Rodó et al. 10.1073/pnas.1400380111

Supporting Information

SI Data and Methods

Due to the low incidence of disease, particular attention was given to the most populated cities in Japan, spanning the country from North (Sapporo) to the center (Tokyo, Yokohama, Saitama, and Chiba, forming the Greater Tokyo Area) and more to the south (Nagoya). Days within the three main epidemics (1979, 1982, and 1986, hereafter referred to as “epidemics”) were treated separately from the rest of the Kawasaki disease (KD) time series (postepidemic), so that a comparison between results of analyses from epidemic periods and nonepidemic periods could be made. Incidences were also calculated from population censuses, for a comparison with results from KD cases time series. To this end, the KD case record at Tokyo was used as the reference series and the 95% level of the distribution of KD cases there was used as the reference value defining a 95% KD anomaly (five cases per day). Comparisons were also made between analyses derived using extreme case occurrence (95%) vs. those using the less extreme (70%) anomalies within each epidemic, to identify the different regions contributing at different phases of the epidemics. For the analysis in the postepidemic period (1987–2010), dates with low KD cases were selected to compare with situations with high KD. The choice of lows was made among days having zero cases and being preceded by at least 10 d with no H events. A comparison was also made to assess stability of results with dates selected from a KD series for Tokyo that was filtered to remove any potential influence of both low- and high-frequency variability. For the filtered series, periods shorter than the weekly cycle were removed with a low-pass filter, and periods longer than 1 mo were filtered with the application of a recursive Butterworth filter with a cutoff period of 30 d (1). Results obtained for Tokyo did not significantly differ from those with the raw data due to the small contribution to KD variability arising from the seasonal and longer-term cycles, in comparison with day-to-day variance (1% and 2.5% extreme data did not significantly differ either between the raw and the filtered series).

KD incidence in the Heilongjiang province of northeastern China has been reported to be much lower than that for Japan (*, 2). However, when population densities and areas are made comparable, relative KD numbers appear less divergent than reported despite the large uncertainty in the reported values according to different sources. Population in the whole Heilongjiang prefecture (38.2 million people) is spread throughout a vast floodplain of more than 454,000 km², with average population densities of 83 people per square kilometer in Heilongjiang. Instead, Japan’s population is by comparison much larger (>127 million people in 2012) and population density oscillates around 330–345 per square kilometer. In contrast to Heilongjiang, the most densely populated area in Japan (that of the Tokyo 23-ward area) covers about 622 km² with the total population showing an upward trend (8.97 million residents—a population density of about 14,422 people per square kilometer as of October 1, 2011). People in the Tokyo “23-area” live in a very small extension of land, which makes it around 174 times the mean difference in population density between the two areas (83 people per square kilometer in Heilongjiang for 14,422 per square kilometer in the Tokyo 23-area or 8,817 per square kilometer in the Greater Tokyo Area, or 72.22 times denser if we consider density for the central Harbin metropolitan area to be 200 people per square kilometer).

Trajectories traced back in time 10 d for each dataset and location were generated using the flexible particle Lagrangian dispersion model (FLEXPART Version 8.23) run in backward mode. The particles modeled were air tracers, with 10,000 particles used on each model run. Residence time is the collective amount of time that a particular area upstream of, or including, the selected location was overlain by any of the air parcels from the trajectories in the sample set. The model output used was residence time, with an output grid of 0.5° latitude × longitude and a time resolution of 3 h. The data input used for the FLEXPART model was gridded atmospheric wind velocity from the European Centre for Medium-Range Weather Forecasts Re-Analysis (REA)—Interim at 1°. Atmospheric REA (3) is a global atmospheric model, with assimilated historical observational data, run over a multidecade span to provide a dynamically consistent, global representation of the weather and climate. Atmospheric REA has captured 1979–2012 time-varying principal modes of cool season wind and weather during the cool season in Eurasia, along with the associated patterns of anomalous precipitation and anomalous temperature (4).

