

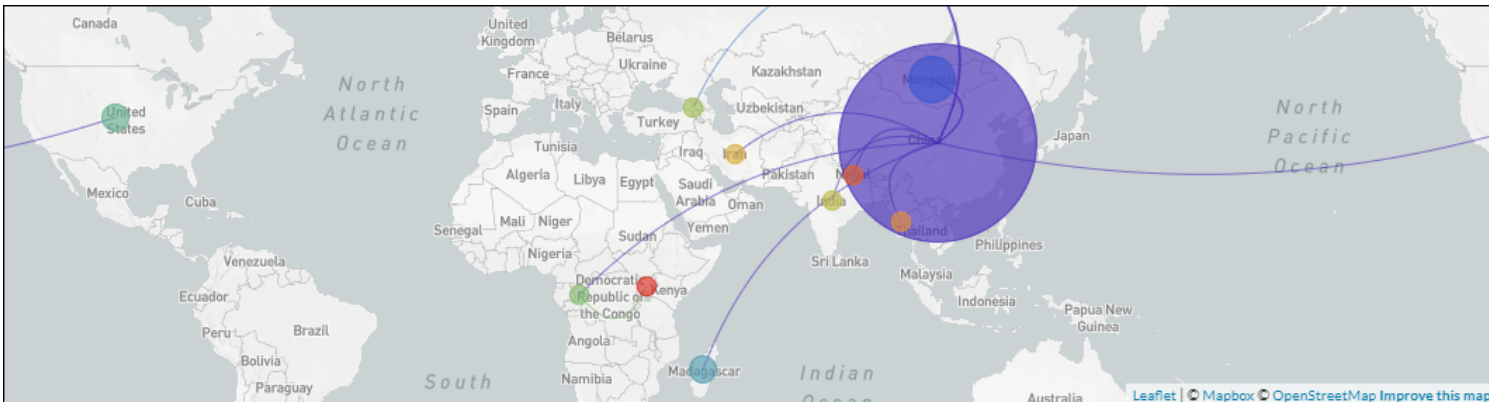
1000 PLAGUES IN THE GENOMICS AGE

New insights from the global phylogeny of *Yersinia pestis*

Katherine Eaton, Hendrik Poinar, Sebastian Duchene, Leo Featherstone

Poinar Lab Meeting

08 June 2021



MOLECULAR CLOCK MODELS

- “When did a past evolutionary event happen?”
- A molecular clock model estimates the rate at which DNA sequences change.
- Rate estimate + Phylogeny = Node Dating

