Pattern of Tick Aggregation on Mice: Larger Than Expected Distribution Tail Enhances the Spread of Tick-Borne Pathogens

Luca Ferreri1,2+, Mario Giacobini1,2,3, Paolo Bajardi1,2, Luigi Bertolotti1, Luca Bolzoni4,5, Valentina Tagliapietra5, Anna Paola Rizzoli5, Roberto Rosà5

1 Computational Epidemiology Group, Department of Veterinary Sciences, University of Torino, Torino, Italy, 2 Applied Research on Computational Complex Systems Group, Department of Computer Science, University of Torino, Torino, Italy, 3 Complex Systems Unit, Molecular Biotechnology Centre, University of Torino, Torino, Italy, 4 Istituto Zooprofilattico Sperimentale della Lombardia e dell’Emilia Romagna, Parma, Italy, 5 Dipartimento Biodiversità ed Ecologia molecolare, Centro Ricerca e Innovazione, Fondazione Edmund Mach, San Michele all’Adige, Italy

Abstract

The spread of tick-borne pathogens represents an important threat to human and animal health in many parts of Eurasia. Here, we analysed a 9-year time series of Ixodes ricinus ticks feeding on Apodemus flavicollis mice (main reservoir-competent host for tick-borne encephalitis, TBE) sampled in Trentino (Northern Italy). The tail of the distribution of the number of ticks per host was fitted by three theoretical distributions: Negative Binomial (NB), Poisson-LogNormal (PoiLN), and Power-Law (PL). The fit with theoretical distributions indicated that the tail of the tick infestation pattern on mice is better described by the PL distribution. Moreover, we found that the tail of the distribution significantly changes with seasonal variations in host abundance. In order to investigate the effect of different tails of tick distribution on the invasion of a non-systemically transmitted pathogen, we simulated the transmission of a TBE-like virus between susceptible and infective ticks using a stochastic model. Model simulations indicated different outcomes of disease spreading when considering different distribution laws of ticks among hosts. Specifically, we found that the epidemic threshold and the prevalence equilibria obtained in epidemiological simulations with PL distribution are a good approximation of those observed in simulations feed by the empirical distribution. Moreover, we also found that the epidemic threshold for disease invasion was lower when considering the seasonal variation of tick aggregation.

Introduction

Several ecological studies have shown that the distribution of ticks on their hosts is often highly aggregated, with a large number of hosts harbouring few parasites and a small number harbouring a large number of them [11–5; other interesting references could be found in [6]]. In addition, the distribution of tick development stages is coincident, rather than independent [7]. Specifically, those hosts feeding larval tick stages were simultaneously feeding the greatest number of nymphs. As a result, about 20% of all hosts feed 80% of both larvae and nymphs and the number of larvae feeding alongside nymphs is twice as many as it would be if the distributions were independent [7,8]. The aggregation of parasites on hosts bears important implications for vector-borne disease dynamics, since the small fraction of hosts supporting the bulk of the vector population is also responsible for the majority of the pathogen transmission [9].

The transmission of tick-borne diseases is characterised by an intricate set of ecological and epidemiological relationships between pathogen, tick vector, vertebrate hosts and humans that largely determine their temporal and spatial dynamics [10]. Tick-borne disease dynamics feature several complexities, due to the presence of a number of heterogeneities in the system coupled with non-linear phenomena operating in the transmission processes between ticks, host and pathogen [11]. The transmission of pathogens from one tick to another, a prerequisite for the establishment of cycles of infection, may occur via three different pathways depending on the pathogen (see [12] for a comprehensive review). First, adult female ticks may transmit the pathogen to eggs trans-ovarially. Second, ticks may infect a host during their blood meal, leading to a systemic infection in the host; ticks might then acquire the infection by feeding on an infected host, maintaining the infection trans-stadially. Third, ticks may become infected by co-feeding with infected ticks on the same host. Co-feeding transmission is also called non-systemic as it does not require the host to have a systemic infection, since pathogens are transmitted from one tick to another as they feed in close proximity. Vertebrate hosts may vary in their competency to
The degree of aggregation of ticks can be measured in a number of ways. Since the appearance of influential works by Randolph [23] and Shaw et al. ([6] and [26]) the negative binomial (NB) distribution has been extensively used to describe tick aggregation on hosts (see e.g. [23,27,20]). Alternatively, other works suggested that different distributions characterised by larger tails than NB (i.e., predicting more rodents with very large tick burden than expected with NB), can be effective in describing tick aggregations. Specifically, a Poisson-LogNormal (PoiLN) mixed model has been successfully used to describe tick distribution on red grouse chicks [29], while Bisanzio and collaborators [30] showed the first evidence that the distribution heterogeneity of ticks on hosts seemed to be better described by a power-law (PL) than a negative binomial distribution. A suitable description of the distribution tail might have important consequences on the dynamics of the pathogen spreading process. Modelling the spread of vector-borne diseases through bipartite networks [30] showed that the extreme aggregation of ticks on hosts has dramatic consequences on the behaviour of the epidemic threshold. In the current study we used an extensive data set of *Ixodes ricinus* ticks feeding on mice (a total of 4722 parasitised hosts collected in 9 years) to detect the best fit for the distribution of tick burden on mice by testing the performance of NB and PoiLN versus PL distribution, with particular interest in the shape of the distribution tail which is crucial to suitably describe the fraction of co-feeding ticks necessary for TBEv transmission. Then, we used a stochastic model to simulate the effect of fitting different tick distributions on the infection dynamics of a tick-borne pathogen. Specifically, we investigated the spread of a non-systemically transmitted pathogen (e.g. TBEv) by modelling the pathogen transmission between susceptible and infective ticks, considering only co-feeding transmission and distributing ticks on mice under the hypotheses of NB, PoiLN, and PL distributions. Finally, we investigated the seasonal variations in the pattern of tick burden distribution on mice and its implication on TBE-like infection dynamics.

