

Bayesian Textual Criticism since Hort

A Synthesis and Demonstration

**2023 SBL Annual Meeting
21 November 2023**

Joey McCollum

Australian Catholic University
Institute for Religion and Critical
Inquiry

-  james.mccollum@myacu.edu.au
-  [@JoeyMcCollum](https://twitter.com/JoeyMcCollum)
-  [jjmccollum](https://github.com/jjmccollum)



ACU

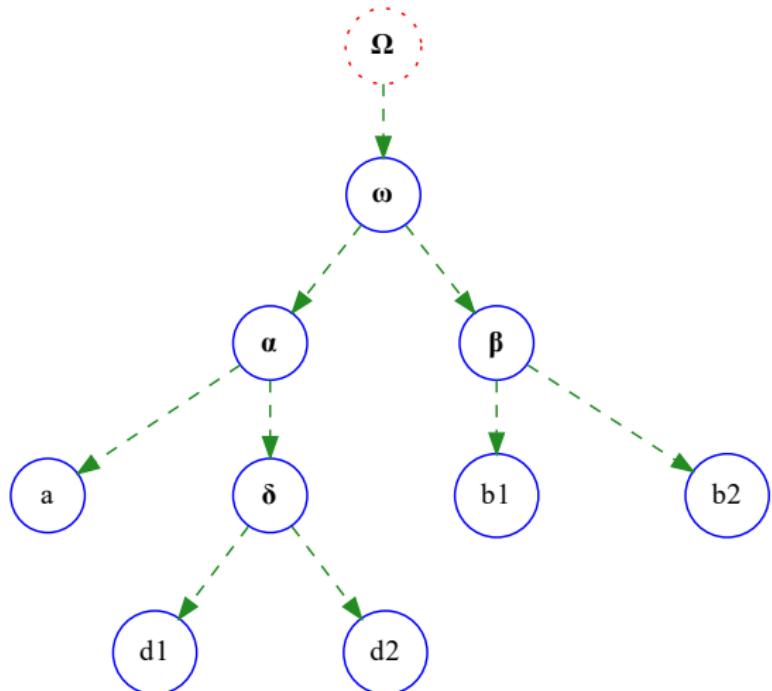
INSTITUTE FOR
RELIGION &
CRITICAL INQUIRY



Wescott & Hort

Source: Wikimedia Commons

- Red, dotted = *intrinsic probability*
- Green, dashed = *transcriptional probability*
- Blue, solid = *genealogical evidence*

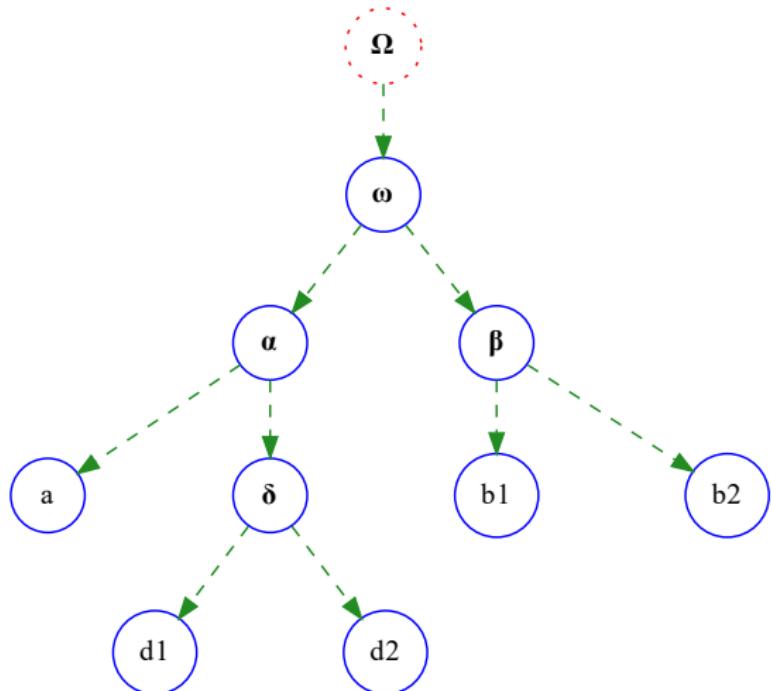




BOOKENDS/HORT & WESTCOTT

Source: Wikimedia Commons
(with adaptations)

