysis tools: RKK TC. Wrote the paper: RKK.

References
1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860–921. PMID:11237011
2. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520–562. PMID:12466850
3. Herniou E, Martin J, Miller K, Cook J, Wilkinson M, et al. (1998) Retroviral diversity and distribution in vertebrates. J Virol 72: 5955–5966. PMID:9621058
4. Feschotte C, Gilbert C (2012) Endogenous viruses: insights into viral evolution and impact on host biology. Nature Reviews Genetics 13: 283–U288. doi:10.1038/nrg3199 PMID: 22421730
5. Blikstad V, Benachenhou F, Sperber GO, Blomberg J (2008) Evolution of human endogenous retroviral sequences: a conceptual account. Cellular and Molecular Life Sciences 65: 3348–3365. doi:10.1007/s00018-008-8495-2 PMID: 18818874
6. Stoye JP (2001) Endogenous retroviruses: still active after all these years? Curr Biol 11: R914–916. PMID: 11719237
7. Belshaw R, Dawson AL, Woolven-Allen J, Redding J, Burt A, et al. (2005) Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K (HML2): implications for present-day activity. J Virol 79: 12507–12514. PMID:16160178
8. Ellenberger L, Kim O, Padhi A, Bankert JG, Simeonov I, et al. (2012) Polymorphic integrations of an endogenous gammaretrovirus in the mule deer genome. J Virol 86: 2787–2796. doi:10.1128/JVI.06859-11 PMID: 22190723
9. Hughes JF, Coffin JM (2004) Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution. Proc Natl Acad Sci U S A 101: 1668–1672. PMID: 14757818
10. Palmarini M, Mura M, Spencer TE (2004) Endogenous betaretroviruses of sheep: teaching new lessons in retroviral interference and adaptation. J Gen Virol 85: 1–13. PMID: 14719813
11. Tarlinton RE, Meers J, Young PR (2006) Retroviral invasion of the koala genome. Nature 442: 79–81. PMID: 16623453
The Effect of Life History on Retroviral Genome Invasions

12. Zhang Y, Maksakova IA, Gagnier L, van de Lagemaat LN, Mager DL (2008) Genome-wide assessments reveal extremely high levels of polymorphism of two active families of mouse endogenous retroviral elements. PLOS Genet 4: e1000007. doi: 10.1371/journal.pgen.1000007 PMID: 18454193

13. Arnaud F, Caporale M, Varela M, Biek R, Chessa B, et al. (2007) A paradigm for virus-host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses. PLOS Pathog 3: e170. PMID: 17997604

14. Arnaud F, Varela M, Spencer TE, Palmarini M (2008) Coevolution of endogenous betaretroviruses of sheep and their host. Cellular and Molecular Life Sciences 65: 3422–3432. doi: 10.1007/s00018-008-8500-9 PMID: 18818869

15. Ruprecht K, Ferreira H, Flockerzi A, Wahl S, Sauter M, et al. (2008) Human endogenous retrovirus family HERV-K(HML-2) RNA transcripts are selectively packaged into retroviral particles produced by the human germ cell tumor line Tera-1 and originate mainly from a provirus on chromosome 22q11.21. J Virol 82: 10008–10016. doi:10.1128/JVI.01016-08 PMID: 18684837

16. Stocking C, Kozak CA (2008) Murine endogenous retroviruses. Cellular and Molecular Life Sciences 65: 3383–3398. doi: 10.1007/s00018-008-8497-0 PMID: 18818872

17. Weiss RA (2006) The discovery of endogenous retroviruses. Retrovirology 3: 67. PMID:17018135

18. Lee A, Huntley D, Aiewsakun P, Kanda RK, Lynn C, et al. (2014) Novel Denisovan and Neanderthal Retroviruses. J Virol 88.