From the trajectories, we calculated ensembles averaging the residence time of FLEXPART simulations from dates with high KD and dates with low KD for each location. An arbitrary residence time selection threshold of 30 s representing the 95% value of the overall distribution of KD cases is considered for taking into consideration extreme KD residence time in each 0.5° grid point. Simulations integrated the residence time for the total column, 0–3,000 m. The high (H) and low (L) KD date samples were constructed with the same overall number of days per months for H and L dates to remove effects of the seasonality of the wind circulation. For nonepidemic events, H dates were dates from winter months (October–March) from the period 1987–2010 where the KD index is greater than the 95th percentile. L dates were again days from winter months (October–March) from the period 1987–2010 for which the KD cases were zero. Anomalies were calculated by subtracting the residence time of L ensembles from the residence time of H ensembles, considering for representation only grid points with concentrations over 30 s in the ensemble. From the daily trajectories, we calculated ensembles averaging the residence time of FLEXPART from dates with H and dates with L separately. Again, a threshold of 30 s was used in counting residence time exposure at each grid point. For epidemic conditions case selection was as follows. For percentile 70 (p70), days with KD values above this threshold were chosen and only from the rising phase so that no simulated day lay outside the p70 interval. In this case, days selected should also have a minimum separation of 12 d from the first p95 d. A first threshold for graphic representation is also set at 30 s, corresponding approximately to the 95% of residence time in each case. Simulations were averaged over the whole ensemble of back-trajectory days and the residence time plotted on a logarithmic scale. For p95 dates, selection was made from the maximum values in each epidemic, as in p70, but with no simulated day being out of the p95 interval. In the composite p95/p70 plots, all positive residence time in the three epidemics (for convenience, this time over the 20-s threshold) were pooled together and their average plotted with different colors denoting the different groups (p70, green; p95, blue). Orange areas denote overlap and brown dots indicate croplands.

Atmospheric data were derived from the National Centers For Environmental Prediction/National Center for Atmospheric Research monthly reanalysis (5) for the construction of Fig. S1. In this case, a Butterworth filter of 18 mo was used to calculate the interannual component of detrended sea-level pressure (SLP) and meridional winds. This component was then standardized by removing the local grid-point mean and the difference divided by the local grid-point SD.

Scale-dependent correlation (SDC) analysis (6, 7), a technique designed to pick transitory signatures in the time domain, was used with both the wind and the KD data because most of these observational time series are nonstationary and may contain temporary signals resulting from noncontinuous external forces. Under such conditions, classical time-series analysis techniques are not accurate enough to extract all of the signals in the proportion they contribute, and therefore time-scale decompositions such as those resulting from SDC analysis are better suited. A range of scales between 1 d and 1 mo was inspected both in one-way SDC (for the individual series of KD in the four cities of the Greater Tokyo Area) and with a two-way SDC among pairs of locations. Significance of SDC results shown in Table S1 is at least $P < 0.04$.

**SI Exploration of SEIR Potential Framework for KD**

We assume that the disease can be modeled using a susceptible–exposed–infected–recovered (SEIR) framework. Individuals are assumed to be homogeneously distributed and their movement is described by the diffusion equation. The 2D system is

$$\frac{\partial S}{\partial t} = -\beta IS/N + D \left( \frac{\partial^2 S}{\partial x^2} + \frac{\partial^2 S}{\partial z^2} \right)$$

$$\frac{\partial E}{\partial t} = \beta IS/N - \sigma E + D \left( \frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial z^2} \right)$$

and

$$\frac{\partial I}{\partial t} = \sigma E - \gamma I + D \left( \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial z^2} \right).$$

We assume an incubation period of 2 h ($\sigma = 12/d$) and a diffusion coefficient of $D = 1 \text{ km}^2/d$. We also used what is possibly the shortest incubation time recorded for an acute respiratory disease, namely 0.6 d (95% confidence interval 0.5–0.7) for influenza B. The system is solved on a 120-km grid, with a single infected individual placed at the center. The equations are discretized using the implicit Crank–Nicolson method with a spatial step of $\Delta x = 2 \text{ km}$ and a time step of $\Delta t = 0.1 \text{ d}$. The constant population of $N = 31$ million individuals are equally distributed among all of the grid points (see Fig. 4 and Fig. S7 for further results). Other disease configurations were also checked as alternative settings, but they always resulted in larger times to cover these same distances.