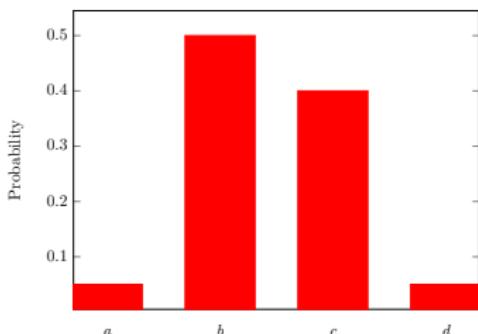
- Red, dotted = *intrinsic probability*
- Green, dashed = *transcriptional probability*
- Blue, solid = *genealogical evidence*



- Similar to the rating system used in the UBS Commentary

<i>Rating</i>	<i>Description</i>
$a \overset{A}{\gg} b$	Reading a is <i>absolutely more likely</i> than reading b .
$a \overset{B}{\gg} b$	Reading a is <i>highly more likely</i> than reading b .
$a \overset{C}{\gg} b$	Reading a is <i>more likely</i> than reading b .
$a \overset{D}{\gg} b$	Reading a is <i>slightly more likely</i> than reading b .
$a = b$	Readings a and b have equal probability.

- Odds ratios between one reading and the next most likely one
- e.g., $b \overset{D}{\gg} c \overset{A}{\gg} a = d$ (right)
- Example values: $A = 19$, $B = 4$, $C = 1.5$, and $D = 1.1$

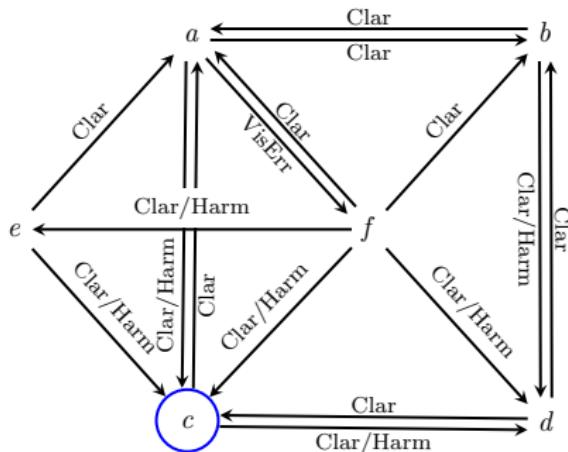


- A system of labels whose relative frequencies will be evaluated as part of the model

<i>Tag</i>	<i>Description</i>
Clar	Clarification in terms of grammar, style, or theology
AurConf	Confusion of similar sounds ($\iota/\epsilon\iota/\eta/o\iota/v$, $a\iota/\epsilon$, o/ω , $\beta/\text{consonantal } v$, π/ϕ , $\kappa\tau\lambda$)
LingConf	Changes due to unfamiliarity with Greek grammar or changes to its rules over time
VisErr	Visual confusion of similar letters, skips and duplications of letters or words
Harm	Harmonization to parallel passage or near context
Byz	Assimilation to the dominant Byzantine text

- This allows us to model (and measure) different classes of scribal habits and a common type of mixture

- We can tag transitions between variant readings in a given unit according to their potential causes
- A stemma's probability is calculated along its branches in terms of transcriptional rates using *Markov chains* (see below)



