#### **Supplementary Text**:

#### Details of mechanistic model

 Each infected host (who developed symptoms) was assumed to progress through latent (*E*), presymptomatic infectious (*P*) and symptomatic infectious (*I*) stages. We denote the 5 infectiousness of hosts in the *P* and *I* stages by  $\beta_P$  and  $\beta_I$ , respectively, and define the ratio  $\alpha_P = \beta_P/\beta_I$ . The duration of each stage, denoted  $y_{E/P/I}$ , was assumed to be gamma distributed:

$$
y_E \sim \text{Gamma}(k_E, 1/(k_{inc}\gamma)),
$$

$$
y_P \sim \text{Gamma}(k_P, 1/(k_{inc}\gamma)),
$$

$$
y_{I} \sim \text{Gamma}(k_{I}, 1/(k_{I}\mu)),
$$

11 where we write  $X \sim \text{Gamma}(a, b)$  for a gamma distributed random variable with shape 12 parameter *a* and scale parameter *b*. We assumed that  $k_E + k_P = k_{inc}$ , so that the incubation

13 period,  $\tau_{inc} = y_E + y_P$ , was gamma distributed, with

$$
\tau_{inc} \sim \text{Gamma}(k_{inc}, 1/(k_{inc}\gamma)).
$$

15 Here, the parameters  $k_{inc}$  and  $1/\gamma$  (which represent the shape parameter of the incubation period distribution and the reciprocal of the mean incubation period, respectively) were fixed in order to obtain the specified incubation period distribution (the exact values that we assumed are given in Table S1).

20 For simplicity, we assumed that  $k<sub>I</sub> = 1$ , so the symptomatic infectious period was

21 exponentially distributed. The parameters  $k_E$  (the shape parameter of the latent  $(E)$  period

22 distribution),  $1/\mu$  (the mean symptomatic infectious (*I*) period) and  $\alpha_p$  (the ratio between the

transmission rates of hosts in the *P* and *I* stages) were estimated when we fitted the model to

the household transmission data.

 Hosts who remained asymptomatic throughout infection were assumed to follow the same *E*/*P*/*I* stages, although in this case the distinction between the *P* and *I* stages has no 28 epidemiological meaning. Stage durations, as well as the value of  $\alpha_p$ , were assumed to be identical for entirely asymptomatic hosts and those who developed symptoms, so that the generation time distribution was the same for all hosts.

31

#### 32 *Conditional infectiousness*

33 For a host who develops symptoms, conditional on incubation period  $\tau_{inc}$ , the expected

34 infectiousness at time since infection  $\tau$  is [1]

35 
$$
\beta(\tau | \tau_{inc}) = \begin{cases} \alpha_P C(\beta_0/n)(1 - F_{Beta}(1 - \tau/\tau_{inc}; k_P, k_E)), & 0 < \tau < \tau_{inc}, \\ C(\beta_0/n)(1 - F_I(\tau - \tau_{inc})), & \tau > \tau_{inc}. \end{cases}
$$

36 Here,  $\beta_0$  is the overall infectiousness parameter (see Methods in the main text), *n* is the 37 household size,  $F_I(y)$  is the cumulative distribution of the duration of the *I* stage, 38  $F_{Beta}(x; a, b)$  is the cumulative distribution of a beta distributed random variable with shape 39 parameters  $a$  and  $b$ , and

40 
$$
C = \frac{k_{inc}\gamma\mu}{\alpha_p k_p \mu + k_{inc}\gamma}.
$$

41

42 The cumulative conditional infectiousness can therefore be calculated to be

43 
$$
B(\tau | \tau_{inc}) = \int_0^{\tau} \beta(\tilde{\tau} | \tau_{inc}) d\tilde{\tau}
$$
  