**Materials and Methods**

**Ethics Statement**

All animal handling procedures and ethical issues were approved by the Provincial Wildlife Management Committee (renewed authorisation n. 595 issued on 04.05.2011)

**Tick Burden Data**

Rodent tick burden data was collected by trapping mice using capture-marking-recapture techniques during 2000–2008. The study area was a mixed broadleaf woodland [7,21], located in Valle dei Laghi within the Autonomous Province of Trento, in the north-eastern Italian Alps (grid reference 1652050E 5093750N, altitude 750–800 m a.s.l.). In the year 2000, mice were monitored in nine selected areas through placement of 8×8 trapping grids with a 15-m inter-trap interval. In 2001 and 2002 the number of trapping grids was reduced to eight, while from 2003 onward their number was further reduced to four. In summary, the trapping effort consisted of 129 twice-daily trap sessions with at least one capture, resulting in a total number of 4722 *A. flavescens* ticks captured with at least one tick attached. For each captured rodent the number and life stage of feeding ticks was carefully assessed and registered, without removal [7,21]. A total number of 55411 ticks were counted of which 98.64% were larvae, 1.30% were nymphs, and 0.04% were adults. The number of ticks [nymphs] per rodent was between 1 [0] and 111 [15] with a median number of ticks per rodent equals to 8.
Detailed data, on a yearly scale, are reported in Table 1, while the fraction of nymphs observed in different year and grids is reported in Table 2. In Figure 1 the number of captured Apodemus flavicollis per trapping session is shown for the whole nine year period and for different grids (A to I).

**Data Analysis**

**Tick burden distribution.** Ticks patterns have usually been described as highly aggregated. Therefore, since the seminal works by Crofton [2], Plowright et al. [3], and those by Anderson and May [4] and [5], the negative binomial (NB) probability distribution,

\[
q(k) = \binom{k+r-1}{k} (1-p)^r p^k, \quad k = 0,1,2,\ldots \quad (1)
\]

has been considered suitable for describing macroparasite distribution on hosts. Here we used a maximum-likelihood-estimation (MLE) method to estimate the parameters \(p\) and \(r\) of the probability distribution of the tick burden on the entire dataset obtained by aggregating capture sessions and grids. In addition, we considered subsets of the original dataset composed by mice with large numbers of feeding ticks to evaluate the capability of the NB distribution to fit the tail of the parasite distribution. In particular, we estimated the parameters of the NB distribution on data characterised by \(k \geq k_{\min}\), where \(k_{\min}\) represents the threshold value of ticks per host above which the distribution is fitted. To evaluate the performance of the obtained fits we used Kolmogorov-Smirnov (KS) statistics. The goodness of each fit (GOF) was also evaluated through a bootstrap resampling procedure, generating \(10^3\) synthetic data sets. The obtained \(p\)-value is defined as the relative number of times that the KS statistic of the fitted distributions on synthetic data exceeds that measured on real data. Therefore, the larger the \(p\)-value, the lower our confidence in rejecting the fit. We considered the conservative value 0.1 as our threshold value, as suggested by Clauset and collaborators [31].

As a first alternative to the NB distribution, we considered a power-law probability distribution (PL), in its discrete version

\[
q(k) = Ak^{-\alpha}, \quad k = k_{\min}, k_{\min} + 1, k_{\min} + 2,\ldots \quad (2)
\]

since it may represent a good candidate to describe the tail of the distribution [30]. We recall that \(A^{-1} = \sum_{n=0}^{\infty} \frac{1}{(k_{\min} + n)^{\alpha}}\) represents the normalising factor of the probability distribution [31]. To estimate the scaling parameter \(\alpha\) of the distribution in such a way that the PL fits the data for \(k \geq k_{\min}\), we followed the algorithm proposed by Clauset et al. [31]. In short, the fitting procedure provides the best estimate for the parameters \(k_{\min}\) (called \(k_{\text{PL min}}\)) and \(\alpha\) by means of MLE and minimisation of KS statistics. Furthermore, bootstrap techniques were used to assess parameter standard deviations (std). We generated synthetic data and obtained a \(p\)-value through KS statistics to indicate the goodness of the fit, as for the NB distribution [31].