- Looks and functions like a local stemma, but is less constrained

- Given a hypothesis \mathcal{H} (i.e., a stemma with its parameters describing intrinsic probabilities, transcriptional probabilities, etc.) about how our collation data \mathcal{D} arose, we can calculate its *explanatory power* or *likelihood* $\Pr(\mathcal{D} | \mathcal{H})$ using phylogenetic algorithms
- The *posterior probability* $\Pr(\mathcal{H} | \mathcal{D})$ tells us how certain we can be about \mathcal{H} ; we can estimate it by cleverly sampling different stemmata and combinations of parameters

$$\underbrace{\Pr(\mathcal{H} | \mathcal{D})}_{\text{Posterior}} = \frac{\underbrace{\Pr(\mathcal{D} | \mathcal{H})}_{\text{Likelihood}} \underbrace{\Pr(\mathcal{H})}_{\text{Prior}}}{\underbrace{\Pr(\mathcal{D})}_{\text{Probability of Data}}}$$



Source: Rembrandt van Rijn, *The Apostle Paul* (Wikimedia Commons)

Much in many ways!

Much in many ways!

- A clean separation of concerns between different types of evidence

Much in many ways!

- A clean separation of concerns between different types of evidence
- Accommodation and resolution of tension between intrinsic and transcriptional evidence

Much in many ways!

- A clean separation of concerns between different types of evidence
- Accommodation and resolution of tension between intrinsic and transcriptional evidence
- Built-in estimation of scribal habits

Much in many ways!

- A clean separation of concerns between different types of evidence
- Accommodation and resolution of tension between intrinsic and transcriptional evidence
- Built-in estimation of scribal habits
- Support for witness dates

Much in many ways!

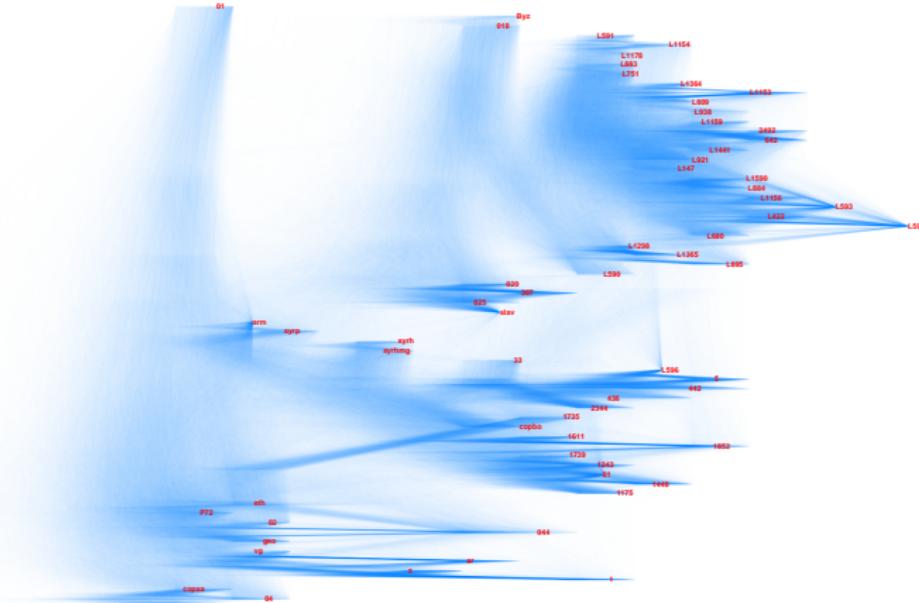
- A clean separation of concerns between different types of evidence
- Accommodation and resolution of tension between intrinsic and transcriptional evidence
- Built-in estimation of scribal habits
- Support for witness dates
- Inclusion of (sufficiently extant) versional and patristic sources

Much in many ways!