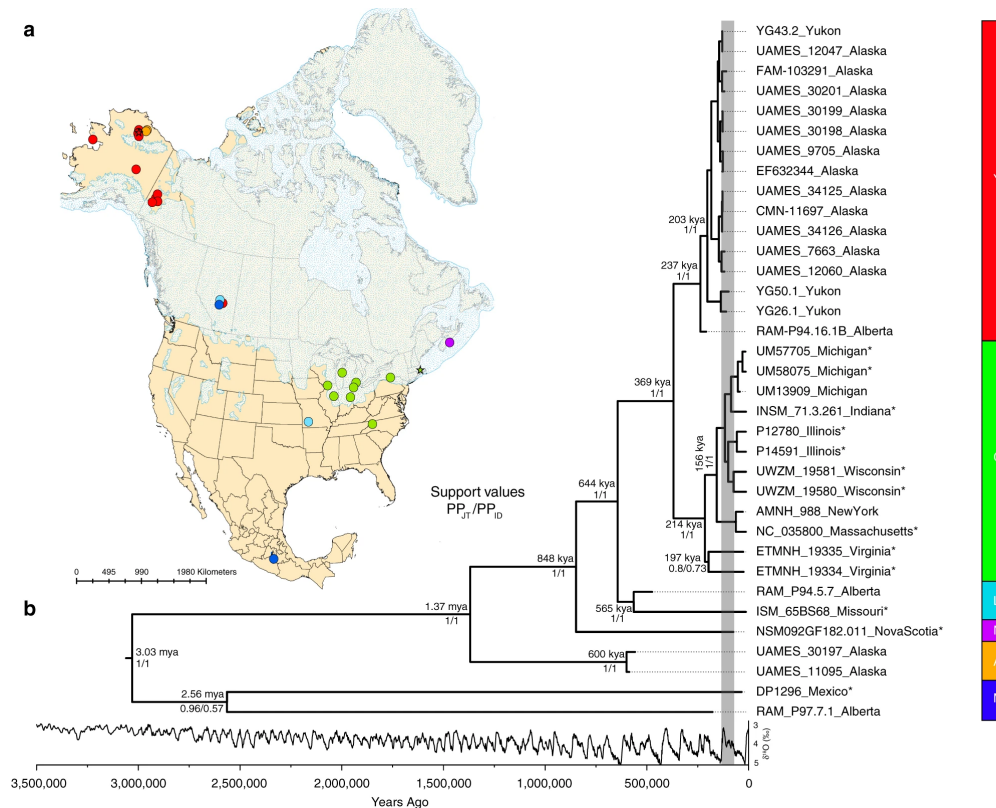


Figure 1: Phylogeographic relationships of American mastodons (Karpinski et al., 2020)

DOES PLAGUE HAVE TEMPORAL SIGNAL?

It's highly debated whether plague (*Yersinia pestis*) has detectable temporal signal.

"It's complicated." - Cui et al. (2013)

"Hard No." - Wagner et al. (2014)

"Sometimes?" - (Duchêne et al., 2016)

"Maybe the Second Pandemic does?" - Spyrou et al. (2019)

LET'S REVIVE THE DEBATE!

Data Collection

1. Downloaded metadata for 1657 *Y. pestis* genome sequencing projects.
 2. Reviewed the original publications to curate collection location, date and host.
 3. Downloaded the curated sequence data for 600 genomes.
-

Phylogeny

1. Constructed a multiple sequence alignment of the *Y. pestis* chromosome (~4 MB).
2. Estimated a maximum likelihood phylogeny.
3. Test for temporal signal, using a variety of methods.

