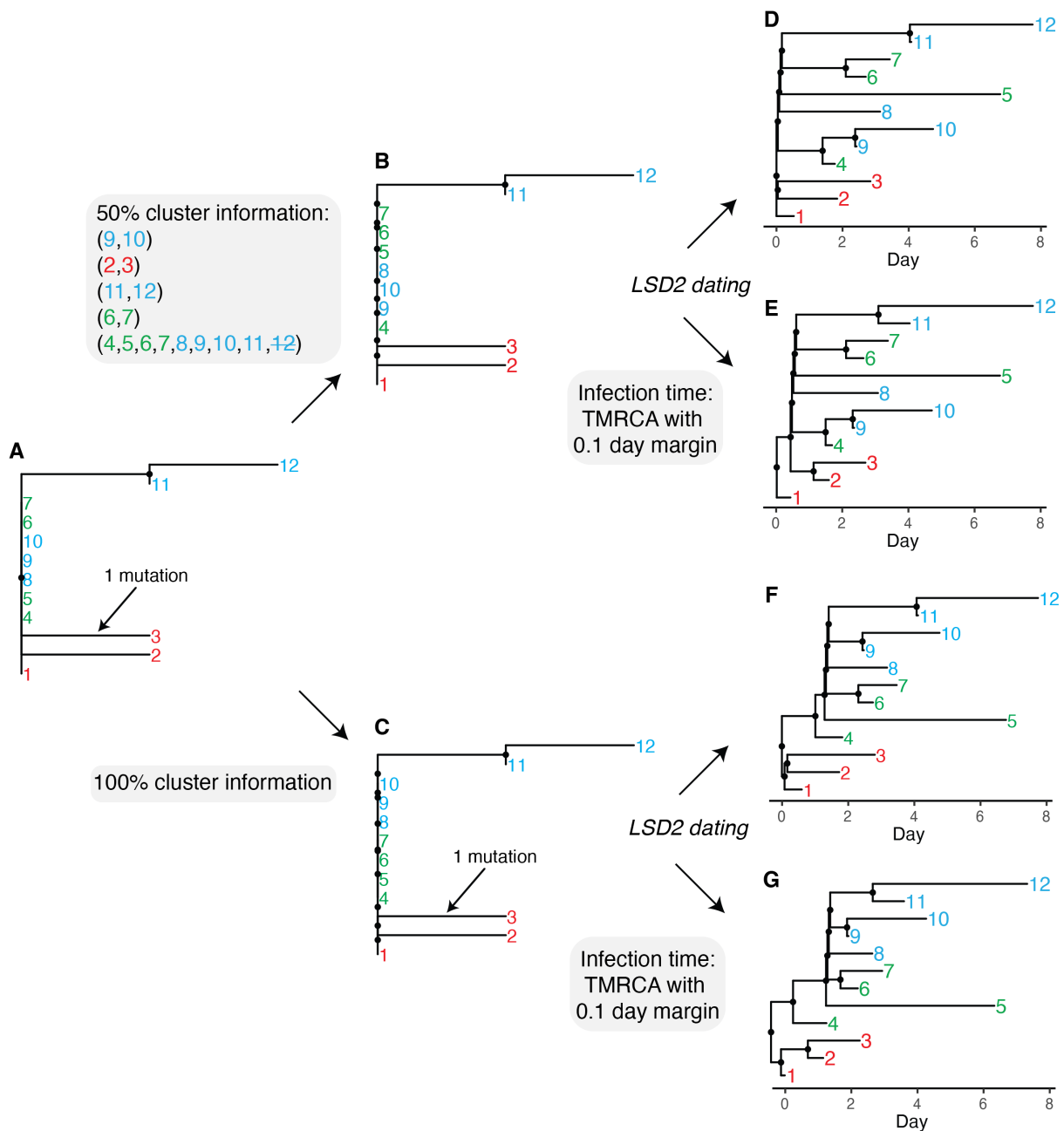
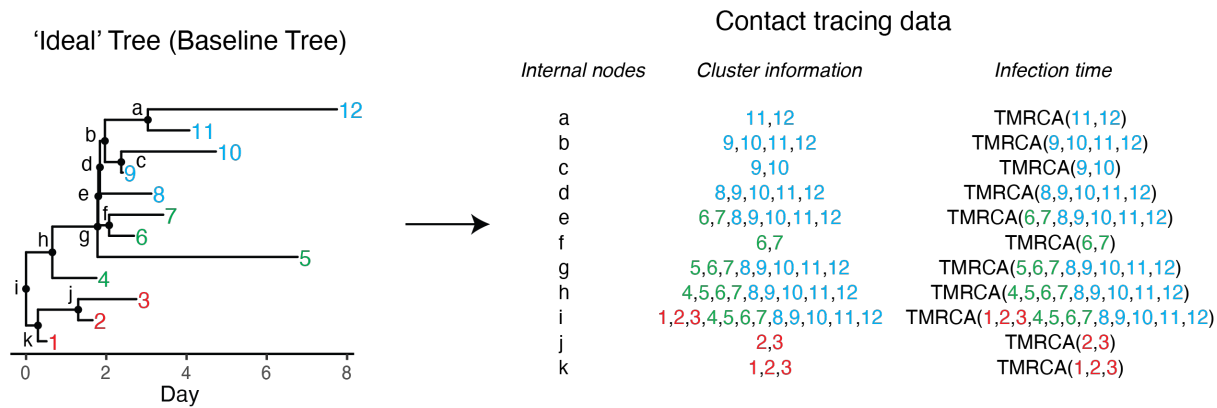


