lescent population genetic analysis yielded a second set of estimates for R_0: posterior median = 1.22; 95% CrI: 1.05 to 1.60 (Fig. 2C).

Third, R_0 can also be estimated from analysis of the dynamics of the epidemic within defined settings. Detailed data collected by the Mexican authorities investigating the La Gloria outbreak indicate that 616 individuals from a resident population of 1575 had acute respiratory infection between 15 February and 14 April 2009 (Fig. 3A). Data on the age distribution of cases and the dates of disease onset were used in our analysis. Figure 3B shows that the clinical attack rate varied markedly as a function of age, with 61% of individuals under 15 years old affected, dropping to 29% of people over that age. The corresponding relative risk is 2.13, with a 95% CI of 1.89 to 2.39. Based on all confirmed cases in Mexico as reported on 5 May 2009 (*1*), the corresponding relative risk is 1.52 (95% C I: 1.33 to 1.73). The overall community attack rates seen in La Gloria are comparable to (or higher than) those seen in previous pandemics (*12*).

Fitting alternative epidemic models to the La Gloria data (*8*) demonstrated that a model with heterogeneous mixing by age plus age-dependent susceptibility to infection was required to adequately fit the data with plausible parameter estimates. The resulting maximum-likelihood estimate of R_0 was 1.58 with a 95% CI of 1.34 to 2.04 (Table 2). This analysis also provided the only independent estimate of the mean generation time, T_g (1.91 days; 95% CI: 1.30 to 2.71 days) (Table 2), shorter than earlier estimates for influenza (*9*, *10*), though not significantly so. It is biologically plausible that R_0 and T_g could be correlated, because both are linked to the underlying replicative fitness of the virus. More data are needed. Owing to parameter identifiability issues, it was not possible to estimate age-dependent infectiousness, as well as age-dependent mixing, from these data. Although these estimates are informative, it should be emphasized that some uncertainties remain regarding the denominator population and that a range of other models may fit the data as well as the model choice shown here. Household data would be particularly useful in reducing remaining uncertainty.

Fourth, the time-dependent reproduction number (R_t) can be estimated from the time series of reported disease onsets among confirmed cases in Mexico (Fig. 4). These data are subject to much uncertainty because of marked changes in surveillance over the reporting interval, plus the nonspecificity of symptoms that are similar to existing and perhaps simultaneously circulating strains of influenza. However, we developed methods for analyzing such data (*8*) that account for substantial underreporting, with a change in the underreporting rate from 17 April when surveillance within Mexico was intensified. The average value of R_t estimated for Mexico up until the end of April was 1.37 (95% CrI: 1.24 to 1.59) for a model with Poisson case counts, and 1.47 (95% CrI: 1.21 to 1.88) for a perhaps more plausible negative binomial model allowing daily case counts to be overdispersed (*8*).

Given estimates of R_0 and the current epidemic size x, we can estimate the number of generations N_t of transmission of the virus among humans that is necessary to explain the current epidemic. Assuming a simple branching process with reproduction number R_0, the mean number of generations of transmission is given by $N_t = \ln(x/x_0)/\ln(R_0)$, assuming the epidemic was started by x_0 humans being infected from animal sources. Assuming $x_0 = 1$ gives estimates of N_t between 14 and 73. But even if we assume that 5% of cases were infected directly from animal sources, we obtain an estimated 5 to 22 generations of transmission, indicating sustained human-to-human transmission in Mexico.