Sebastian Duchene

Leo Featherstone

GLOBAL PHYLOGENY

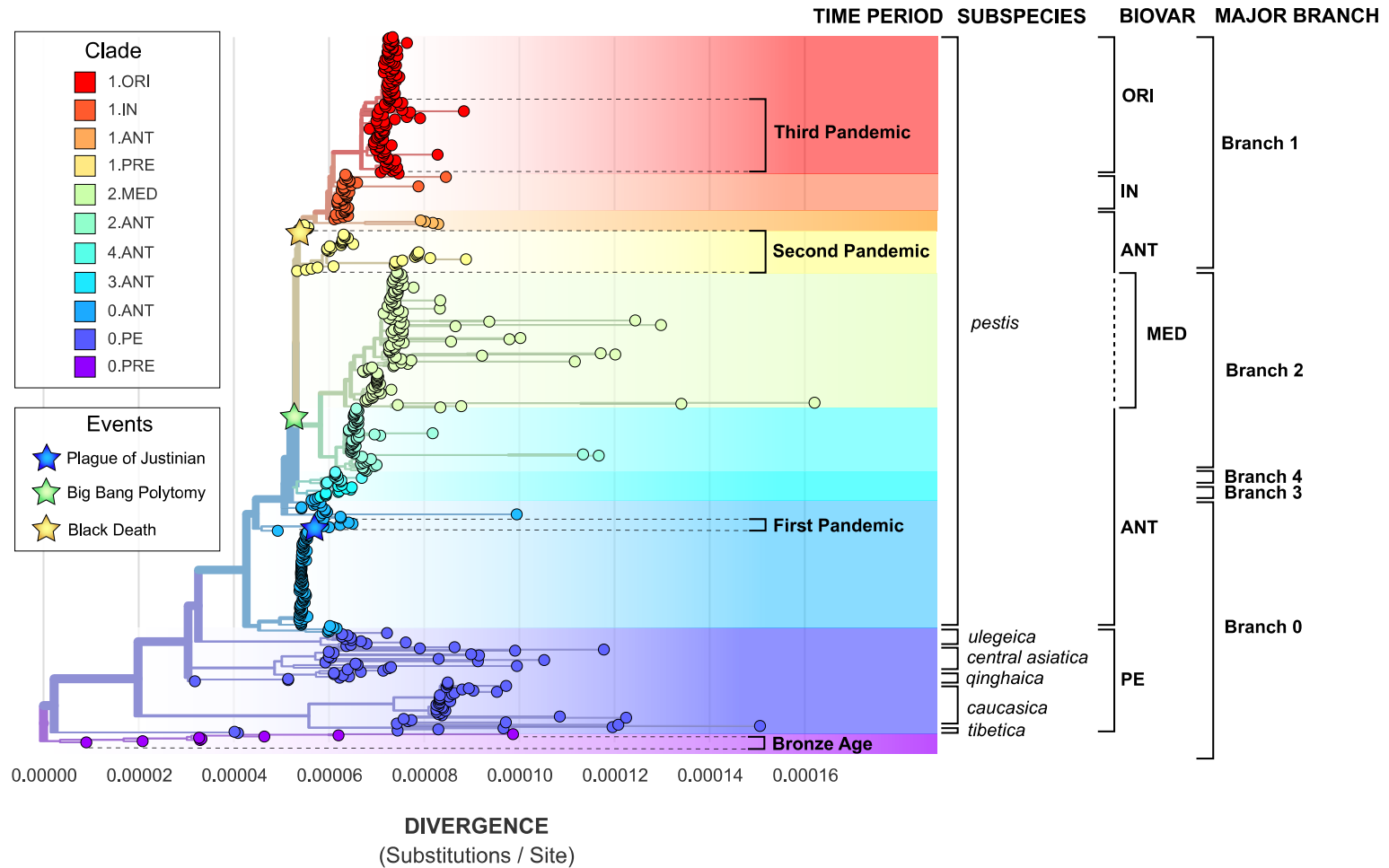


Figure 2: The maximum-likelihood phylogeny depicts the global population structure of *Y. pestis*.

PHYLODYNAMICS

What's the best method to estimate a molecular clock model?

1. Sequential (Dellicour et al., 2020)

- First estimate a phylogeny then the clock model.
- IQTREE, TreeTime, LSD2

Pros	Cons
Handles large datasets	Dependent on phylogeny being 'correct'
Super fast	Struggles with relaxed clock models

2. Joint (Duchene et al., 2020)

- Estimate a phylogeny and clock model simultaneously.
- BEAST, specifically, BETS.

Pros	Cons
Wide variety of models available	Can't handle large datasets
Not dependent on input tree	Slow

SEQUENTIAL (N=600)

- All terminal branches are long.
- Large number of polytomies.
- Divergences are older than expected.



Figure 3: The time-scaled phylogeny of the modern pandemic clade 1.ORI.

SEQUENTIAL (N=191)

- The MCMC analysis does not converge when estimating parameters of the clock model.
- The MCMC trace shows poor mixing throughout the run.