\n44 
$$
= \begin{cases} (\tau - \tau_{inc})\beta(\tau | \tau_{inc}) + \frac{\alpha_p C \beta_0}{n} \left[ \frac{k_p \tau_{inc}}{k_{inc}} (1 - F_{Beta}(1 - \tau/\tau_{inc}; k_p + 1, k_E)) \right], & 0 \le \tau < \tau_{inc}, \\ (\tau - \tau_{inc})\beta(\tau | \tau_{inc}) + \frac{C \beta_0}{n} \left[ \frac{\alpha k_p \tau_{inc}}{k_{inc}} + \frac{1}{\mu} F_{Gamma} \left( \tau - \tau_{inc}; k_l + 1, \frac{1}{k_l \mu} \right) \right], & \tau \ge \tau_{inc}, \end{cases}
$$

45 where  $F_{Gamma}(x; a, b)$  is the cumulative distribution of a gamma distributed random variable 46 with shape parameter  $a$  and scale parameter  $b$ . The total force of infection on each household 47 member (over the course of infection) is then

48 
$$
B(\infty | \tau_{inc}) = \frac{\beta_0}{n} \left( \frac{\alpha k_P \gamma \mu \tau_{inc} + k_{inc} \gamma}{\alpha k_P \mu + k_{inc} \gamma} \right).
$$

49 The mean of this expression over the incubation period distribution is then  $\beta_0/n$ .

51 For hosts who remained asymptomatic throughout infection, conditional on the combined 52 duration of the *E* and *P* stages,  $\tau_{inc} = y_E + y_P$ , the infectiousness,  $\beta(\tau | \tau_{inc})$ , was given by 53  $\alpha_A$  times the corresponding expression for those who developed symptoms. We note that in 54 this case,  $\tau_{inc}$  has no epidemiological interpretation, but this conditional infectiousness was 55 useful when fitting parameters (see "Parameter fitting" below).

56

## 57 *Generation time distribution*

58 The generation time,  $\tau_{gen}$ , can be written as

$$
\tau_{gen} = y_E + y^*,
$$

60 where  $y_E$  is the length of the latent (*E*) stage, and  $y^*$  is the time from the start of the

61 presymptomatic infectious ( $P$ ) stage to the transmission occurring. As shown in [1],  $y^*$  has

62 density

63 
$$
f^*(y^*) = C\bigg(\alpha_P\big(1 - F_P(y^*)\big) + \int_0^{y^*} \big(1 - F_I(y^* - y_P)\big) f_P(y_P) \mathrm{d}y_P\bigg).
$$

64 Using this density, it can be shown that the moments of this distribution are

65 
$$
E[(y^*)^m] = \frac{C}{m+1} (\alpha_P E[y_P^{m+1}] + E[(y_P + y_I)^{m+1} - y_P^{m+1}]).
$$

66 In particular, we have

67 
$$
E[y^*] = \frac{C}{2} (\alpha_P E[y_P^2] + 2E[y_P]E[y_I] + E[y_I^2]),
$$

68 
$$
Var[y^*] = \frac{C}{3}(\alpha_P E[y_P^3] + 3E[y_P^2]E[y_I] + 3E[y_P]E[y_I^2] + E[y_I^2]) - (E[y^*])^2.
$$

69 Note that for a gamma distributed random variable,  $X \sim \text{Gamma}(a, b)$ , we have

70 
$$
E[X^m] = \frac{\Gamma(a+m)}{\Gamma(a)} b^m = a(a+1)...(a+(m-1))b^m.
$$

71 Therefore, for gamma distributed stage durations, explicit expressions can be obtained for the 72 mean and variance of the generation time distribution,

$$
E[\tau_{gen}] = E[y_E] + E[y^*],
$$

$$
Var[\tau_{gen}] = Var[y_E] + Var[y^*],
$$

75 since  $y_F$  and  $y^*$  are assumed to be independent.

76

## 77 *Proportion of presymptomatic transmissions*

78 Among infectors who develop symptoms, the proportion of transmissions occurring prior to 79 symptom onset is given by [1]

80 
$$
q_P = \frac{\left(\frac{\beta_P k_P}{k_{inc}\gamma}\right)}{\left(\frac{\beta_P k_P}{k_{inc}\gamma} + \frac{\beta_I}{\mu}\right)} = \frac{\alpha_P k_P \mu}{\alpha_P k_P \mu + k_{inc}\gamma}.
$$

81

82 *Parameter fitting*

83 The vector of model parameters,

84  $\theta = (k_F/k_{inc}, 1/\mu, \alpha_P, \beta_0),$ 

85 was estimated by fitting the mechanistic model to the household transmission data.