The system admits traveling wave solutions for the infected population, which spread out from the initial infectious seed. We define the wave speed as the distance traveled by the peak of the wave in unit time. This is calculated by determining the time from when the peak is achieved at the initial infected site and the time that the peak appears at a point 30 km from the initial site, to mimic average distances between different cities in the Tokyo metropolitan area. The wave speed is thus calculated for various values of the parameters $\gamma$ and $\beta$ (Fig. S7 A and B). Maximum wave speeds are observed (for the range of values considered) when $\beta = 10$ and as $\gamma \rightarrow 0$ and the maximum speed is 5.66 km/d (Fig. 4).

The number of infected individuals at the peak of the wave is shown in Fig. S7B. The severity of the epidemic decreases as the basic reproductive number, $R_0 = \beta/\gamma \rightarrow 1$, with no wave detected (absence of a wave is defined as $I < 1$ at the wave peak) when $R_0 < 1$. The evolution of the epidemic is shown at a fixed point in the spatial grid ($x = 30 \text{ km}$ from the initial site) in Fig. 4 for $\beta = 10 \text{ d}^{-1}$ and $\gamma = 0.1 \text{ d}^{-1}$. Initially no infection is present. The infection reaches the 30-km mark at $t \sim 5 \text{ d}$ into the outbreak and achieves a maximum on day 7. The evolution of the infectious wave shows that during the first 2 d of the outbreak the wave increases to a maximum before maintaining this value for $\sim 10 \text{ d}$, as it travels away from the initial site, and then finally decreasing as the epidemic burns out.

**SI Atmospheric Pollutants**

Trend, interannual variability, and seasonality decompositions were applied to daily data series of atmospheric pollutants over Tokyo ($\text{SO}_2$, $\text{O}_3$, NO, NO$_2$, NO$_x$, CO, and NMHC) from three stations in the Tokyo metropolitan area (Kokusetsu Tokyo, Higashikoujiya Oota-ku, and Sisibone Edogawa-ku) from April 1983 to January 2012 with negative results (Data are provided by the Environmental Numerical Databases at National Institute for Environmental Studies, and the Bureau of Environment, Tokyo Metropolitan Government). As no atmospheric measurements of Hg were made available to us at the time of the submission for Tokyo, direct comparison with KD data series was not possible. Despite the former, correlations between Hg have been shown to be very high with NMHC, $\text{SO}_2$, NO, NO$_2$, NO$_x$, and CO in different seasons in Korea (5).

However, despite some controversy existing between increased urinary mercury excretion levels and the possibility of a clinical diagnosis for KD (9–11), Tokyo had the lowest total gaseous mercury concentrations (averaged 2.88 ng m$^{-3}$ in a study comparing different locations in East Asia (12) (Guizhou in China averaged 196.95 ng m$^{-3}$). noteworthy, the gaseous mercury concentrations in Korea have also decreased in the past few years due to reduced use of coal in Korea, whereas KD has steadily increased in the region. Although gaseous mercury concentrations were not anomalous at the time of the three large KD epidemics (1979, 1982, and 1986), concomitant side effects contributing to the severity of KD in some patients of any of the pollutants above cannot, however, be fully discarded.