All of the R_0 estimates are comparable with, but perhaps on the low end of, R_0 estimates

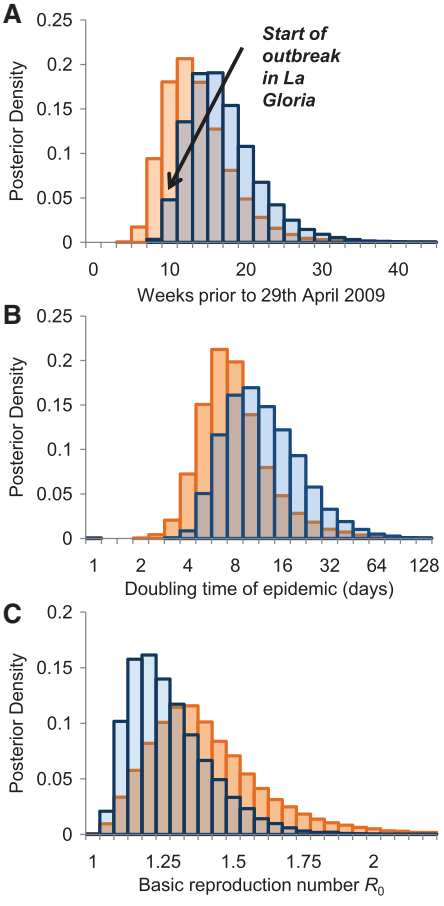

Fig. 2. (**A**) Starting from publicly available HA viral sequences, a posterior distribution of the estimated TMRCA was derived using a Bayesian coalescent model, which assumes exponential population growth (coded in BEAST 1.4), with the date of the first known human case highlighted. Details of the BEAST analysis and parameter estimates are presented in (*8*). Posterior distribution of the doubling time of the epidemic (**B**) and of R_0 (**C**). The bar charts show the results obtained from the first 11 sequences available on 2 May 2009 (orange) and from an updated analysis with 23 epidemiologically unlinked sequences available on 7 May 2009 (blue). The differences in estimates arise due to some sequences in the smaller sample being from epidemiological clusters, highlighting the importance of careful sampling.

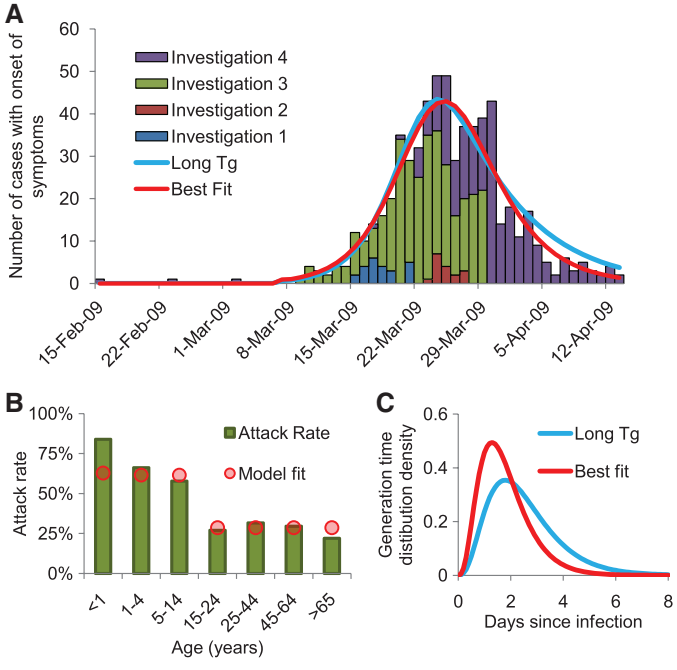

Fig. 3. Results of a detailed investigation into an outbreak in the village of La Gloria. (**A**) The time series of cases based on repeat rounds of investigation into the outbreak, and the best fit of an age-stratified transmission model (see Table 2 for estimates). The graph also shows the best fit of a model where the generation time is constrained to be consistent with earlier estimates for influenza (2.6 days), which does not fit significantly worse than the unconstrained best fit (see Table 2 legend). (**B**) Observed (bars) and fitted (using best fit, circles) age-specific attack rates; (**C**) best fit and constrained estimate of the generation time distribution.