86

87 We assumed independent prior distributions for each entry of  $\theta$ . Lognormal priors were 88 assumed for  $1/\mu$ ,  $\alpha_p$  and  $\beta_0$ . Since  $\alpha_p$  represents the ratio between the transmission rate of hosts in the *P* and *I* stages, a prior with median 0 was used to ensure equal prior probabilities of values above and below 1. This prior was also chosen to limit the prior probability of extreme values, with a prior 95% credible interval of [0.2,5]. A beta prior was used for  $k_E/k_{inc}$  (which was constrained to lie between 0 and 1), and was chosen to restrict the prior

93 probability of values very close to either 0 or 1. The exact priors we used are given in Table 94 S3.

95

96 A slightly amended version of the parameter fitting algorithm described in the main text for 97 the independent transmission and symptoms model was used. In particular, we augmented the 98 observed data with:

99 i. The infection time,  $t_i$ , of each infected host.

100 ii. The time,  $t_{s,i}$ , at which each infected host transitioned from the *P* to *I* stage.

101 Note that for hosts who developed symptoms, the time of entry into the *I* stage corresponds to

102 the symptom onset time. The data were also augmented with this transition time for entirely

103 asymptomatic infected hosts because the conditional infectiousness,  $\beta(\tau | t_{s,j} - t_j)$ , was

104 relatively easy to calculate compared to  $\beta(\tau)$ .

105

#### 106 In each step of the chain, we carried out (in turn) one of the following:

107 1. Propose new values for the vector of model parameters,  $\theta$ , using a multivariate 108 normal proposal distribution (around the value of  $\theta$  in the previous step of the 109 chain; a correlation of 0.5 was used between the proposal distributions of 110  $k_E/k_{inc}$  and  $\alpha_P$ , and between those of  $1/\mu$  and  $\alpha_P$ ). Accept the proposed 111 **parameters,**  $\theta_{pron}$ **, with probability** 

112 
$$
\min\left(\frac{L(\theta_{prop};t)\pi(\theta_{prop})}{L(\theta_{old};t)\pi(\theta_{old})},1\right),\right)
$$

113 where  $\theta_{old}$  denotes the vector of parameter values from the previous step of 114 the chain, and where the augmented data, **t**, remain unchanged in this step. 115 2. Propose new values for the precise symptom onset times of each symptomatic 116 infected host, using independent uniform proposal distributions (within the

117 day of symptom of onset for each host). For each household,  $m$ , accept the 118 **Solution** 118 hoposed augmented data,  $t_{prop}^{(m)}$ , from that household with probability

119 
$$
\min\left(\frac{L^{(m)}\left(\theta;\mathbf{t}_{prop}^{(m)}\right)}{L^{(m)}\left(\theta;\mathbf{t}_{old}^{(m)}\right)},1\right),\right)
$$

120 where  $t_{old}^{(m)}$  denotes the corresponding augmented data from the previous step 121 of the chain, and where the model parameters,  $\theta$ , remain unchanged in this 122 step (i.e., proposed times are accepted/rejected independently for each 123 household, according to the likelihood contribution from that household). 124 3. Propose new values for the infection time of one randomly chosen infected 125 host in each household (either symptomatic or asymptomatic), using 126 independent normal proposal distributions (around the equivalent times in the 127 previous step of the chain). For each household,  $m$ , accept the proposed 128 augmented data,  $t_{prop}^{(m)}$ , from that household with probability

129 
$$
\min\left(\frac{L^{(m)}\left(\theta;\mathbf{t}_{prop}^{(m)}\right)}{L^{(m)}\left(\theta;\mathbf{t}_{old}^{(m)}\right)},1\right).
$$

130 4. Propose new values for both the infection time,  $t$ , and the time of the start of the *I* stage,  $t_s$ , holding  $t_s - t$  constant, for one randomly chosen asymptomatic 132 infected host in each household (in households where there was at least one), 133 using independent normal proposal distributions (around the equivalent times 134 in the previous step of the chain). For each household,  $m$ , accept the proposed 135 augmented data,  $t_{prop}^{(m)}$ , from that household with probability