Supplementary Figure 1 | A comparison of the distribution of internal node children between 2000 simulated trees and trees generated from SARS-CoV-2 sequences during waves 3 and 4 in Hong Kong. (A) Contrasts children per internal node in simulated trees (left) with those observed in SARS-CoV-2 trees (right). (B) Depicts internal node offspring frequency across trees, with each line representing one simulated tree. 3rd and 4th wave occurrences are highlighted in red.

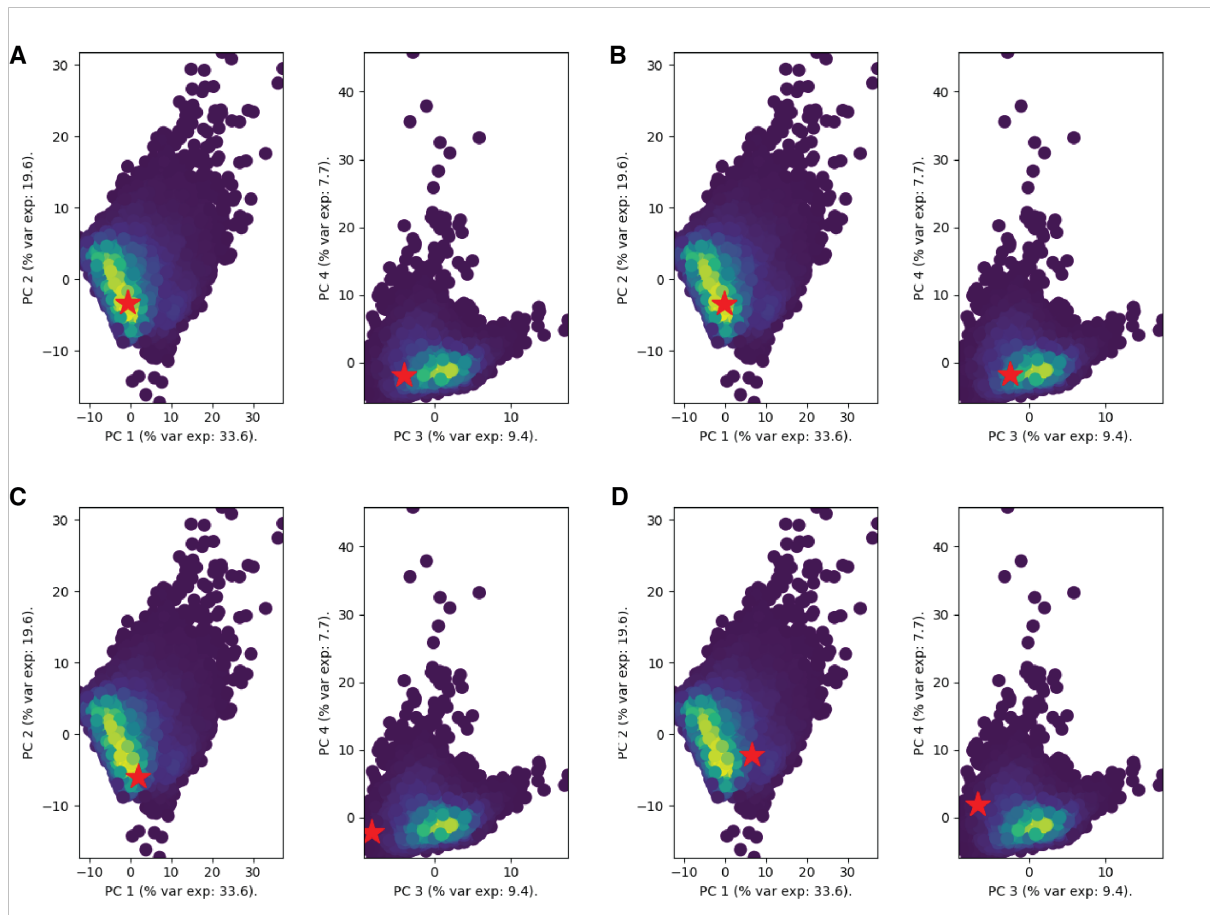


Supplementary Figure 2 | Resolving polytomies in the Genetic Tree using contact tracing data. (A) The initial Genetic Polytomous Tree, as in Fig. 1. (B) Tree resolved with 50% cluster information; notably, the inclusion of sample '12' in the fifth cluster does not alter the resolution outcome per our methodology. Polytomies lacking cluster information were randomly resolved into a binary tree. (C) Fully resolved tree with 100% cluster information, matching the Genetic Baseline Tree in Fig. 1. (D, E) Dated trees from tree in Panel (B) using LSD2, without and with temporal constraints from known 50% clusters, respectively. (F) Tree from Panel (C) dated using LSD2, corresponding to the Dated Baseline Tree in Fig. 1. (G)