Another aggregated distribution used for describe pattern of macroparasites [29] is the Poisson-LogNormal (PoiLN) distribution,

\[
q(k) = \frac{(2\pi\sigma)^{\frac{1}{2}}}{k!} \int_0^{\infty} \lambda^{k-1} e^{-\lambda} e^{-\frac{((\log(\lambda) - \mu)^2}{2\sigma}} \, d\lambda, \quad k = 0,1,2,\ldots \quad (3)
\]

firstly introduced by Bulmer [32] and used in several fields for its capability in describing aggregated data, e.g. [33–35]. As for the NB distribution, we used a MLE method to estimate parameters \(\mu\) and \(\sigma\) on the entire data set. Uncertainty on the parameter estimation was assessed by bootstrap techniques. Moreover, in line with analysis performed for the NB distribution, we also explored the capability of the PoiLN distribution to describe a tick burden larger than a certain threshold \(k_{\min}\) by coupling KS statistics and bootstrap procedures.

### Table 1. Basic descriptive statistics for empirical data.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># of grids</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td># of sessions</td>
<td>16</td>
<td>15</td>
<td>15</td>
<td>13</td>
<td>10</td>
<td>15</td>
<td>15</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>sum of feeding ticks</td>
<td>14376</td>
<td>7073</td>
<td>6550</td>
<td>2426</td>
<td>3063</td>
<td>7361</td>
<td>4077</td>
<td>5821</td>
<td>4685</td>
</tr>
<tr>
<td>median of ticks per rodent</td>
<td>9</td>
<td>14</td>
<td>11</td>
<td>14</td>
<td>11</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>ranges of ticks per rodent (1,103)</td>
<td>(1,102)</td>
<td>(1,78)</td>
<td>(3,88)</td>
<td>(1,95)</td>
<td>(1,111)</td>
<td>(1,93)</td>
<td>(1,85)</td>
<td>(1,77)</td>
<td></td>
</tr>
<tr>
<td>nymphs fraction</td>
<td>0.5%</td>
<td>2.2%</td>
<td>2.7%</td>
<td>1.1%</td>
<td>3.0%</td>
<td>0.8%</td>
<td>2.0%</td>
<td>0.7%</td>
<td>1.0%</td>
</tr>
</tbody>
</table>

Number of trapping grids, trapping sessions, total number of *A. flavicollis* captures for different years, sum of feeding ticks, median and ranges of the number of ticks per rodent, and mean number of nymphs fraction among feeding ticks.

DOI:10.1371/journal.pcbi.1003931.t001

### Table 2. Nymphs to total ticks ratio for observed feeding ticks on mice.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.1</td>
<td>1.0</td>
<td>4.0</td>
<td>0.3</td>
<td>2.0</td>
<td>1.1</td>
<td>1.8</td>
<td>0.6</td>
<td>1.6</td>
</tr>
<tr>
<td>B</td>
<td>0.5</td>
<td>1.6</td>
<td>0.7</td>
<td>1.0</td>
<td>4.6</td>
<td>0.6</td>
<td>1.8</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td>C</td>
<td>0.4</td>
<td>1.2</td>
<td>2.8</td>
<td>1.3</td>
<td>1.7</td>
<td>0.4</td>
<td>3.1</td>
<td>1.2</td>
<td>1.4</td>
</tr>
<tr>
<td>D</td>
<td>0.2</td>
<td>1.8</td>
<td>3.6</td>
<td>1.0</td>
<td>3.9</td>
<td>0.9</td>
<td>1.4</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>E</td>
<td>0.4</td>
<td>4.4</td>
<td>1.6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F</td>
<td>0.9</td>
<td>2.8</td>
<td>2.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>G</td>
<td>0.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>H</td>
<td>1.1</td>
<td>1.5</td>
<td>3.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>I</td>
<td>0.7</td>
<td>10.4</td>
<td>3.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Percentage of feeding nymphs on the total feeding ticks observed on mice in different years (2000–2008) and grids (A–I).

DOI:10.1371/journal.pcbi.1003931.t002
Finally, we compared the PL hypothesis in fitting the tail of the real data distribution with the two alternatives NB and PoiLN by a log-likelihood ratio (LLR) test for different values of $k_{\text{min}}$. In particular, since the distribution models are non-nested, we used the method proposed by Vuong [36] to understand whether the sign of such test was statistically significant or not.

Beyond the estimate of the ticks-per-host distribution, we also investigated how the tick burden distributions vary over time and whether a significant difference was observed when different time periods were considered. In particular, we investigated the tick aggregation patterns during periods characterised by low and high $A.\ flavicollis$ abundance. To achieve this goal we smoothed the time series of captured mice with a quadratic polynomial curve. The parabola describing the mice abundance in a specific year and grid was normalised between 0 and 1 before isolating the time window where this normalised parabola was higher than a threshold value $h \in [0,1]$, thus identifying the peak time of mice abundance, as reported in Figure 2. The distribution of ticks feeding on mice has been evaluated and compared considering in- and out-of-peak time periods for different values of $h$. We calculated the KS statistic between the in- (high abundance) and out-of- (low abundance) peak time distributions of tick burden, and we then compared the value observed in real data to a bootstrapped data set in order to establish whether this measure was statistically significant. For this purpose we generated $10^5$ synthetic in-and out-of-peak samples having the same size as the observed ones. As a test of soundness, we then calculated the fraction of the KS statistic that is larger in synthetic data than on real data.