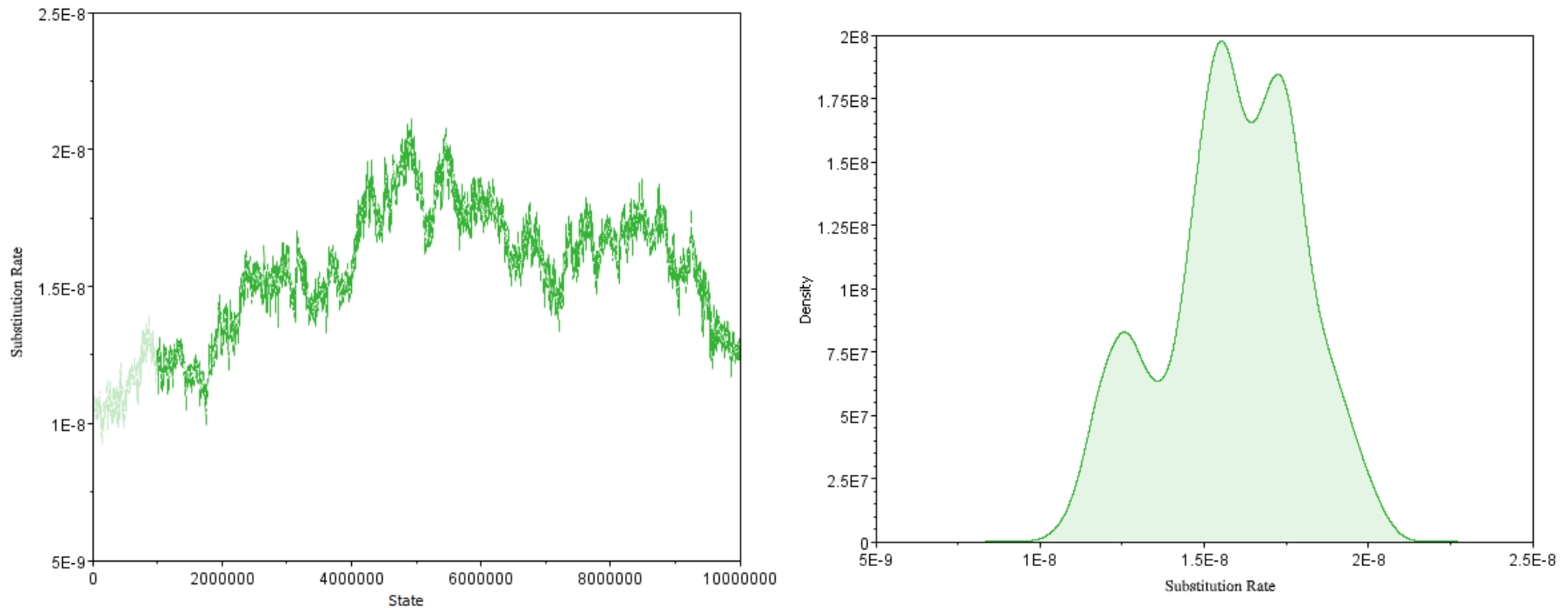


Figure 4: MCMC parameter estimation of the mean substitution rate. Left: Poor chain mixing, Right: Multimodal distribution of the rate.

WHAT TO DO?

- Visually inspected the root-to-tip regression to look for patterns.

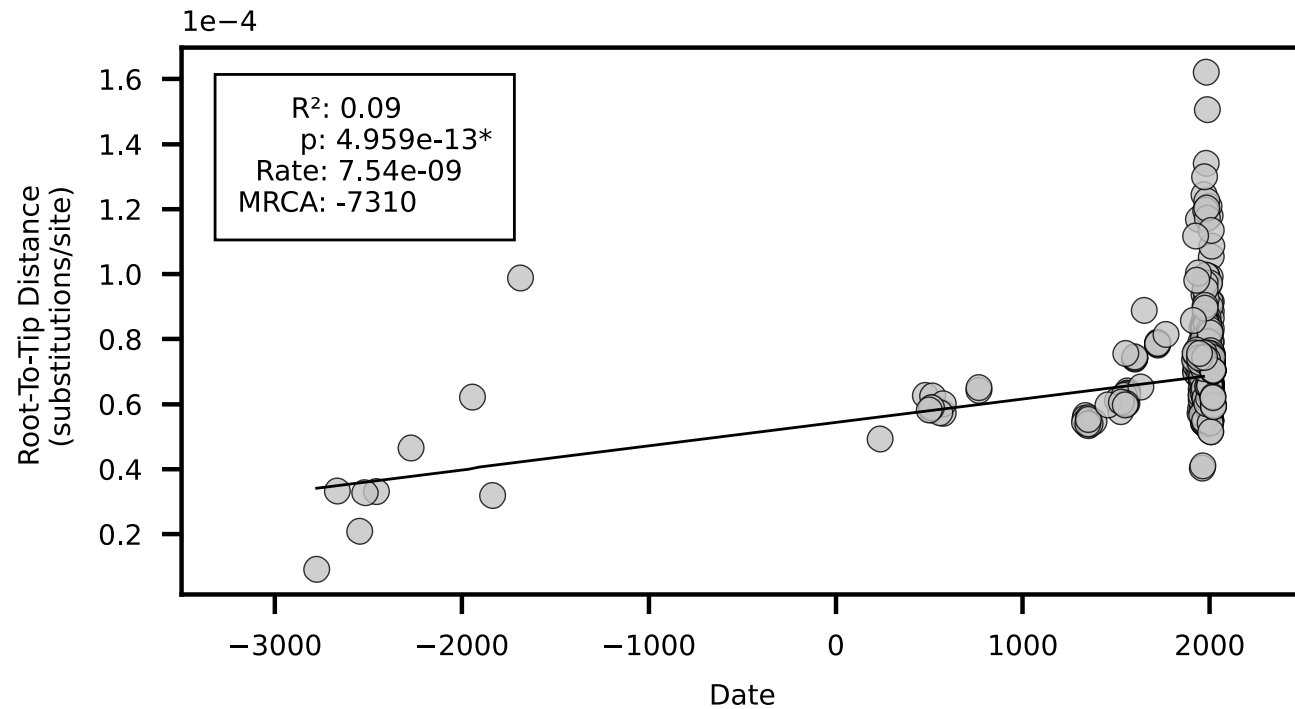


Figure 5: Root-to-tip regression on sampling date.

WHAT TO DO?

- Several clades appear to have their own linear models.