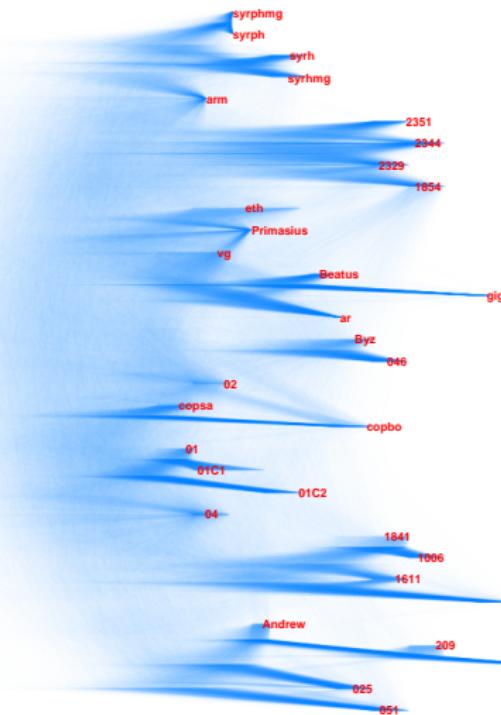
- A clean separation of concerns between different types of evidence
- Accommodation and resolution of tension between intrinsic and transcriptional evidence
- Built-in estimation of scribal habits
- Support for witness dates
- Inclusion of (sufficiently extant) versional and patristic sources
- A measure of (un)certainty about stemma branches, scribal habits, initial readings

Much in many ways!

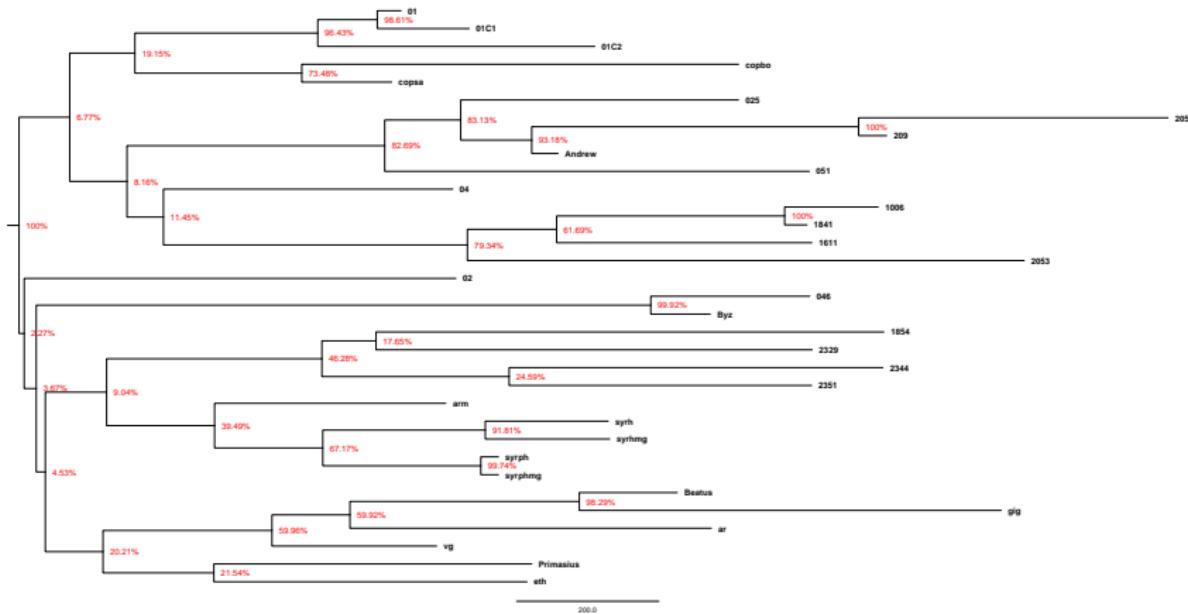
- A clean separation of concerns between different types of evidence
- Accommodation and resolution of tension between intrinsic and transcriptional evidence
- Built-in estimation of scribal habits
- Support for witness dates
- Inclusion of (sufficiently extant) versional and patristic sources
- A measure of (un)certainty about stemma branches, scribal habits, initial readings
- A set of candidate stemmata that can be reconciled to address contamination in a local-genealogical way