Tree from Panel (C) dated using LSD2 with temporal constraints applied to all clusters, closely resembling the “ideal” tree in **Fig. 1**.



Supplementary Figure 3 | Extraction of contact tracing data from an "ideal" tree.



Supplementary Figure 4 | PCA plots illustrating the third and fourth waves of SARS-CoV-2 in Hong Kong. For the third wave: Dated Resolved (A) and Dated Resolved-Cluster (B). For the fourth wave: Dated Resolved (C) and Dated Resolved-Cluster (D).

Supplementary Table 1 | Parameter ranges used for simulations.

Parameters	Range in PhyloDeep	Range in this study	Relation to parameters in <i>tre simulator</i>
R_0	(1, 5)	(0.4, 6)*	$= (l_{a_ss} + l_{a_nn}) / \psi$
$1/\gamma$	(1, 10)	(1, 14)	$= 1 / \psi$
X_{ss}	(3, 10)	(2, 40)	$= l_{a_ss} / l_{s_ns} = l_{a_sn} / l_{a_nn}$
f_{ss}	(0.05, 0.20)	(0.05, 0.25)	$= l_{a_ss} / (l_{a_ss} + l_{a_sn}) = 1 - l_{a_nn} / (l_{a_nn} + l_{a_ns})$
k	(0.33, 0.90)	(0.09, 1.18)	-
<i>Tips</i>	(200, 500)	(200, 1000)	<i>min_tips, max_tips</i>
<i>Sampling proportion</i>	(0.01, 1)	(0.01, 0.4)	p

*Values highlighted in bold indicate the adjusted boundary values.

Supplementary Table 2 | Performance comparison (mean relative error) of models trained on seven types of phylogenetic trees.