Table 1. Parameter estimates (and 95% confidence intervals) for the cumulative number of influenza A (H1N1) infections in Mexico among Mexican residents by late April 2009, along with corresponding estimates for the case fatality ratio (CFR), the basic reproduction number, and the exponential growth rate, assuming that infections in travelers occurred in Mexico.

Assumed average duration of stay for visitors arriving by flights	Estimated number of infections among Mexican residents	Case fatality ratio*		Estimated reproduction number R_0‡	Estimated exponential growth rate‡ (doubling time in days‡)
		Assuming 9 deaths§	Assuming 101 deaths§		
Based on analysis of interval-censored country counts					
6 days	32,000 (26,000, 39,000)	0.03% (0.02%, 0.03%)	0.32% (0.26%, 0.39%)	1.42	0.123 (5.6)
9 days (*23*)	23,000 (20,000, 280,00)	0.04% (0.03%, 0.05%)	0.44% (0.37%, 0.52%)	1.40	0.118 (5.9)
12 days	180,00 (150,00, 220,00)	0.05% (0.04%, 0.06%)	0.55% (0.46%, 0.66%)	1.39	0.114 (6.1)
Based on analysis of presence or absence of confirmed disease in countries					
6 days	11,000 (5,000, 27,000)	0.08% (0.03%, 0.20%)	0.90% (0.37%, 2.20%)	1.36	0.106 (6.6)
9 days (*23*)	7,000 (3,000, 18,000)	0.12% (0.05%, 0.29%)	1.36% (0.56%, 3.30%)	1.33	0.098 (7.0)
12 days	6,000 (2,000, 14,000)	0.16% (0.07%, 0.39%)	1.81% (0.74%, 4.40%)	1.31	0.093 (7.4)

*This is a simple estimate of the CFR (*4*). The numbers of deaths and the data on cases in countries were both obtained as reported on 30 April. Although censoring is important to consider when estimating the CFR (due to the time interval between case report and death), in these data we have both the time between clinical onset and death within Mexico and the time between clinical onset and case confirmation in other countries to consider. Given uncertainty surrounding both of these distributions, we have made the simplifying assumption that these time intervals are similar and thus that the CFR can be estimated by the deaths within Mexico attributed to the novel influenza A (H1N1) divided by the estimated number of infections among Mexican residents based on data reported up to 30 April. §Based on 9 confirmed and 92 suspected deaths (101 total) that were reported by 30 April 2009 (*6*). ‡These estimates assume that the estimated cumulative number of influenza A (H1N1) infections in Mexico among Mexican residents relates to the cumulative number of infections up to 23 April 2009 (i.e., 1 week before the data were reported).

Table 2. Epidemiological parameters estimated by fitting an age-stratified mathematical model to the outbreak in the village of La Gloria (Fig. 3). For sensitivity analysis and model selection, we tested several reduced model variants. None fitted significantly worse, but several produced implausible estimates of the generation time. The respective best-fit values are as follows: 1. No asymptomatics and no misreporting, no assortative mixing: $p_{symp} \equiv 1$, $\theta \equiv 0$, $R_0 = 1.37$, $T_g = 1.39$, and $\rho_{child} = 2.80$. 2. No asymptomatics and no misreporting: $p_{symp} \equiv 1$, $R_0 = 1.41$, $T_g = 1.53$ days, $\theta = 0.31$, and $\rho_{child} = 2.22$. 3. No assortative mixing: same as variant 1. 4, model with long generation time consistent with previous estimates from influenza (also shown in Fig. 3), $T_g \equiv 2.60$ days, $R_0 = 1.97$, $\rho_{child} = 2.52$, $\theta = 0.51$, and $p_{symp} = 0.72$. (The symbol "$\equiv$" denotes parameters defined to take fixed values.)