136 
$$
\min\left(\frac{L^{(m)}\left(\theta;\mathbf{t}_{prop}^{(m)}\right)}{L^{(m)}\left(\theta;\mathbf{t}_{old}^{(m)}\right)},1\right).
$$

- Relationship between generation time, TOST and serial interval
- Here, we consider a randomly chosen infector-infectee pair (in which both the infector and

140 the infectee develop symptoms) within a large, well-mixed population. In that setting, the

- observed distribution of generation times is equal to the normalised infectiousness profile,
- which will not be true within a household (cf. Figure 1 and Figure S4). We define:

143 
$$
\tau_{inc,1} = (incubation period of the infector),
$$

144 
$$
\tau_{inc,2} = (incubation period of the infectee),
$$

145  $\tau_{gen} =$  (generation time),

146 
$$
x_{tost} = (time from onset of symptoms (of infector) to transmission (TOST))
$$

$$
x_{ser} = (serial interval),
$$

148 where we use  $\tau$  for time intervals relative to the time of infection and x for those relative to the time of symptom onset. We denote the probability density functions of these time periods

150 by  $f_{inc,1}$ ,  $f_{inc,2}$ ,  $f_{gen}$ ,  $f_{tost}$  and  $f_{ser}$ , respectively. Note that we have

$$
x_{\text{tost}} = \tau_{\text{gen}} - \tau_{\text{inc},1}
$$

- and
- 153  $x_{ser} = x_{tost} + \tau_{inc,2}$
- so that
- 155  $x_{ser} = \tau_{gen} + \tau_{inc,2} \tau_{inc,1}$ .
- 
- 157 In the independent transmission and symptoms model,  $\tau_{gen}$  and  $\tau_{inc,1}$  are assumed to be
- independent, and the incubation periods of the infector and infectee are assumed to be drawn
- 159 independently from the population incubation period distribution,  $f_{inc} = f_{inc,1} = f_{inc,2}$ .
- Therefore, the TOST distribution is given by the convolution

161 
$$
f_{\text{tost}}(x_{\text{tost}}) = \int_0^\infty f_{\text{gen}}(x_{\text{tost}} + \tau) f_{\text{inc}}(\tau) d\tau.
$$
 (1)

162 We also assume that  $x_{\text{test}}$  and  $\tau_{\text{inc,2}}$  are independent, so the serial interval distribution can be 163 calculated from the TOST distribution as

164 
$$
f_{ser}(x_{ser}) = \int_0^\infty f_{tost}(x_{ser} - \tau) f_{inc}(\tau) d\tau.
$$
 (2)

165 Note that

166 
$$
E[x_{ser}] = E[\tau_{gen}] + E[\tau_{inc,2}] - E[\tau_{inc,1}] = E[\tau_{gen}],
$$

167 i.e., the generation time and serial interval distributions have the same mean.

168

169 For the mechanistic model, we still have  $f_{inc,2} = f_{inc}$ , and the serial interval distribution can 170 be calculated from the TOST distribution using equation (2). On the other hand,  $\tau_{gen}$  and  $\tau_{inc,1}$  are not independent, so equation (1) connecting the TOST and generation time distributions for the independent transmission and symptoms model does not hold for the mechanistic model. As shown in [1], the TOST distribution for the mechanistic model is, instead, given by

175 
$$
f_{tost}(x_{tost}) = \begin{cases} \alpha_p C (1 - F_p(-x_{tost})), & x_{tost} < 0, \\ C (1 - F_l(x_{tost})), & x_{tost} \ge 0. \end{cases}
$$

176

 Further, under the mechanistic model, the expected number of presymptomatic transmissions generated by an infected host is dependent on their incubation period. As a result, the infector's incubation period does not follow the same distribution as that of the infectee. In particular, by Bayes' theorem, we have