Model	R_0	$1/\gamma$	X_{ss}	f_{ss}
Baseline-Model	0.095 (0.382)*	0.092 (0.561)	0.215 (0.721)	0.167 (0.598)
Dated Baseline-Model	0.103 (0.306)	0.122 (0.566)	0.268 (0.648)	0.200 (0.465)
Dated Resolved-Model	0.119	0.155	0.444	0.367
Dated Polytomous-Model	0.136	0.196	0.449	0.357
Genetic Baseline-Model	0.205 (0.664)	0.191 (0.523)	0.314 (0.798)	0.265 (0.602)
Genetic Resolved-Model	0.261	0.214	0.480	0.360
Genetic Polytomous-Model	0.232	0.192	0.454	0.361

**Values in parentheses represent performance metrics tested with realistic phylogenetic trees: the Baseline-Model and Dated Baseline-Model analyses utilized Dated Resolved trees, whereas the Genetic Baseline-Model employed Genetic Resolved trees for evaluation.*

Supplementary Table 3 | Performance comparison (mean relative error) by incorporating varying levels of contact tracing data based on Baseline-Model and Dated Baseline-Model.

Model	Percentage of contact tracing data	R_0	$1/\gamma$	X_{ss}	f_{ss}
Baseline-Model	0%	0.382	0.561	0.721	0.598
	25%	0.198	0.193	0.667	0.544
	50%	0.154	0.165	0.524	0.467
	75%	0.154	0.178	0.395	0.370
	100%	0.165	0.196	0.302	0.308
			(0.101)*	(0.109)	(0.222)
Dated Baseline-Model	0%	0.306	0.566	0.648	0.465
	25%	0.156	0.200	0.485	0.364
	50%	0.124	0.150	0.392	0.285
	75%	0.111	0.132	0.323	0.229
	100%	0.103	0.120	0.266	0.187

*Values in parentheses represent performance metrics tested with a time constraint margin of 0.1 day.

Supplementary Table 4 | Performance comparison (mean relative error) by incorporating varying levels of contact tracing data based on Baseline-Model, Dated Baseline-Model and BEAST2 using 100 simulated trees.

Model	Percentage of contact tracing data	R_0	$1/\gamma$	X_{ss}	f_{ss}
Baseline-Model	0%	0.419	0.595	0.675	0.596
	50%	0.142	0.162	0.469	0.459
	100%	0.140	0.178	0.283	0.266
		(0.083)*	(0.086)	(0.211)	(0.150)
BEAST2 (incorporating cluster information and infection times)	0%	0.384	0.295	1.086	0.728
	50%	0.271	0.282	0.682	0.723
	100%	0.236	0.311	0.492	0.715
		(0.187)	(0.223)	(0.451)	(0.684)
Dated Baseline-Model	0%	0.336	0.596	0.658	0.464
	50%	0.137	0.171	0.364	0.248
	100%	0.105	0.133	0.277	0.206
BEAST2 (incorporating cluster information)	0%	0.384	0.295	1.086	0.728
	50%	0.38	0.269	1.089	0.685
	100%	0.349	0.271	1.051	0.676

*Values in parentheses represent performance metrics tested with a time constraint margin of 0.1 day.

Supplementary Table 5 | Summary of out-of-range SS for waves 3 and 4 compared to simulated values in Dated Baseline-Model.

Features	Max	Min	<i>Dated Resolved</i> (wave 3)	<i>Dated Resolved-Cluster</i> (wave 3)	<i>Dated Resolved</i> (wave 4)	<i>Dated Resolved-Cluster</i> (wave 4)
Mean length of transmission chain	2.979	0.548	0.643	0.676	0.247*	0.452
The 4 th deciles of transmission chain	2.155	0.019	0.050	0.053	0.015	0.029
The 5 th deciles of transmission chain	2.552	0.028	0.050	0.053	0.015	0.029
The 6 th deciles of transmission chain	2.948	0.043	0.050	0.288	0.015	0.177
The 7 th deciles of transmission chain	3.778	0.065	0.156	0.559	0.015	0.456
The 8 th deciles of transmission chain	4.789	0.102	0.651	0.883	0.016	0.621
Max length of transmission chain	29.365	2.146	5.372	4.453	2.115	3.270

**Values highlighted in bold exceed the established maximum and minimum range thresholds.*

Supplementary Table 6 | Performance impact of training dataset size on Dated Resolved models.

Training dataset size	R_0	$1/\gamma$	X_{ss}	f_{ss}
200,000 trees	0.126	0.162	0.456	0.364
20,000 trees	0.133	0.175	0.479	0.394
5,000 trees	0.149	0.195	0.533	0.451
1,000 trees	0.186	0.218	0.762	0.587

Supplementary Table 7 | Summary of out-of-range SS for waves 3 and 4 compared to simulated values in PhyloDeep.