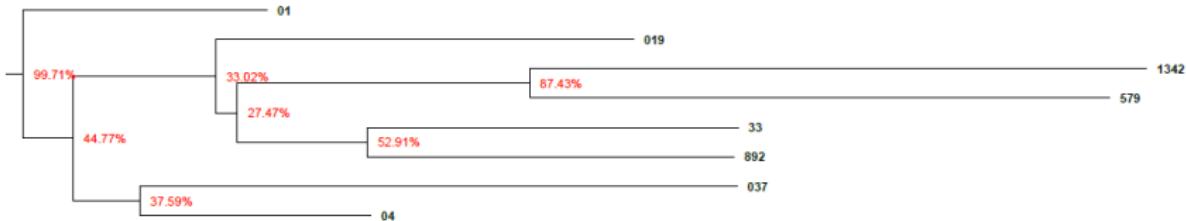
Posterior distribution of stemmata for subset of UBS 1 Peter data at UBS variation units (64 witnesses, 36 variation units, 20,000,000 iterations) with local clock model



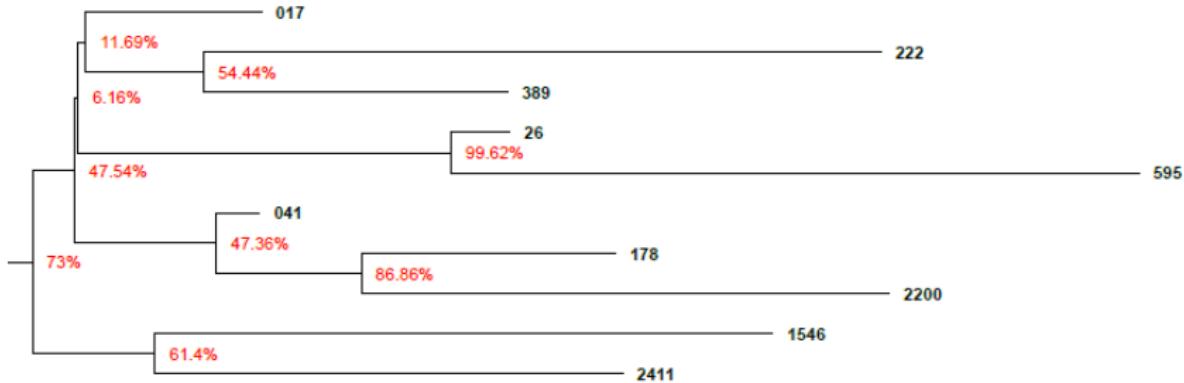
Posterior distribution of stemmata for subset of UBS Revelation data at UBS variation units (33 witnesses, 70 variation units, 20,000,000 iterations) with local clock model



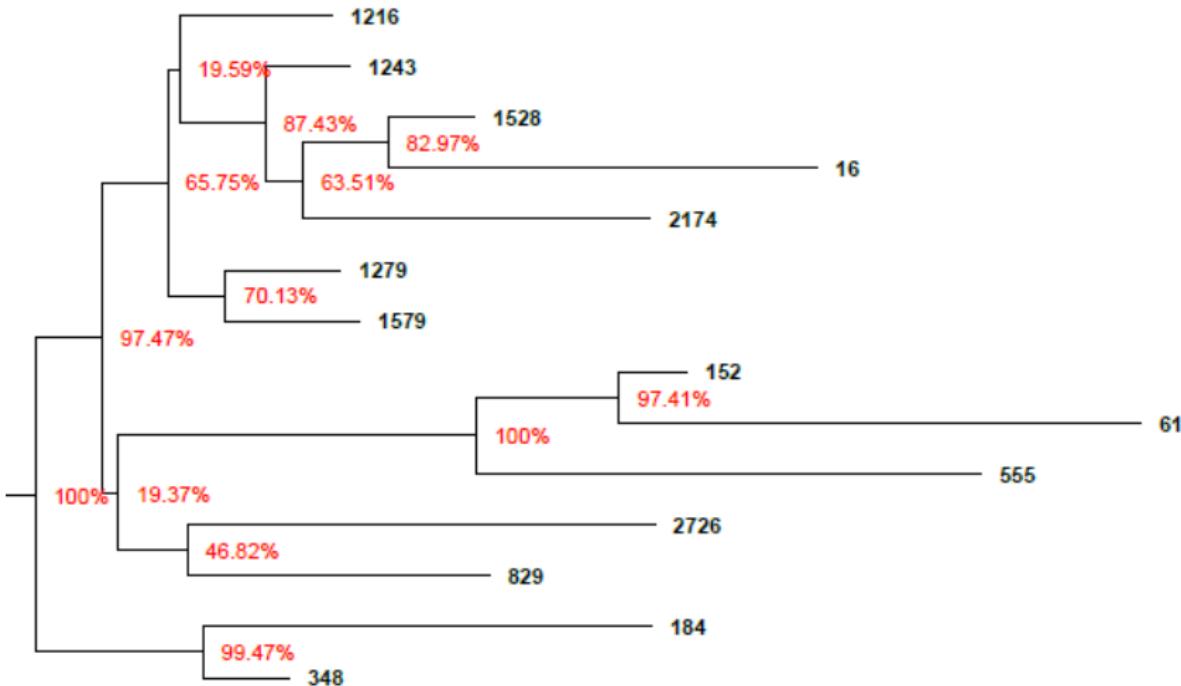
Maximum clade credibility tree for subset of UBS Revelation data at UBS variation units (33 witnesses, 70 variation units, 20,000,000 iterations) with local clock model