19. Gifford R, Tristem M (2003) The evolution, distribution and diversity of endogenous retroviruses. Virus Genes 26: 291–315. PMID:12876457

20. Hartl DL, Clark AG (1997) Principles of Population Genetics: Sinauer Associates.

21. Anderson RM, May RM, Anderson B (1991) Infectious Diseases of Humans: Dynamics and Control: Oxford University Press. PMID: 25144101

22. Eble P, de Koeijer A, Bouma A, Stegeman A, Dekker A (2006) Quantification of within- and between-pen transmission of Foot-and-Mouth disease virus in pigs. Vet Res 37: 647–654. PMID: 16777036

23. Lawson AB, Onicescu G, Ellerbe C (2011) Foot and mouth disease revisited: re-analysis using Bayesian spatial susceptible-infectious-removed models. Spat Spatiotemporal Epidemiol 2: 185–194. doi: 10.1016/j.sste.2011.07.004 PMID: 22748177

24. Arnaud F, Varela M, Spencer TE, Palmarini M (2008) Coevolution of endogenous betaretroviruses of sheep and their host. Cellular and Molecular Life Sciences 65: 3422–3432. doi:10.1007/s00018-008-8500-9 PMID: 18818869

25. Stocking C, Kozak CA (2008) Murine endogenous retroviruses. Cellular and Molecular Life Sciences 65: 3383–3398. doi:10.1007/s00018-008-8497-0 PMID: 18818872

26. Weiss RA (2006) The discovery of endogenous retroviruses. Retrovirology 3: 67. PMID:17018135

27. Gifford R, Tristem M (2003) The evolution, distribution and diversity of endogenous retroviruses. Virus Genes 26: 291–315. PMID:12876457

28. Hartl DL, Clark AG (1997) Principles of Population Genetics: Sinauer Associates.

21. Anderson RM, May RM, Anderson B (1991) Infectious Diseases of Humans: Dynamics and Control: Oxford University Press. PMID: 25144101

22. Eble P, de Koeijer A, Bouma A, Stegeman A, Dekker A (2006) Quantification of within- and between-pen transmission of Foot-and-Mouth disease virus in pigs. Vet Res 37: 647–654. PMID: 16777036

23. Lawson AB, Onicescu G, Ellerbe C (2011) Foot and mouth disease revisited: re-analysis using Bayesian spatial susceptible-infectious-removed models. Spat Spatiotemporal Epidemiol 2: 185–194. doi: 10.1016/j.sste.2011.07.004 PMID: 22748177

24. Orsel K, Bouma A, Dekker A, Stegeman JA, de Jong MC (2009) Foot and mouth disease virus transmission during the incubation period of the disease in piglets, lambs, calves, and dairy cows. Prev Vet Med 88: 158–163. doi:10.1016/j.prevetmed.2008.09.001 PMID: 18929417

25. Krylova O, Earn DJ (2013) Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. J R Soc Interface 10: 20130098. doi:10.1098/rsif.2013.0098 PMID: 23676892

26. Burton J, Billings L, Cummings DA, Schwartz IB (2012) Disease persistence in epidemiological models: the interplay between vaccination and migration. Mathematical Biosciences 239: 91–96. doi: 10.1016/j.mbs.2012.05.003 PMID: 22652034

27. Uziel A, Stone L (2012) Determinants of periodicity in seasonally driven epidemics. J Theor Biol 305: 88–95. doi:10.1016/j.jtbi.2012.02.031 PMID: 22465112

28. Kanda RK, Tristem M, Coulson T (2013) Exploring the effects of immunity and life history on the dynamics of an endogenous retrovirus. Philos Trans R Soc Lond B Biol Sci 368: 20120505. doi:10.1098/rstb.2012.0505 PMID: 23938754

29. Katzourakis A, Magiorkinis G, Lim AG, Gupta S, Belshaw R, et al. (2014) Larger mammalian body size leads to lower retroviral activity. PLOS Pathog 10: e1004214. doi:10.1371/journal.ppat.1004214 PMID: 25033295

30. Jones KE, Biebl J, Cardillo M, Fritz SA, O’Dell J, et al. (2009) PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology 90: 2648–2648.

31. Biebl J, Mace GM, Bininda-Emonds OR, Cardillo M, Gittleman JL, et al. (2007) The fast-slow continuum in mammalian life history: an empirical reevaluation. Am Nat 169: 748–757. PMID: 17479461

32. Anderson RM, May RM (1979) Population biology of infectious diseases: Part I. Nature 280: 361–367. PMID: 460412

33. Vynnycky E, White RG (2010) An Introduction to Infectious Disease Modelling. Oxford: Oxford University Press. doi:10.14219/jada.archive.2010.0269 PMID: 25555789

34. Steams SC (1977) The Evolution of Life History Traits: A Critique of the Theory and a Review of the Data. Annual Review of Ecology and Systematics: 145–171.