**Larval and nymphal aggregations patterns on mice.** A necessary condition for an effective non-systemic transmission of a pathogen is the coincidence of the larval and nymphal aggregation distributions on hosts [8]. Therefore, our first step was to examine the association between the number of larvae and that of nymphs on each host using Spearman’s rank-correlation coefficient [37]. In particular, we preferred a non-parametric method rather than the more commonly used Pearson’s correlation coefficient since tick distributions are aggregated (i.e. deviate from normal distribution) and we were more interested in any monotonic relations of our variables than in the linear relation depicted by the Pearson’s coefficient. More in detail, a positive [negative] Spearman’s coefficient would indicate that an increase in the number of nymphs per mouse is associated with an increase [decrease] of the number of larvae per mouse. Therefore, a positive Spearman’s correlation coefficient could be interpreted as an indicator of the coincidence of the distributions, a zero coefficient could suggest the independence of the two distributions, and a negative coefficient, an uncommon result, would be an indicator of having two unimodal distributions with two asynchronous peaks. Moreover, to evaluate the significance of Spearman’s coefficient (i.e. the probability that the same coefficient could be obtained by chance) we implemented a permutation test. In particular, we compared the evaluations on synthetic datasets with a reshuffled number of nymphs and on the original data and counted the number of times that the absolute value of Spearman’s coefficient was larger than

---

**Figure 1. Temporal variation of $A.\ flavicollis$ mice abundance recorded in different grids (labelled in different colours from A to I).**

**Figure 2. Detection of seasonal abundance time-windows.** The time series of captured mice has been interpolated by a quadratic polynomial curve. By normalising the obtained parabola to unity and setting a threshold $\theta$ ($\approx 0.5$ in the example), we identify mice captured in high abundance season, those above the threshold $\theta$ (triangles), and mice captured in low abundance period, those below the threshold (circles).
Simulations of Tick-Borne Disease Spreading via Non-systemic Transmission

In order to explore the impact of different parasite aggregation distributions on the spread of a TBEv-like pathogen where the main transmission route is through co-feeding, we performed extensive numerical simulations informed by the data about tick aggregation on mice. In this setting, tick larvae were not infective (transovaric transmission has been indicated as negligible [38]), and adults only rarely feed on mice (on our data set, adults ticks are about 0.05% of the total number of ticks feeding on mice), and the only transmission link that we considered was the co-feeding between infective nymphs and larvae. Therefore, the only actors in our model were nymphs and larvae feeding on hosts. Moreover, Rosà and collaborators suggested in a recent work devoted to the same geographical area [21] that the larvae that feed in one year generally quest and feed as nymphs in the following year.

Rosa and collaborators suggested in a recent work devoted to the main transmission route is through co-feeding, we performed simulations in tick aggregation as a function of seasonality. At each year, we classified mice as observed during the mice peak activity (the remaining mice, with 0 < γ < 1) and observed out of the peak (γ = (1 − γ)N0). Therefore, we assigned the number of ticks feeding on mice according to the respective aggregated distributions qIN and qOUT. Moreover, since the larvae obtaining a blood meal at year t will be nymphs at year t + 1 without any other involvement in the epidemic spreading at year t, [21], the meta-code is sufficient to suitably describe the seasonal variation in the epidemic process. More explicitly, the epidemic dynamic in the presence of seasonality in tick aggregation may be described by the following meta-code:

\[ k(i) \text{ is the number of ticks it feeds, being } k(i) \text{ a number drawn from the probability distribution } q_{iN}, \text{ if the mouse was labelled as observed during the mice peak activity, or } q_{OUT}, \text{ if not} \]

\[ \text{of the } k(i) \text{ ticks, } f(k) \text{ are nymphs, the remaining larvae} \]

\[ \text{of the } f(k) \text{ nymphs, a fraction } \pi_L(t−1)/f(k) \text{ are infective, the other susceptible} \]

\[ \text{non-systemic transmission between infective nymphs and larvae on the same host occurs with probability } \beta \]

\[ \pi_L(t) \text{ is updated as the fraction of larvae infected} \]

(c) if } \pi_L(t) \text{ is equal to zero we stop the loop}

It is worth stressing that in the previous meta-code we did not consider ticks recovering from the infection, since we assumed that a feeding infective nymph at time t will exit the infectious dynamics by molting to the adult stage or dying.