	Best estimate	95% confidence interval	Description
R_0	1.58	1.34–2.04	Basic reproduction number
T_g	1.91	1.30–2.71	Mean generation time (days)
p_{symp}	86%	69–100%	Proportion of cases that are symptomatic and ascertained
ρ_{child}	2.06	1.60–3.31	Susceptibility of children relative to adults
θ	0.50	0.00–0.72	Assortativity of mixing between children and adults (0 = random, 1 = fully assortative)
	Assumed value		
f_L	1/3	Assumed	Fraction of the generation time that is latent (uninfectious)
ϕ_{child}	1	Assumed	Infectiousness of children relative to adults

obtained from analysis of previous pandemics [1.4 to 2.0 for 1918, 1957, and 1968 (*9*, *13*–*15*)].

Overall, our transmissibility estimates are consistent with the lowest values used in earlier detailed computer simulations used to study scenarios in pandemic mitigation (*16*, *17*), indicating that the conclusions regarding control policy effectiveness reached by those analyses could be relevant to the current epidemic. However, the key trade-off remains the balancing of the economic and societal cost of interventions, such as school closure, against the numbers of lives saved through use of such measures. Where substantial antiviral stockpiles are available, a secondary trade-off is the extent to which large-scale prophylaxis is justified, given the potential risks of high-level resistance developing (*18*–*21*). At present, estimates of disease severity are insufficiently robust to allow these trade-offs to be properly evaluated, but that uncertainty should diminish rapidly in coming weeks as more data on severe cases in the United States and other countries become available.

As the situation develops, a key issue is to optimize study designs and surveillance protocols to be most informative in estimating some of these unknown factors, thus potentially informing and refining the public health response. Clearly, detailed investigations of transmission in households and schools will be useful, as would be the consistent collection and dissemination of electronic patient records, which could be used to detect cofactors in the severity of infection.

In conclusion, while the emerging data from Mexico and other countries have enabled important insights into the origin, extent, transmissibility, and severity of the unfolding pandemic [including detailed epidemiological analysis of data from the U.S. outbreak recently published (*22*)], many uncertainties remain and should not be underestimated. The incubation and infectious periods have not yet been reliably ascertained, leaving uncertainty in estimates of the generation time. Much remains to be done to estimate clinical severity of infection, to understand regional variations seen so far (or indeed, whether they exist). As the epidemic spreads further, it is likely that severity will vary from country to country depending on health care resources and the public health measures adopted to mitigate impact. The existence of any cross-immunity (perhaps not mediated via HA-specific antibodies) from past exposure to prior influenza A subtypes is unknown, but the strong age dependence in clinical attack rates seen in La Gloria is intriguing. Cross-immunity would imply that R_0 could be higher in fully susceptible populations than estimated here.

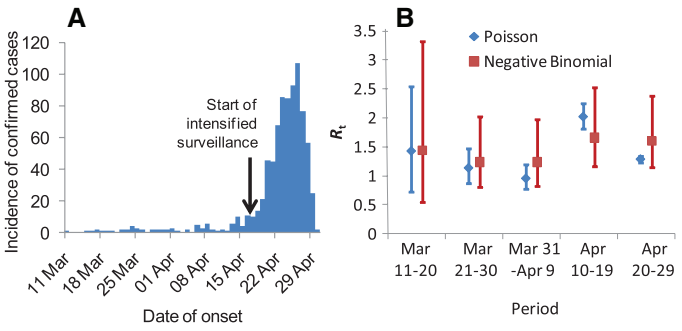

Fig. 4. (**A**) Time course of the Mexican epidemic with (**B**) the posterior estimates (median and 95% CrI) of the reproduction number over time obtained under Poisson and negative binomial models from the analysis of confirmed cases. The estimate of the negative binomial dispersion parameter k is for a low-to-moderate overdispersion, but this is enough to greatly increase the uncertainty in $R(t)$.