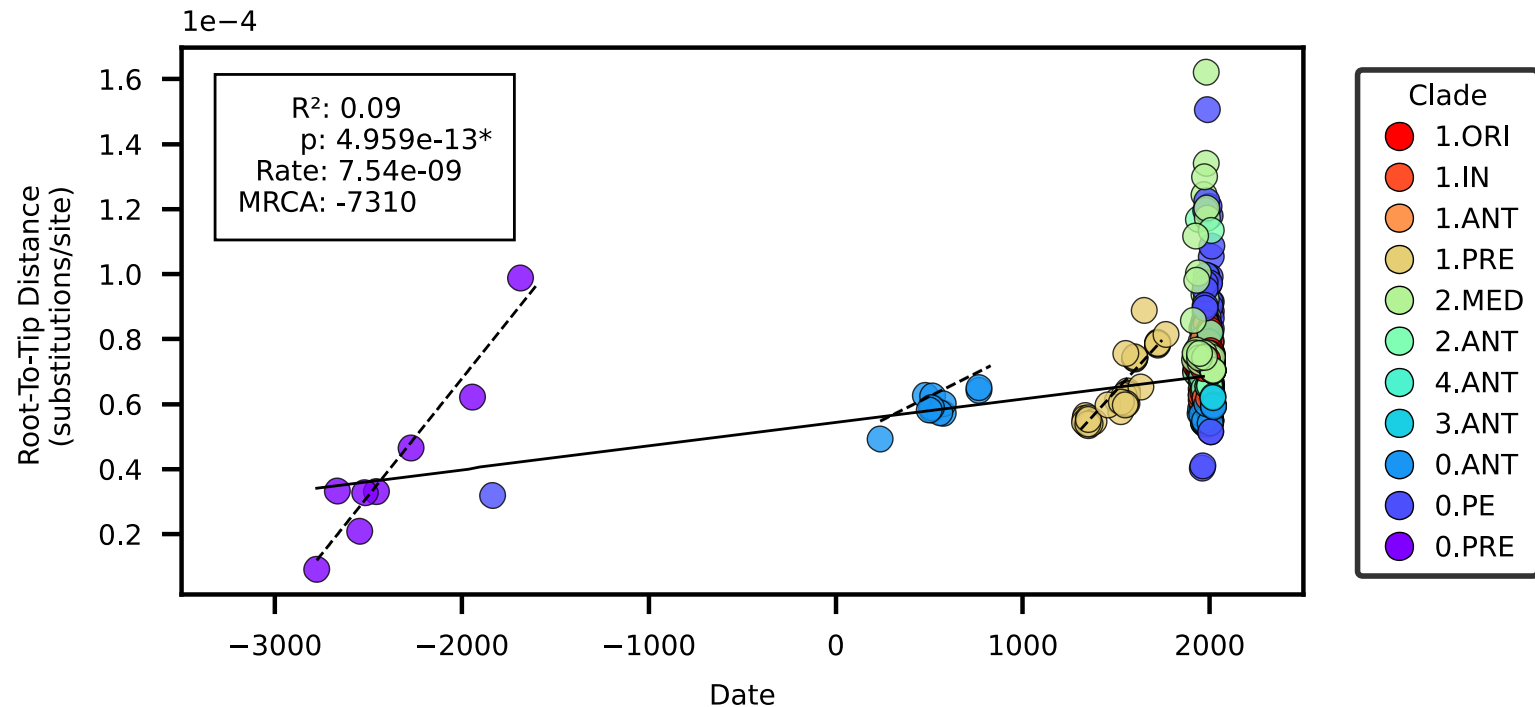


Figure 6: Root-to-tip regression on sampling date by clade.

JOINT (N=600, SPLIT BY CLADE)

- 7 out of 12 clades (46% of samples) have temporal signal using a relaxed clock.
- The remaining clades have no detectable temporal signal.

Table 1: Bayesian Evaluation of Temporal Signal (BETS) statistics.

Clade	N	Bayes Factor	Temporal Signal	Substitution Rate
1.ORI	117	35.7	Yes	1.29×10^{-7}
1.IN	39	-10.2	No	-
1.ANT	4	12.6	Yes	6.48×10^{-8}
1.PRE	40	44.1	Yes	4.77×10^{-8}
2.MED	116	-	No	-
2.ANT	54	-13.7	No	-
4.ANT	11	3.7	Yes	9.38×10^{-8}
3.ANT	11	-11.4	No	-
0.ANT	103	-6.5	No	-
0.ANT4	12	5.9	Yes	3.51×10^{-8}
0.PE	83	12.4	Yes	6.10×10^{-7}
0.PRE	8	83.0	Yes	5.23×10^{-8}