We also modified the previous dynamics to deal with different distributions in tick aggregation as a function of seasonality. At each year t, we classified mice as observed during the mice peak activity (γ = γN0 mice, with 0 < γ < 1) and observed out of the peak (γ = (1 − γ)N0). Therefore, we assigned the number of ticks feeding on mice according to the respective aggregated distributions qIN and qOUT. Moreover, since the larvae obtaining a blood meal at year t will be nymphs at year t + 1 without any other involvement in the epidemic spreading at year t, [21], these modifications to the meta-code are sufficient to suitably describe the seasonal variation in the epidemic process. More explicitly, the epidemic dynamic in the presence of seasonality in tick aggregation may be described by the following meta-code:

1. for } t \text{ between } t_0 \text{ and } t_{\text{max}}:

(a) a fraction } \gamma \text{ of the } N_0 \text{ mice are labelled as observed during mice peak activity (the remaining } = (1 − \gamma)N_0 \text{ as observed out of the peak window} \]

(b) for each mouse } i, \text{ with } i \text{ between } 1 \text{ and } N_0:

\[ k(i) \text{ is the number of ticks it feeds, being } k(i) \text{ a number drawn from the probability distribution } q_{iN}, \text{ if the mouse was labelled as observed during the mice peak activity, or } q_{OUT}, \text{ if not} \]

\[ \text{of the } k(i) \text{ ticks, } f(k) \text{ are nymphs, the remaining larvae} \]

\[ \text{of the } f(k) \text{ nymphs, a fraction } \pi_L(t−1)/f(k) \text{ are infective, the other susceptible} \]

\[ \text{non-systemic transmission between infective nymphs and larvae on the same host occurs with probability } \beta \]

\[ \pi_L(t) \text{ is updated as the fraction of larvae infected} \]

(d) if } \pi_L(t) \text{ is equal to zero we stop the loop}

Results

Ticks Burdens

The probability distribution of tick burden on mice was skewed and showed a heavy tail. The best fit of the NB distribution was obtained on the largest available subsets of data, i.e. with } \kappa_{\text{min}} = \kappa_{\text{max}} = 1, \text{ see left panel of Figure 3. In this setting, the MLE method estimated } r = 1.30 (95\% confidence intervals (CI) = 1.25,1.35) \text{ and } p = 0.10 (95\%CI = 0.09,0.10). \text{ However, the GOF of the NB distribution was very low (p < 10^{-3}) for any value of } \kappa_{\text{min}}, \text{ see central panel of Figure 3, thus giving evidence for rejecting the hypothesis of the NB functional form. Similarly, the best fit of PoI LN distribution was achieved on the largest subsets of data, } \kappa_{\text{min}} = \kappa_{\text{max}} = 1, \text{ see left panel of Figure 3. In this case the estimated parameters were } \mu = 1.96 (95\%CI ± 1.92,1.99) \text{ and } \sigma = 0.99 (95\%CI = 0.96,1.02). \text{ The GOF of the PoI LN, central panel of Figure 3, suggested that PoI LN was acceptable only for } \kappa_{\text{max}} > 38. \text{ However, for } \kappa_{\text{min}} > 38, \text{ the KS statistic displayed values that were too large to consider the PoI LN distribution appropriate for describing real data.}

On the other hand, by fitting the tail of the distribution to a PL distribution, we found that the best fit was obtained for } \kappa_{\text{min}} = \kappa_{\text{max}} = 38 \\text{ (with a standard deviation of 5.83), see left panel of Figure 3. This } \kappa_{\text{min}} \text{ value is matched with an estimated scaling}
parameter $\alpha = 4.27$ (with standard deviation = 0.41). The GOF test ($p$-value larger than 0.1) suggested that the optimum PL fit on the tail of the distribution should not be ruled out, and that the result holds for every PL fit with $k_{\text{min}} > 35$ see center panel of Figure 3. Finally, the LLR test highlighted that the PL fitting is to be preferred ($p < 10^{-3}$) to the NB in describing the tail of the distribution for a large range of lower bounds, $k_{\text{min}} \in [8,44]$, see right panel of Figure 3. Similarly, the PL is to be preferred to the PoiLN for $k_{\text{min}} \in [5,54]$. Moreover, it is worth to stress that for values above 44 (55) the sign of the LLR test still indicates the PL fit as the preferred one compared to the NB (and PoiLN), although the indication loses statistical significance due to the scarcity of available data.

In Figure 4 we show the complementary cumulative probability distribution of the best fits resulting from $k_{\text{min}}^{\text{NB}} = k_{\text{min}}^{\text{PoiLN}} = 1$ for NB and PoiLN distributions and $k_{\text{min}}^{\text{PL}} = 38$ for PL distribution against field data of the number of ticks per mouse. From this plot we noticed that above a certain number of ticks per mouse NB [PoiLN] under-estimates [over-estimates] the tail of the distribution (indeed both fits were statistically evaluated as very poor). At the same time, in agreement with statistical results summarised in Figure 3, we noticed that the PL fit in Figure 4 more appropriately describes the right tail of the data distribution.

The number of mice captured in different years and grids showed strong seasonal patterns as reported in Figure 1. For each grid and each year we defined two separate periods depending on the mice abundance as defined in section “Data Analysis” and sketched in Figure 2. Imposing a threshold $\theta$, for each year and grid we identified a time window of high mice abundance. With $\theta = 0.5$ we found significant evidence that the distribution of ticks on mice within the abundance peak was different from that observed outside. Indeed, the fraction of the KS measures calculated on the synthetic samples lower than the real-data KS statistic was almost 98%, thus indicating very low confidence in obtaining the same measurement by chance. The same statistical evidence was also obtained by using different time window thresholds (such as, $\theta = 0.4$ and 0.6).