181 
$$
f_{inc,1}(\tau_{inc,1}) = p(\tau_{inc,1} \mid 1 \to 2) = \frac{p(1 \to 2 \mid \tau_{inc,1})f_{inc}(\tau_{inc,1})}{p(1 \to 2)},
$$

182 where we write  $1 \rightarrow 2$  to denote the occurrence of the transmission from the infector to the 183 infectee. Because we are here considering a large population, the probability of the

184 transmission occurring is proportional to the overall infectiousness of the infector (integrated 185 over the course of infection),  $B(\infty)$ , so we have

186 
$$
f_{inc,1}(\tau_{inc,1}) = \frac{B(\infty \mid \tau_{inc})f_{inc}(\tau_{inc,1})}{B(\infty)} = \left(\frac{\alpha_P k_P \gamma \mu \tau_{inc,1} + k_{inc} \gamma}{\alpha_P k_P \mu + k_{inc} \gamma}\right) f_{inc}(\tau_{inc,1}).
$$

187 The expected incubation period of the infector is then

188 
$$
E[\tau_{inc,1}] = \frac{1}{\gamma} + \frac{\alpha_P k_P \mu}{k_{inc} \gamma (\alpha_P k_P \mu + k_{inc} \gamma)} = E[\tau_{inc,2}] + \frac{q_P}{k_{inc} \gamma'}
$$

189 where  $q_p$  is the proportion of transmissions occurring prior to symptom onset.

190

191 As a result of the above, the expected values of the generation time and serial interval in the

192 mechanistic model are not equal. Instead, we have

$$
E[x_{ser}] = E\big[\tau_{gen}\big] - \frac{q_P}{k_{inc}\gamma}.
$$

194 Under the values of  $k_{inc}$  and  $\gamma$  that we assumed (Table S1), this gives a mean generation time

195 that is approximately  $(1.6 \times q_P)$  days longer than the mean serial interval.

## **Supplementary Figures:**





**Figure S1. Posterior distributions for the independent transmission and symptoms model.** Prior

distributions (black dotted lines), posterior distributions (blue bars), and posterior means (vertical red dotted

201 lines) for fitted parameters in the independent transmission and symptoms model. A. Mean generation time. B.

202 Standard deviation of generation times. C. Overall infectiousness,  $\beta_0$ .



 **Figure S2. Posterior distributions for the mechanistic model.** Prior distributions (black dotted lines), posterior distributions (blue bars), and posterior means (vertical red dotted lines) for fitted parameters in the 207 mechanistic model. A. Ratio of mean durations of the latent (E) and incubation (E+P) periods,  $k_E/k_{inc}$ . B. 208 Mean symptomatic infectious (I) period,  $1/\mu$ . C. Ratio of transmission rates in the P and I stages,  $\alpha_p$ . D. Overall 209 infectiousness,  $\beta_0$ .









 **Figure S4. Generation times within study households.** Violin plots indicating posterior distributions of the 222 mean (A) and standard deviation (B) of realised generation times in the study households, and the proportion of 223 transmissions occurring prior to symptom onset (C), for the independent transmission and symptoms model (blue) and mechanistic model (red).



**Figure S5. Number of study households and household members by recruitment month.** A. Bars indicating

228 the number of households recruited each month into the study from which we obtained the household

transmission data used in our analyses, from March to November 2020. B. Equivalent panel showing the total

number of individuals within the households recruited each month.



**Figure S6. Temporal changes in the generation time distribution for the mechanistic model.** Violin plots

- 234 indicating posterior distributions of the mean generation time (A), standard deviation of generation times (B),
- 235 and overall infectiousness parameter,  $\beta_0$  (C), for the mechanistic model fitted to data from March-April (blue),
- May-August (red) or September-November 2020 (orange).





**Figure S7. Monthly changes in the generation time distribution from September-November 2020.** Violin

240 plots indicating posterior distributions of the mean generation time (A), standard deviation of generation times

- 241 (B), and overall infectiousness parameter,  $\beta_0$  (C), for the independent transmission and symptoms model fitted
- to data from September (blue), October (red) or November 2020 (orange).