SEQUENTIAL VS. JOINT

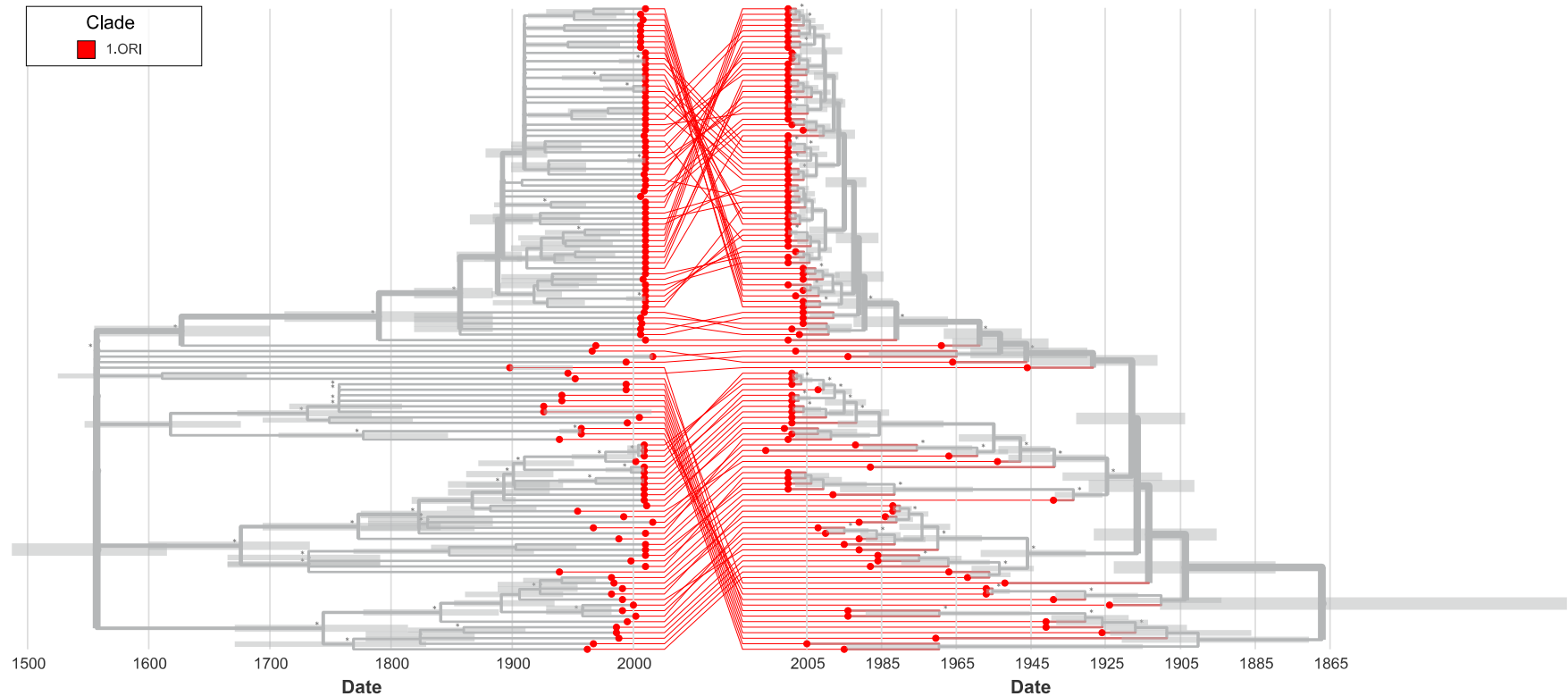


Figure 7: Left: Sequential method (ML) timetree using a single model. Right: Joint method (Bayesian) timetree using clade-specific models

CONCLUSIONS

1. Fitting a single clock model to all of *Y. pestis* is not statistically supported.

The MCMC analyses are relatively unstable and fail to converge in parameter space.

2. *Y. pestis* has much more temporal signal than previously thought.

Separating alignments by clade recovers robust temporal signal.

3. The true substitution rate of *Y. pestis* is much higher than previously thought.

Prior rate estimates were 2×10^{-8} . This study finds it ranges from 3×10^{-8} to 6×10^{-7} .

4. Root-to-tip regression is a poor statistical test of temporal signal.

DON'T TRUST THE RTT

Even in clades with seemingly good strict-clock like behavior!