**SI Air Sampling**

In these experiments, a Cessna aircraft (Cessna 172, maximum speed 220–230 km/h, climbing rate 770 feet per minute) was equipped with an aerosol-sampling device constructed by the Institut Catalá de Ciències del Clima (IC3) team in Barcelona that used 150-mm QM-H, pure quartz fiber Whatman filters (GE Healthcare). The pore size was chosen to retain >99% of particles <2.5 μm [size of inhaled particles, particulate matter (PM)2.5] and >0.3 μm. Filters were sterilized at 270 °C for 30 min and handled using sterile precautions. Before takeoff, meticulous cleaning of the air inlet and sterilization of all parts in contact with the airflow was performed. The flight path was determined by the IC3 and National Institute for Environmental Studies teams to trace a path to the northwest of Japan, against the prevailing wind (Fig. S8). After attaining an altitude of 1,000 m (~30 min), the internal manifold was opened and the sampling initiated. The aircraft gradually ascended to 3,000 m with continuous air sampling and then returned to base with the internal manifold closed on the descent at 1,000 m. The total duration of the sampling was ~3 h from 1,000 to 3,000 m and filter samples of ~90 m$^2$ (90,000 L) of air per filter were collected and sent on dry ice to the Center for Infection and Immunity at Columbia University. Aerosol samples were also collected at ground level on the same day (Fig. 5, surface filters). Negative control filters were handled in the same manner as the experimental filters but without aerosol collection also sent for analysis (Fig. 5, blank filters). The handling of filters was conducted under biosafety level 2 conditions to avoid contamination of the filters with ambient organisms or nucleic acid.
DNA was extracted from the filters using a modified PowerSoil DNA Isolation Kit (MoBio Laboratories). Extraction tubes were supplemented with UV-irradiated 0.1- and 0.5-mm-diameter glass beads. All reagents used in the extraction were first aliquoted in volumes less than 1 mL. Aliquots of reagents were UV irradiated twice at a distance of 1 inch from the UV bulbs in a Spectolinker XL-1500 UV cross-linker at 300,000 μmJ/cm² to prevent downstream amplification of contaminating DNA that may be present in the buffers. Sections (1 cm²) of the filter were cut in a sterile UV-cabinet using scissors and forceps that had been autoclaved and were baked at 95 °C overnight, UV irradiated, and flamed just before use. Each 1-cm² filter section was processed separately until the final stage in the protocol where precipitated DNA from 4 x 1-cm² filter sections from each filter were collected on individual columns to concentrate the DNA. Finally, the DNA was eluted with 50 μL of either the flight filter DNA, surface filter DNA, blank filter section, or reagent control extraction, or UV-irradiated water (PCR control). As PCR reagents can be contaminated with background DNA, we applied a protocol for decontamination of the PCR components using restriction enzyme digestion prior amplification. Briefly, a master mix was applied to all of the PCR components, except the DNA, and 2 U of Sau3A1 was added to the mixture. The master mix containing Sau3A1 was aliquoted into PCR tubes (19 μL). The reaction tubes were then incubated at 37 °C for 30 min and then placed on ice for 5 min. Then 6 μL of either the flight filter DNA, surface filter DNA, or blank filter section, reagent control extraction, or UV-irradiated water was added directly into the master mix and immediately placed into a prewarmed thermocycler at 95 °C for 15 min (this step fully inactivates any remaining activity of the restriction enzyme while activating the HotStart polymerase), followed by 40 cycles of 94 °C for 1 min, 56 °C for 1 min, and 72 °C for 1 min; followed by a final elongation step at 72 °C for 5 min. We have demonstrated that this protocol fully eliminates amplification of background DNA in PCR reagents using both fungal and bacterial 16S assays without any detectable effect on PCR efficiency. PCR products were examined with agarose gel electrophoresis and products were excised and purified from the agarose gels. PCR products were cloned into the vector pGEM-T Easy (Promega) and clone libraries were sequenced (100 clones from the flight filter and 100 clones from the surface filter). For sequence classification, database queries using the Basic Local Alignment Search Tool (BLAST) were performed using the website of the National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov). Each sequence was identified to the lowest taxonomic rank common to the top BLAST hits. Species-level identification was determined based on the closest match with greater than 97% sequence identity to sequences in the NCBI database. Where a sequence could not discriminate between two or more species in the database, a genus level classification was applied.
Fig. S1. (A) One-month leading standardized SLP and (B) wind anomalies ($V_x$) over the entire Pacific basin 1 mo before all major interannual KD peaks in the interval 1987–2010. Scale is 1 m/s.
Fig. S2. Same as in Fig. 1 but computed for the 645 dates with KD cases above the 95th percentile of the whole KD distribution of cases in the interval 1977–2010. The total atmospheric column (0–3,000 m) is included in the simulations. Brown dots indicate cropland coverage from the MODIS dataset.
Fig. S3. (A) Same as Fig. 1 but averaged for the KD rising phase days (p70; green) and the peak phases (p95; blue) in all three epidemics (p70 and p95 are calculated separately for each of the three epidemics). Orange denotes overlap between p70 and p95. (B) Same as for A but for a lag of 15 d in both p70 (green) and p95 (magenta) within epidemics. Epidemics include dates corresponding to the three largest epidemics (1979, 1982, and 1986). Orange color indicates common area and dots indicate crops.