The future evolution of the transmissibility, antigenicity, virulence, and antiviral resistance profile of this or any influenza virus is difficult to predict. It is also unclear whether this strain will displace existing influenza A subtypes from the human population, as occurred in the past three pandemics. The extent to which seasonal damping of transmission in North America and Europe is responsible for the moderate transmissibility seen to date is uncertain; the progress of transmission in the Southern Hemisphere (which is just entering its influenza season) needs to be carefully monitored in the next few months. To reduce all these uncertainties, it is essential that public health agencies around the world continue to collect high-quality epidemiological data in a focused, resource-efficient manner despite the expected increases in case numbers in coming weeks. Epidemiological analysis and modeling are useful tools for guiding such efforts and interpreting the resulting data.

Note added in proof: We cited two sources (*1*, *6*) for confirmed and suspected deaths in Mexico, reported by 4 May 2009 and 30 April 2009, respectively. These sources are not publicly available at present. However, similar reports are publicly available: The Mexican government Web site (*24*) gives some data on the 5 May situation report (*25*) documenting 26 confirmed deaths and 114 suspected deaths (77 without samples for analysis), and *Morbidity and Mortality Weekly Report* (*26*) lists 7 confirmed and 77 suspected deaths posted on 30 April. Since this article appeared online, the number of deaths in Mexico up to 23 April has been determined to be 21, resulting in a revised estimate of the CFR of 0.091% (range: 0.066 to 0.35%) (*24*).

References and Notes

1. México Dirección General Adjunta de Epidemiología, *Brote de influenza humana A H1N1 México* (4 and 5 May 2009).
2. WHO, *Swine Influenza—Update 15* (www.who.int/csr/don/2009_05_05/en/index.html; accessed 5 May 2009).
3. W. P. Glezen, A. A. Payne, D. N. Snyder, T. D. Downs, *J. Infect. Dis.* **146**, 313 (1982).
4. A. C. Ghani *et al.*, *Am. J. Epidemiol.* **162**, 479 (2005).
5. T. D. Hollingsworth, N. M. Ferguson, R. M. Anderson, *Nat. Med.* **12**, 497 (2006).
6. Mexico Dirección General Adjunta de Epidemiología, *Brote de influenza porcina México* (30 April 2009).
7. A. J. Drummond, A. Rambaut, *BMC Evol. Biol.* **7**, 214 (2007).
8. Additional details on methods, data, and results are in the Supporting Online Material.
9. N. M. Ferguson *et al.*, *Nature* **437**, 209 (2005).
10. J. Wallinga, M. Lipsitch, *Proc. Biol. Sci.* **274**, 599 (2007).
11. M. Lipsitch *et al.*, *Science* **300**, 1966 (2003).
12. W. H. Frost, E. Sydenstricker, *Public Health Rep.* **34**, 491 (1919).
13. C. E. Mills, J. M. Robins, M. Lipsitch, *Nature* **432**, 904 (2004).
14. R. Gani *et al.*, *Emerg. Infect. Dis.* **11**, 1355 (2005).
15. C. Viboud *et al.*, *Vaccine* **24**, 6701 (2006).
16. N. M. Ferguson *et al.*, *Nature* **442**, 448 (2006).
17. T. C. Germann, K. Kadau, I. M. Longini Jr., C. A. Macken, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5935 (2006).
18. M. D. de Jong *et al.*, *N. Engl. J. Med.* **353**, 2667 (2005).
19. A. Lackenby *et al.*, *Euro Surveill.* **13** (2008); available at www.eurosurveillance.org/images/dynamic/EE/V13N05/art8026.pdf.
20. A. Moscona, *N. Engl. J. Med.* **353**, 2633 (2005).
21. S. H. Hauge, S. Dudman, K. Borgen, A. Lackenby, O. Hungnes, *Emerg. Infect. Dis.* **15**, 155 (2009).
22. Novel Swine-Origin Influenza A Virus Investigation Team, *N. Engl. J. Med.* 10.1056/NEJMoa0903810 (2009).
23. L. E. Hudman, R. H. Jackson, *Geography of Travel and Tourism* (Cengage Learning, Thomson Learning, Clifton Park, NY, ed. 4, 2002).
24. http://portal.salud.gob.mx/contenidos/noticias/influenza/estadisticas.html.
25. http://portal.salud.gob.mx/sites/salud/descargas/pdf/influenza/presentacion20090505.pdf.
26. www.cdc.gov/mmwr/preview/mmwrhtml/mm5817a5.htm.