On the data sets classified as inside (IN) and outside (OUT) the time window of mice abundance peak, we fitted for different time-window lengths ($\theta = 0.4, 0.5, 0.6$) the parameters $t$ and $p$ for NB distribution (Figure 5, left panels) and $\alpha$ and $k_{\text{min}}$ for PL distribution (Figure 5, right panels). We observed a larger PL scaling parameter $\alpha$ inside the mice abundance peak than outside (two-sample $t$-test output: for $\theta = 0.5$, $t$-statistic = $-74.95$, df = 1931, $p < 10^{-3}$) indicating a larger heterogeneity in tick burden outside the abundance peak time. Moreover the GOF test indicated a rejection of the NB fit in both sets (IN and OUT) with $\theta = 0.5$. On the other hand, the GOF test with $\theta = 0.5$ showed that the PL model cannot be ruled out in both sets ($p$-value > 0.1) and the LLR test indicated that the PL fitting outperforms the NB model ($p$-value < 0.05) in the estimates both inside and outside the peak time window.

The distribution of larvae and nymphs on mice are coincident rather than independent, and indeed the same 20% most infested hosts feed both 55% of the nymphs and 54% of the larvae. Moreover, Spearman’s correlation coefficient measured on the number of larvae and nymphs on mice was positive (0.24) and the probability that this coefficient was detected by chance was very low (the empirical value was the largest if compared to those evaluated in $10^3$ reshuffled samples). In addition, the mean number of larval co-feeding with a nymph is about 23 which is almost double the value that would be seen if the distributions were independent (mean equal to 12).
Non-systemic Disease Spreading Simulations

To start, we simulated the non-systemic disease spreading of a TBE-like pathogen with a fraction \( f \) of nymphs among ticks equals to 2\%, close to the one observed in our real data (cfr. Table 2), 5\%, and 10\%, as in literature [8,40]. We consider the empirical distribution observed on the entire data set. We fixed the number of hosts to \( N_h = 10^4 \) which, together with the considered distribution, resulted in a number of vectors pairs equal to \( N_T \sim 10^5 \). In our simulations, we explored the effects of \( \beta \), the infection probability, on the observed prevalence at the final time step, \( \pi_L(t_{\text{max}}) \), with \( t_{\text{max}} = 1000 \). (We observed that \( t_{\text{max}} = 1000 \) was larger enough to allow the prevalence to converge toward an endemic pseudo-equilibrium or the disease-free equilibrium). For each \( \beta \) we allowed 200 simulations to run starting from an initial prevalence of \( \pi_L(t_0) = 1\% \). In Figure 6 we plotted the prevalences (median value, interquartile intervals and the 95\%CI) observed at equilibrium as a function of the transmission probabilities, \( \beta \). Results showed that the larger the fraction of nymphs among ticks feeding on mice, the larger the probability of pathogen invasion and the infection prevalence.

Then, we explored the effects of different tick burden distributions on the spread of infection. To this end we considered four distributions: PL, NB, PoiLN, and the empirical distribution on the entire data set (aggregated on capture sessions and grids). For synthetic distributions we considered the actual observed distribution below the estimated \( k_{\text{min}} \), while we used the best fit of synthetic distributions to describe values greater than \( k_{\text{min}} \). Again, we fixed the number of hosts to \( N_h = 10^4 \). It is worth stressing that in the synthetic samples generated from these distributions we observed some features similar to those observed in real-data. For instance, the number of nymphs was positively associated with that of larvae and more particularly a nymph co-fed with a mean number of larvae similar to that observed in reality (for PL the mean number was 23, for NB 20, and for PoiLN 27).

Results, plotted in Figure 7 for \( f = 2\% \) and in Text S2 for \( f = 5\% \) and \( f = 10\% \), corroborated the hypothesis that the transmission probability needed for the pathogen to become endemic is driven by the shape of the tail of the distributions. In particular, we noticed that for the PoiLN distribution (the one with larger fitted tail) the epidemic threshold is the lowest, while for the NB distribution (the one with smaller fitted tail) the infection probability needed for invasion is the highest. Not surprisingly, the PL, which has the best performances in fitting the tail of the empirical distribution, is the one for which the prevalences at equilibria better resemble those observed in simulations using the empirical distribution. We also performed some sensitivity analysis on parameter distributions, further highlighting that the larger the tail of the distribution, the lower the epidemic threshold (see Text S1). In addition, sensitivity analysis on the fraction of nymphs (\( f \)) showed that \( f \) does not qualitatively influence the epidemic behaviour (see Text S2).