**Figure S8**. **Sensitivity of results to the functional forms of the generation time and incubation period** 

**distributions for the independent transmission and symptoms model.** A-C. Violin plots indicating posterior

247 distributions of the mean generation time (A), standard deviation of generation times (B), and overall

248 infectiousness parameter,  $\beta_0$  (C), for the independent transmission and symptoms model, when the generation

time was assumed to follow either a lognormal (blue, as in the main text) or a gamma (red) distribution, and the

incubation period distribution followed a lognormal distribution (as in the main text). D-F. Equivalent panels to

A-C, instead comparing cases where the incubation period was assumed to follow either a lognormal

distribution (blue, as in the main text) or a gamma distribution with the same mean and standard deviation (red),

and the generation time followed a lognormal distribution (as in the main text).





 **Figure S9. Sensitivity of the results to the incubation period distribution.** A-B. Uncertainty in the incubation 257 period distribution was accounted for by updating the parameters,  $\mu_{inc}$  and  $\sigma_{inc}$ , of a lognormal incubation period distribution (these parameters represent the mean and standard deviation of the natural logarithm of the incubation period distribution, respectively) alongside unknown model parameters in the MCMC procedure. Independent normal prior distributions (truncated at zero) consistent with the 95% confidence intervals obtained 261 in [2] were assumed for  $\mu_{inc}$  (prior mean 1.63 log(day), standard deviation 0.061, 95% CrI 1.51-1.75; panel A) 262 and  $\sigma_{inc}$  (prior mean 0.5 log(day), standard deviation 0.026, 95% CrI 0.45-0.55; panel B). This incubation period was used directly when evaluating the likelihood in the independent transmission and symptoms model. In the mechanistic model, we assumed a gamma distributed incubation period with the same mean and standard 265 deviation as a lognormal distribution with parameters  $\mu_{inc}$  and  $\sigma_{inc}$ . C-E. Violin plots indicating posterior 266 distributions of the mean generation time (C), standard deviation of generation times (D), and overall 267 infectiousness parameter,  $\beta_0$  (E), for the independent transmission and symptoms model with either a fixed (as 268 in the main text; blue) or variable (i.e., accounting for uncertainty in  $\mu_{inc}$  and  $\sigma_{inc}$  as described above; red) incubation period distribution, and for the mechanistic model with a fixed (orange) or variable (purple) incubation period distribution.









**Figure S11. Sensitivity of the results to the relative infectiousness of entirely asymptomatic infected hosts**.

Violin plots indicating posterior distributions of the mean generation time (A), standard deviation of generation

286 times (B), and overall infectiousness parameter,  $\beta_0$  (C), for the independent transmission and symptoms model,

when the relative infectiousness of asymptomatic hosts (compared to a host who develops symptoms, at the

288 same time since infection) was  $\alpha_A = 0.1$  (blue),  $\alpha_A = 0.35$  (red) and  $\alpha_A = 1.27$  (orange).





**Figure S12. Sensitivity of the results to the exclusion of hosts with unknown infection status.** A-C. Violin

292 plots indicating the posterior distributions of the mean generation time (A), standard deviation of generation

- 293 times (B), and overall infectiousness parameter,  $\beta_0$  (C), for the independent transmission and symptoms model,
- 294 when hosts with unknown infection status were assumed all uninfected (blue), excluded (red), or assumed all
- infected (orange).

# 297 **Supplementary Tables:**



298

299 **Table S1.** Assumed (not fitted) parameter values used for the two models that we considered.



- 301 **Table S2.** Fitted parameters in the independent transmission and symptoms model, the prior distributions used,
- 302 and the posterior means and 95% credible intervals obtained.



305 **Table S3.** Fitted parameters in the mechanistic model, the prior distributions used, and the posterior means and

306 95% credible intervals obtained.

## **References:**

- [1] Hart WS, Maini PK, Thompson RN. High infectiousness immediately before COVID- 19 symptom onset highlights the importance of continued contact tracing. *eLife* 2021; 311 10: e65534.
- [2] McAloon C, Collins Á, Hunt K, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open* 2020; 10: e039652.
- [3] Buitrago-Garcia D, Egli-Gany D, Counotte MJ, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLOS Med* 2020; 17: e1003346.