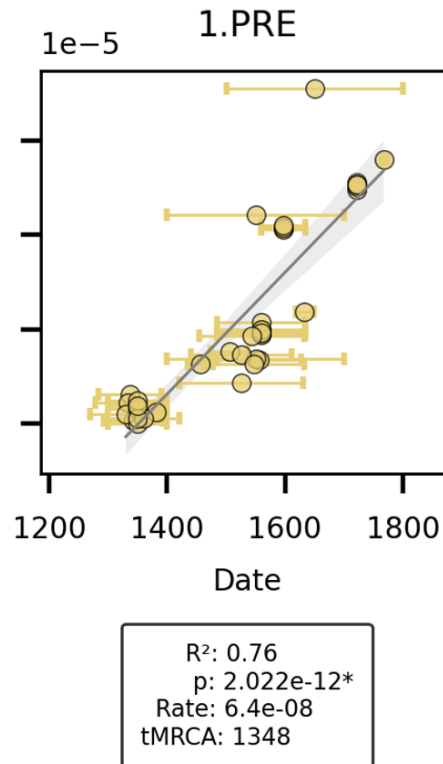


Figure 8: Root-to-tip regression for clade 1.PRE.

WHAT'S NEXT

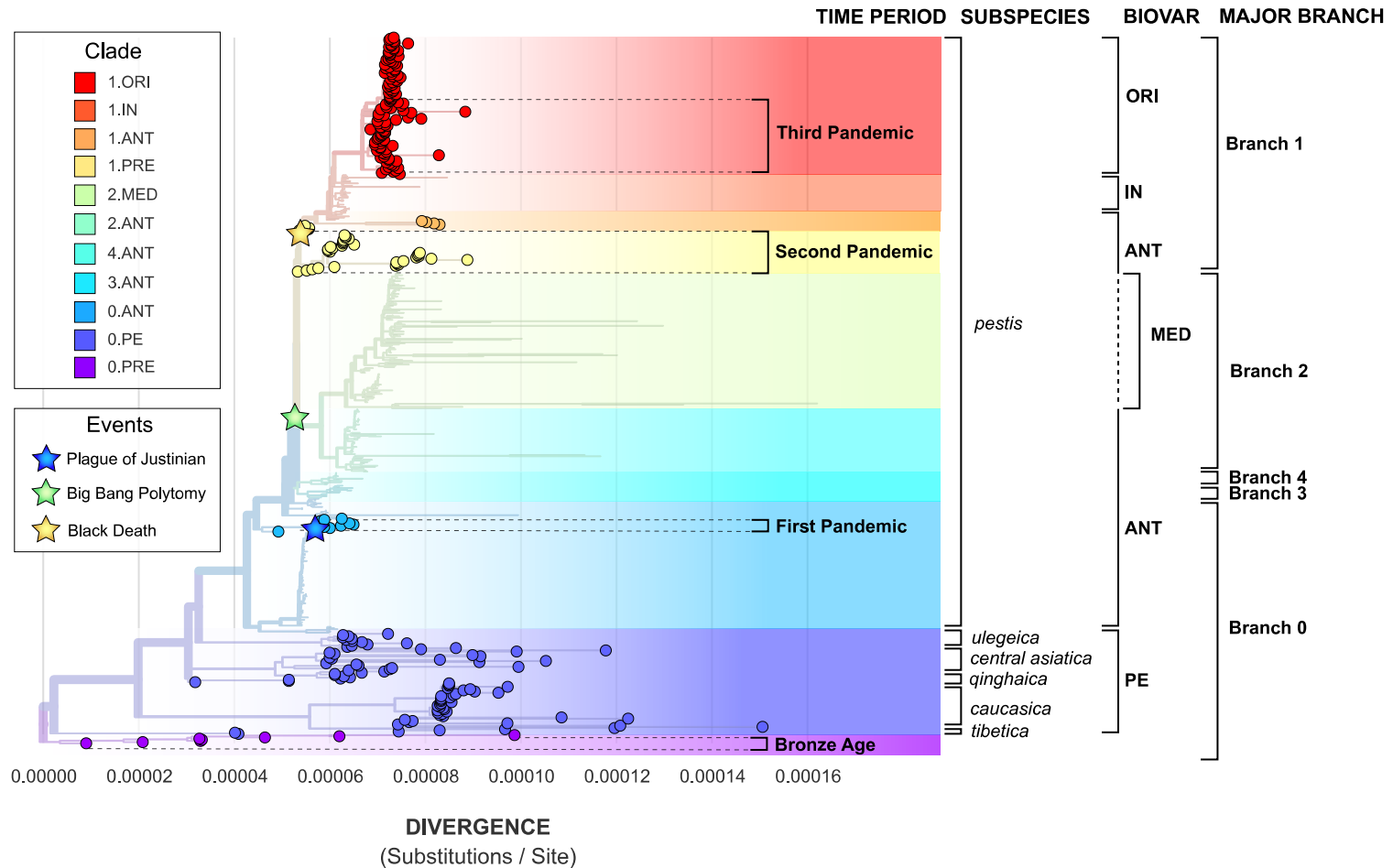


Figure 9: The maximum-likelihood phylogeny showing clades with temporal signal.