Furthermore, we investigated the effect of differences in the distribution of the tick burden as a function of the abundance of mice on the spreading of a non-systemic infectious disease. To this end, we fixed \( \gamma = 0.89 \), as measured in the dataset, and as \( q_{\text{IN}} \) we considered a PL with exponent \( \alpha_{\text{IN}} = 4.39 \) as estimated with \( \theta = 0.5 \). In a similar way, we assumed as \( q_{\text{OUT}} \) a PL distribution with exponent \( \alpha_{\text{OUT}} = 3.48 \). For both \( q_{\text{IN}} \) and \( q_{\text{OUT}} \) we further set \( k_{\text{min}} = 5,10,15 \). Results are summarised in Figure 8, from which it could be inferred that the epidemic outcome was strongly influenced by the different distributions of feeding ticks according to mice abundance. We consistently observed that the transmission probability needed for the pathogen to effectively spread was smaller when the time windows identified by mice abundance are considered.

Discussion

Tick aggregation on hosts is the result of several complex interactions of biotic and abiotic factors, such as host exposure and susceptibility to ticks, ticks’ phenology and host behaviour,
environmental factors, availability of resources, and others [27,41]. Historically, the NB distribution has been preferred to the Poisson distribution to describe parasite heterogeneity across hosts because it suitably reproduces overdispersed observations. It has also been widely used in empirical [6,25,26,28] and theoretical studies [23,24,42]. However, fat tailed distributions other than the NB

**Figure 5.** Estimated parameters of different distributions (NB on left and PL on right) obtained inside (blue) and outside (red) of the mice peak abundance time window. Time windows are defined by $\theta = 0.4, 0.5, 0.6$ (from left to right for each subsets). Vertical bars indicate best model fits (central horizontal lines) with their uncertainties that are 95\% confidence interval for NB models while standard deviations for PL models. doi:10.1371/journal.pcbi.1003931.g005

**Figure 6.** Median (line), interquartile (darker area) and 95\% confidence intervals (lighter area) of the final prevalence as a function of the transmission probability, for different values of $f(=2\%, 5\%, 10\%)$, fraction of nymphs among ticks on a mouse, and by describing the ticks aggregation with the empirical distribution. Other parameters are $N_0 = 10^6$, $t_{\text{max}} = 10^7$. doi:10.1371/journal.pcbi.1003931.g006
one can also adequately reproduce tick aggregation, as shown by Elston et al. [29] and Bisanzio and collaborators [30].

Through the use of an extensive data set of feeding *Ixodes ricinus* ticks on mice, we showed that a PL distribution is better able to describe the right tail of the tick distribution on hosts than a NB or a PoiLN distribution (see Figure 3 and 4). This finding may have relevant epidemiological consequences, since it is well documented that the heterogeneity of contact distributions among individuals has large impacts on pathogen spread and persistence [43–49]. In fact, it has been demonstrated [50] that the minimum transmission probability for a pathogen to spread on a network, the so-called epidemic threshold, is driven by the first and the second moment of this distribution. In particular, Pastor-Satorras et al. [50] demonstrated that the larger the heterogeneity, the

![Figure 7. Median (line), interquartile (darker area) and 95% confidence intervals (lighter area) of the final prevalence as a function of the transmission probability, for different fitting distributions (PL, NB and PoiLN). Other parameters are $N_h = 10^4$, $t_{max} = 10^3$, and $f = 2\%$. doi:10.1371/journal.pcbi.1003931.g007](image1)

![Figure 8. Median of the final prevalence as a function of the transmission probability. A PL distribution of vectors-per-host has been considered in all scenarios. Simulations that consider different aggregation behaviours according to the temporal window of mice abundance (red) are compared with others with a fixed distribution (blue). Other parameters are $N_h = 10^4$, $t_{max} = 10^3$, and $f = 10\%$. doi:10.1371/journal.pcbi.1003931.g008](image2)
lower the epidemic threshold for the pathogen to spread, with an interesting behaviour in infinite size network showing a zero epidemic threshold [46]. Thus, the epidemiological inferences on the spread of a pathogen are highly influenced by the characterisation of the connectivity distribution and in particular by the distribution tail (i.e. the heterogeneity). Our results corroborate those findings and generalise them in a different framework and for more complex transmission routes, i.e. a vector-host network for non-systemically transmitted diseases. In particular, we found that the tail of the distribution of the number of ticks per rodent highly influences pathogen spreading (see Figure 7 and Text S1). Furthermore, it is worth remarking that although the tail of the distribution as defined here represents about 5% of the entire data set, our simulation findings suggest that this small part of the distribution is crucial for pathogen invasion.

We also confirm that the probability of pathogen invasion and the infection prevalence are strongly influenced by the fraction $f$ of nymphs on the total feeding ticks on mice (Figure 6 and Text S2). The co-occurrence of larvae and nymphs on competent hosts is in fact essential for the horizontal transmission of non-systemic transmitted tick-borne pathogens, such as TBE, and it has been documented, both empirically and theoretically, that it could be a key factor in creating TBE hotspots, [51,52].