Features	Max	Min	<i>Dated Resolved (wave 3)</i>	<i>Dated Resolved (wave 4)</i>
Median of all branch length	0.939	0.520	0.445*	0.402
Variance of all branch length	1.636	0.590	1.737	1.886
Variance of external branch length	2.095	0.363	2.288	2.622
Mean length of transmission chain	2.481	0.929	0.643	0.247
The 2 nd deciles of transmission chain	1.382	0.186	0.050	0.015
The 3 rd deciles of transmission chain	1.663	0.298	0.050	0.015
The 4 th deciles of transmission chain	1.867	0.411	0.050	0.015
The 5 th deciles of transmission chain	2.063	0.522	0.050	0.015
The 6 th deciles of transmission chain	2.257	0.627	0.050	0.015
The 7 th deciles of transmission chain	2.672	0.775	0.156	0.015
The 8 th deciles of transmission chain	3.178	0.862	0.651	0.016
The 9 th deciles of transmission chain	3.635	1.214	1.257	0.345
The 10 th deciles of transmission chain	5.241	1.419	2.071	0.859

Max length of transmission chain	15.296	2.189	5.372	2.115
----------------------------------	--------	-------	-------	--------------

**Values highlighted in bold exceed the established maximum and minimum range thresholds.*

Supplementary Table 8 | Estimation of the initial R_0 during the exponential growth period of SARS-CoV-2 waves 3 and 4 in Hong Kong using line-listed data.

Waves	R_0	Generation Time	Parameters	Reference
3	1.880 (1.825, 1.937)	Exponential	Mean = 7.27; SD= 3.81	Chen et al. 2022
	1.693 (1.649, 1.738)	Exponential	Mean = 5.7; SD= 1.8	Hu et al. 2021
	2.003 (1.930, 2.079)	Gamma	Mean = 7.27; SD= 3.81	Chen et al. 2022
	1.800 (1.741, 1.861)	Gamma	Mean = 5.7; SD= 1.8	Hu et al. 2021
4	2.183 (2.089, 2.282)	Exponential	Mean = 7.27; SD= 3.81	Chen et al. 2022
	1.933 (1.858, 2.012)	Exponential	Mean = 5.7; SD= 1.8	Hu et al. 2021
	2.424 (2.289, 2.570)	Gamma	Mean = 7.27; SD= 3.81	Chen et al. 2022
	2.141 (2.031, 2.260)	Gamma	Mean = 5.7; SD= 1.8	Hu et al. 2021

Values in parentheses are the 95% CI. Estimates in bold are also shown in Table 1.

Supplementary Table 9 | Inference of epidemiological parameters based on waves 3 and 4 of SARS-CoV-2 in Hong Kong using BEAST2.

Waves	Input tree	R_0	Infection-to-sampling period (day)	X_{ss}	f_{ss}	Dispersion k
Wave3	Dated Resolved	2.142 (2.073, 2.281)	4.373 (3.975, 4.852)	31.225 (27.950, 35.002)	0.035 (0.031, 0.040)	0.358 (0.326, 0.392)
	Dated Resolved- Cluster	2.148 (2.096, 2.281)	3.935 (3.634, 4.335)	32.107 (28.950, 35.433)	0.035 (0.032, 0.039)	0.326 (0.297, 0.358)
Wave4	Dated Resolved	2.159 (2.075, 2.314)	9.912 (9.604, 9.998)	34.673 (31.345, 38.219)	0.038 (0.035, 0.044)	0.268 (0.242, 0.296)
	Dated Resolved- Cluster	2.154 (2.040, 2.265)	5.699 (5.085, 6.451)	34.515 (30.911, 38.044)	0.038 (0.035, 0.042)	0.270 (0.243, 0.298)

Note: in the BDSS model, the term “infectious period” specifically refers to the interval from the time of infection to the sampling date. To clarify terminology and avoid confusion in epidemiological contexts, we have replaced the term “infectious period” with “infection-to-sampling period” here and also in Table 1 of the revised manuscript.