35. Caswell H (2001) Matrix population models: construction, analysis and interpretation: Sinauer Associates Inc., U.S. PMID: 25506954
36. Team RDC (2011) R: A language and environment for statistical computing. 2.15.1 ed. R Foundation for Statistical Computing, Vienna, Austria. doi: 10.1080/17437199.2011.587961 PMID: 25473706

37. Aswad A, Katzourakis A (2012) Paleovirology and virally derived immunity. Trends Ecol Evol 27: 627–636. doi: 10.1016/j.tree.2012.07.007 PMID: 22901901

38. Best S, Le Tissier PR, Stoye JP (1997) Endogenous retroviruses and the evolution of resistance to retroviral infection. Trends Microbiol 5: 313–318. PMID: 9263409

39. Muller V, De Boer RJ (2006) The integration hypothesis: an evolutionary pathway to benign SIV infection. PLOS Pathog 2: e15. PMID: 16609728

40. Taylor GM, Gao Y, Sanders DA (2001) Fv-4: identification of the defect in Env and the mechanism of resistance to ecotropic murine leukemia virus. J Virol 75: 11244–11248. PMID:11602766

41. Tikhonenko AT, Lomovskaya OL (1990) Avian endogenous provirus (ev-3) env gene sequencing: implication for pathogenic retrovirus origination. Virus Genes 3: 251–258. PMID:2161159

42. Tarlinton R, Meers J, Young P (1997) Endogenous retroviruses and the evolution of resistance to retroviral infection. Trends Microbiol 5: 313–318. PMID:9263409

43. Muller V, De Boer RJ (2006) The integration hypothesis: an evolutionary pathway to benign SIV infection. PLOS Pathog 2: e15. PMID: 16609728

44. Taylor GM, Gao Y, Sanders DA (2001) Fv-4: identification of the defect in Env and the mechanism of resistance to ecotropic murine leukemia virus. J Virol 75: 11244–11248. PMID:11602766

45. Tikhonenko AT, Lomovskaya OL (1990) Avian endogenous provirus (ev-3) env gene sequencing: implication for pathogenic retrovirus origination. Virus Genes 3: 251–258. PMID:2161159

46. Tikhonenko AT, Lomovskaya OL (1990) Avian endogenous provirus (ev-3) env gene sequencing: implication for pathogenic retrovirus origination. Virus Genes 3: 251–258. PMID:2161159

47. Nexo BA, Christensen T, Frederiksen J, Moller-Larsen A, Oturai AB, et al. (2011) The etiology of multiple sclerosis: genetic evidence for the involvement of the human endogenous retrovirus HERV-Fc1. PLOS One 6: e16652. doi:10.1371/journal.pone.0016652 PMID: 21311761

48. Gold JGR, Maruszak H, Giovannoni G, Yeates D, Goldacre M (2014) HIV and lower risk of multiple sclerosis: beginning to unravel a mystery using a record-linked database study. J Neurol Neurosurg Psychiatry.

49. Hanger JJ, Bromham LD, McKee JJ, O’Brien TM, Robinson WF (2000) The nucleotide sequence of koala (Phascolarctos cinereus) retrovirus: a novel type C endogenous virus related to Gibbon ape leukemia virus. J Virol 74: 4264–4272. PMID: 10756041

50. Boeke JD, Stoye JP (1997) Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements. In: Coffin JM, Hughes SH, Varmus HE, editors. Retroviruses. Cold Spring Harbor (NY).

51. Nellaker C, Keane TM, Yalcin B, Wong K, Agam A, et al. (2012) The genomic landscape shaped by selection on transposable elements across 18 mouse strains. Genome Biol 13: R45. doi: 10.1186/gb-2012-13-6-r45 PMID: 22703977
59. Rebollo R, Zhang Y, Mager DL (2012) Transposable elements: not as quiet as a mouse. Genome Biol 13: 159. doi: 10.1186/gb-2012-13-6-159 PMID: 22715936

60. Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, et al. (2006) Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLOS Genet 2: e2. PMID: 16440055