Our conclusions confirm previous findings showing that the distribution of ticks on rodents may significantly affect the spread of infections [27,30,53], especially for non-viraemic transmitted diseases such as TBE [7,23,24]. Under the hypothesis of a NB distribution of ticks across hosts, both Rosá et al. [24] and Harrison and collaborators [23] showed that highly coincident and aggregated distributions favour the establishment of TBEv. However, highly heterogeneous degree distributions do not necessary imply a higher spread of disease. Indeed, Piccardi et al. [54] showed that scale-free networks can be much less efficient than homogeneous networks in favouring the disease spread in the case of a nonlinear force of infection.

The correct description of tick aggregation on hosts could dramatically affect disease control strategies: for instance, Perkins [7] emphasised that an optimised control effort targeted on highly parasitised mice, also identified as sexually mature males of high body mass, could significantly lower the transmission potential. On the other hand, Brunner and colleagues [27] observed that the identification of individuals which fed a disproportionate number of ticks (and that can therefore act as superspreaders) can be challenging, since simple covariates such as sex, age or mass do not entirely explain the differences in parasite burden.

In order to fully understand the different tick attachment behaviours on hosts, we identified different time windows related to rodent seasonal dynamics. Using this approach we found that the distribution of ticks on mice may vary across the season, with higher aggregation heterogeneity in periods of low rodent abundance and lower aggregation heterogeneity during the peak of host abundance (see Figure 5). We also showed that seasonal aggregation patterns, characterised by larger tails in time periods of low host abundance, enhance the spread of non-viraemic transmitted diseases (see Figure 8). Shaw and collaborators [26] observed significant variations in the degree of aggregation between host subsets – stratified by sex, age, space or time of sampling – in several host-parasite systems. In agreement with our results (lower aggregation in period of high mice abundances as shown by estimated exponents of PL), they found that aggregation in copepod (Lepeophtheirus pectoralis) infesting plaice (Pleuronectes platessa) decreases during summer months. They mainly ascribed the observed variation to significant differences in mean parasite burden among months. On the other hand, we did not find significant differences in tick burden inside and outside the window of high rodent abundance. Specifically, in the case of $\theta=0.4,0.5,0.6$, the average number of ticks per host were 11.96,11.96,11.84 inside the window of high rodent abundance and 12.34,12.23,12.78 outside and the differences between inside and outside are not statistically significant (permutation tests, $p>0.05$). However, the second moment of the number of ticks per host drastically changed between high and low abundance periods, driving the difference in the aggregation distributions observed in the two time windows. Seasonal variations in resource availability and host abundance can have a significant effect on the space used by mice. Males and females tend to respond to these changes in different ways, since space use for females is driven largely by food availability, whereas the distribution of males is related primarily to mating opportunities. Yellow-necked mice (A. flavicollis) females exhibited reduced spatial exclusivity and larger home ranges during lower food availability while males varied their spatial distribution accordingly by also expanding their home ranges [55]. An inverse relationship between population density and home range sizes has also been observed in wood mice (Apodemus sylvaticus) [56]. Consequently, in periods of low rodent abundance more mobile rodents, especially males, are more likely to hit a patch of larval ticks. As a result, these individuals would harbour a large amount of ticks and increase the aggregation of tick distribution among the rodent population. On the other hand, tick density is usually lower in periods of low rodent abundance, and the average tick burden would decrease for the rest of the population, especially females, balancing the overall tick burden. On the contrary, during times of high abundance mice move less and ticks would be distributed more evenly among the rodent population resulting in the observation of a lower aggregation in tick distribution during the peak of rodent abundance.

Our primary goal was to help understand the role of tick aggregation across mice on the spread of non-viraemic transmitted diseases through a simple and general transmission model. Other works – such as [24,42,52,57] – described in very fine detail the transmission of vector-borne diseases, introducing different transmission routes, tick stages and alternative hosts in the epidemic model. For instance, Norman and colleagues [57] demonstrated through an epidemiological model that non-viraemic transmission could have non-negligible effects on the persistence of a disease like the Louping ill. Here, considering the non-systemic transmission only, we explored the effect of using different theoretical functional forms to describe the tick burden on hosts. By estimating parameters of the burden distributions on a very detailed data set, we defined a simple and transparent transmission model that explicitly takes into account the real contact pattern of vectors and hosts in the description of a non-systematically transmitted vector-borne disease. In this way we were able to emphasise that, while the NB and PoisLN models can sufficiently fit the whole real distribution, the PL model represents a better fit for the distribution tail. Furthermore, the vector perspective approach used in our model gives better insights into the dynamics of non-systemically transmitted pathogens respect to host perspective models that were more commonly and widely used in this context [24,42,52,57]. In addition, epidemiological simulations parameterised by the fitted tick burden distributions highlighted the epidemiological consequences of describing tick aggregation on hosts through distributions with different tails, showing that the shape of the tail distribution has a non-negligible influence on pathogen persistence. Future works will be devoted to extend the present findings to more complex transmission dynamics (e.g. including viraemic or transovaric transmission), in order to assess
the effect of a PL decay of the distribution for a wider range of vector-borne diseases.

**Supporting Information**

**Text S1** Sensitivity analysis of distribution parameters on epidemic spreading. (PDF)

**Text S2** Sensitivity analysis of $f_\lambda$ fraction of nymphs among ticks. (PDF)

**References**