Fig. S4. Same as Fig. 2 but for 1986.
Fig. S5. Weekly filtered time series of KD cases in the interval 1988–2008 for (A) Tokyo and Kanagawa (Yokohama) and (B) Tokyo and Saitama. Note the striking similarity in peaks and lows within the different locations in the Greater Tokyo Area (insets).
Fig. S6. (A) Average residence time per unit of time (in h) obtained for Harbin (H) for all simulations in Movie S1 generated from Tokyo (T). Two multiensembles of backward FLEXPART simulations were generated, one that corresponds to all high KD days for the postepidemic interval 1987–2010 and another one for all low KD days (at most 1 KD case per day) in this same interval. The final size of these ensembles is 3,206 dates for the low ensemble and 532 for the high. Residence time in low days corresponds to the normal situation in the winter season with a central time of 48 h. (B) Residence time vs. time (in h) for the peak maxima in both simulations. Both low (black) and high (red) days in this same interval are displayed for comparison. Peak shifts to a longer time (high, red, 66 h), this peak time shift being significant (Wilcoxon rank-sum test, \( P < 0.05 \)). The interval of time during which the air in low conditions resides in excess over H corresponds to an extra recharge time of air over the region’s surface. In A it can be seen that the fastest time a particle can take to travel from H to T is 24 h. From A it can also be concluded that the mean time particles take to travel from H to T ranges from an average maximum of 48 h to an average minimum of 24 h. The maximum confidence interval possible for the incubation time would then result from confronting the extremes of this interval, 84–24 h and 54–48 h, that is, between 60 and 6 h.

Fig. S7. (A) Wave speed of the infectious peak and (B) maximum infected population in the \( \gamma - \beta \) parameter space. The maximum number of infected denotes the time at which the disease reaches the next town (on average after 5.66 d). Conditions are the same as in Fig. 4.
Fig. S8. (A) Residence times (log$_{10}$ s) inferred from the FLEXPART model backward simulations executed from Tokyo, with the time of the atmospheric survey over Japan (March 3, 2011). Map shown is for noon local time and total residence time calculated for that day. Ten thousand particles were used for each simulation. (B) Flight trajectory on March 2011.
Table S1. Results from the SDC analysis (6, 7) between KD occurrences in the most populated prefectures within the Greater Tokyo Area (Tokyo, Kanagawa, Saitama, and Chiba)

<table>
<thead>
<tr>
<th></th>
<th>1979</th>
<th></th>
<th>1982</th>
<th></th>
<th>1986</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TOK</td>
<td>SAI</td>
<td>KAN</td>
<td>CHI</td>
<td>TOK</td>
<td>SAI</td>
</tr>
<tr>
<td>TOK</td>
<td>0.558</td>
<td>0.573</td>
<td>0.651</td>
<td></td>
<td>0.738</td>
<td>0.782</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.03)</td>
<td>(0.004)</td>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>SAI</td>
<td>0.692</td>
<td>0.701</td>
<td>0.75</td>
<td></td>
<td>0.75</td>
<td>0.767</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.012)</td>
<td></td>
<td>(0.005)</td>
<td></td>
</tr>
<tr>
<td>KAN</td>
<td>0.757</td>
<td></td>
<td>0.756</td>
<td></td>
<td>0.756</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td></td>
<td>(0.000)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Each column indicates the correlation value and the boldface denotes the maximum value attained. The window size (WS) used is 21 d for all correlations displayed (only SAI–CHI was not significant in 1979 when varying WS to 7 d; significance for all table values ranges between $0.04 < P < 0.000$, with the individual significance of the maximum correlation denoted by the $P$ value in parentheses). CHI, Chiba; KAN, Kanagawa; SAI, Saitama; TOK, Tokyo.

Movie S1. Average backward trajectory and residence times for air masses arriving at Tokyo (TKY) during the winter months (October–March) for the period 1987–2010. Units are residence time of particles on each grid cell (log$_{10}$ s). Time step representation is 6 h. The number of KD H dates used in the simulations is 532, all of them being maxima within the months October through March. Red dot denotes the location of Harbin in northeastern China.

Movie S1