27. We thank all those in Mexico and WHO (in particular J. Fitzner, K. Vandemaele and A. Merianos) who helped to collate the data used in this analysis. We also thank A. Borquez for help with translation and data collection and R. Eggo for data collation. We thank R. Anderson, K. Fukada, R. Hatchett, M. Lipsitch, D. Shay and L. Wolfson for useful discussions and comments. We thank the U.S. Centers for Disease Control; the Instituto de Salud Carlos III, Spain; Statens Erum Institut, Denmark; Erasmus MC Rotterdam, Netherlands; University of Regensburg, Germany; and the WHO collaborating centre for Reference and Research on Influenza, Australia, for posting viral sequences on GenBank. The work at Imperial College was funded by the Medical Research Council UK Centre grant. We also acknowledge additional support for individual staff members from the National Institute of General Medical Sciences (NIH) Models of Infectious Disease Agent Study (MIDAS) programme, The Royal Society (C.F., W.P.H., N.C.G., A.R., O.G.P.), Research Councils UK (S.C.), Bill and Melinda Gates Foundation (M.V.K., T.D.H., J.G.), The Wellcome Trust (R.F.B., grant GR082623MA), Biotechnology and Biological Sciences Research Council UK (T.J.), Microsoft Research (W.R.H.), and a studentship from the Medical Research Council (H.E.J.). GenBank accession numbers: GQ117067, FJ973557, FJ966082, FJ966952, FJ966960, FJ966974, FJ966971, FJ969511, GQ117040, FJ985753, GQ117119, FJ982430, GQ117097, GQ117059, GQ117103, GQ117112, CY039527, FJ984364, FJ984397, FJ985763, FJ974021, GQ117056, and FJ966982 for the main analysis and CY039527, FJ966082, FJ966959, FJ966960, FJ966974, FJ969509, FJ969511, FJ966952, FJ966982, FJ971076, and FJ973557 for the preliminary analysis.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1176062/DC1
Methods
Figs. S1 to S3
Tables S1 to S12
References
Epidemiological data

5 May 2009; accepted 11 May 2009
Published online 11 May 2009;
10.1126/science.1176062
Include this information when citing this paper.

Rapid and Accurate Large-Scale Coestimation of Sequence Alignments and Phylogenetic Trees

Kevin Liu,[1] Sindhu Raghavan,[1] Serita Nelesen,[1] C. Randal Linder,[2] Tandy Warnow[1]*

Inferring an accurate evolutionary tree of life requires high-quality alignments of molecular sequence data sets from large numbers of species. However, this task is often difficult, slow, and idiosyncratic, especially when the sequences are highly diverged or include high rates of insertions and deletions (collectively known as indels). We present SATé (simultaneous alignment and tree estimation), an automated method to quickly and accurately estimate both DNA alignments and trees with the maximum likelihood criterion. In our study, it improved tree and alignment accuracy compared to the best two-phase methods currently available for data sets of up to 1000 sequences, showing that coestimation can be both rapid and accurate in phylogenetic studies.

Phylogeny estimation from molecular sequences typically has two phases: An alignment is estimated, and then a tree is produced for the alignment. Alignment methods like MAFFT (*1*), Probcons (*2*), Probtree (*3*), Prank (*4*), and Muscle (*5*) provide more accurate alignments than earlier methods (*3*, *4*, *6*), and maximum likelihood (ML) methods of phylogeny estimation [e.g., RAxML (*7*, *8*), GARLI (*9*), and Phyml (*10*)] produce more accurate trees for large data sets than other