REFERENCES

- Cui, Y., Yu, C., Yan, Y., Li, D., Li, Y., Jombart, T., Weinert, L. A., Wang, Z., Guo, Z., Xu, L., Zhang, Y., Zheng, H., Qin, N., Xiao, X., Wu, M., Wang, X., Zhou, D., Qi, Z., Du, Z., ... Yang, R. (2013). Historical variations in mutation rate in an epidemic pathogen, *Yersinia Pestis*. *Proceedings of the National Academy of Sciences*, 110(2), 577–582. <https://doi.org/10.1073/pnas.1205750110>
- Dellicour, S., Durkin, K., Hong, S. L., Vanmechelen, B., Martí-Carreras, J., Gill, M. S., Meex, C., Bontems, S., André, E., Gilbert, M., Walker, C., De Maio, N., Faria, N. R., Hadfield, J., Hayette, M.-P., Bours, V., Wawina-Bokalanga, T., Artesi, M., Baele, G., & Maes, P. (2020). A Phylodynamic Workflow to Rapidly Gain Insights into the Dispersal History and Dynamics of SARS-CoV-2 Lineages. *Molecular Biology and Evolution*, msaa284. <https://doi.org/10.1093/molbev/msaa284>
- Duchene, S., Lemey, P., Stadler, T., Ho, S. Y. W., Duchene, D. A., Dhanasekaran, V., & Baele, G. (2020). Bayesian Evaluation of Temporal Signal in Measurably Evolving Populations. *Molecular Biology and Evolution*, 37(11), 3363–3379. <https://doi.org/10.1093/molbev/msaa163>
- Duchêne, S., Holt, K. E., Weill, F.-X., Le Hello, S., Hawkey, J., Edwards, D. J., Fourment, M., & Holmes, E. C. (2016). Genome-scale rates of evolutionary change in bacteria. *Microbial Genomics*, 2(11). <https://doi.org/10.1099/mgen.0.000094>
- Karpinski, E., Hackenberger, D., Zazula, G., Widga, C., Duggan, A. T., Golding, G. B., Kuch, M., Klunk, J., Jass, C. N., Groves, P., Druckenmiller, P., Schubert, B. W., Arroyo-Cabrales, J., Simpson, W. F., Hoganson, J. W., Fisher, D. C., Ho, S. Y. W., MacPhee, R. D. E., & Poinar, H. N. (2020). American mastodon mitochondrial genomes suggest multiple dispersal events in response to Pleistocene climate oscillations. *Nature Communications*, 11(1), 4048. <https://doi.org/10.1038/s41467-020-17893-z>
- Spyrou, M. A., Keller, M., Tukhbatova, R. I., Scheib, C. L., Nelson, E. A., Andrades Valtueña, A., Neumann, G. U., Walker, D., Alterauge, A., Carty, N., Cessford, C., Fetz, H., Gourvennec, M., Hartle, R., Henderson, M., von Heyking, K., Inskip, S. A., Kacki, S., Key, F. M., ... Krause, J. (2019). Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes. *Nature Communications*, 10(1), 4470. <https://doi.org/10.1038/s41467-019-12154-0>
- Wagner, D. M., Klunk, J., Harbeck, M., Devault, A., Waglechner, N., Sahl, J. W., Enk, J., Birdsell, D. N., Kuch, M., Lumibao, C., Poinar, D., Pearson, T., Fourment, M., Golding, B., Riehm, J. M., Earn, D. J. D., DeWitte, S., Rouillard, J.-M., Grupe, G., ... Poinar, H. (2014). *Yersinia Pestis* and the Plague of Justinian 541 AD: A genomic analysis. *The Lancet Infectious Diseases*, 14(4), 319–326. [https://doi.org/10.1016/S1473-3099\(13\)70323-2](https://doi.org/10.1016/S1473-3099(13)70323-2)

// remote reveal.js plugins // local reveal.js plugins