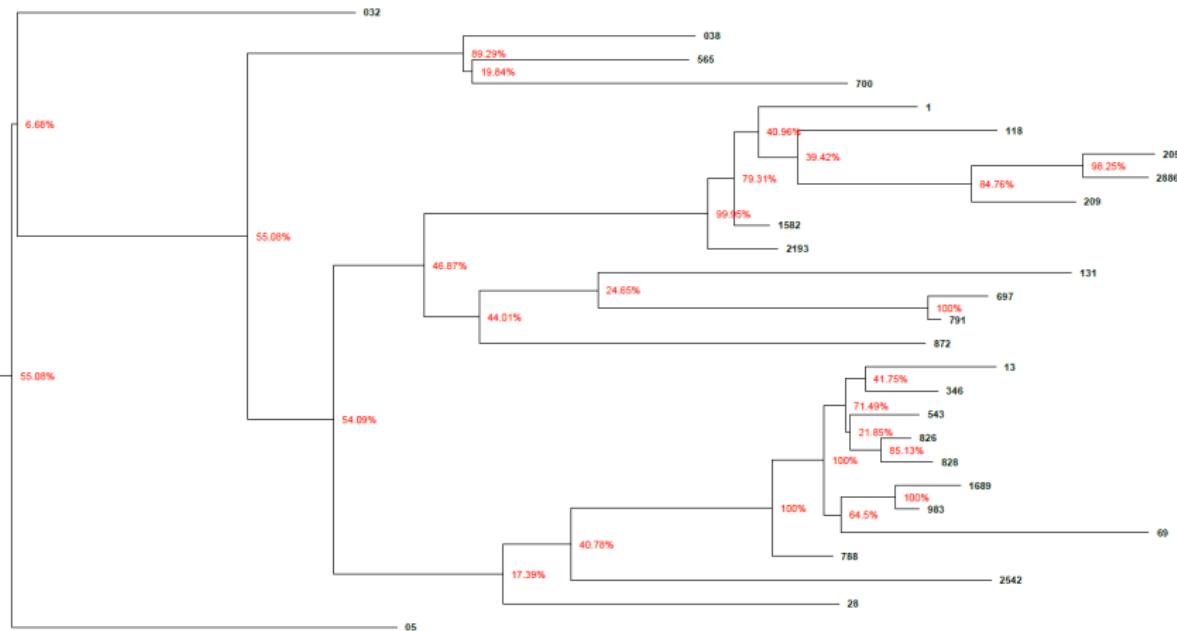
Subset of maximum clade credibility tree for ECM Mark data at UBS variation units (167 witnesses, 148 variation units, 20,000,000 iterations) with local clock model



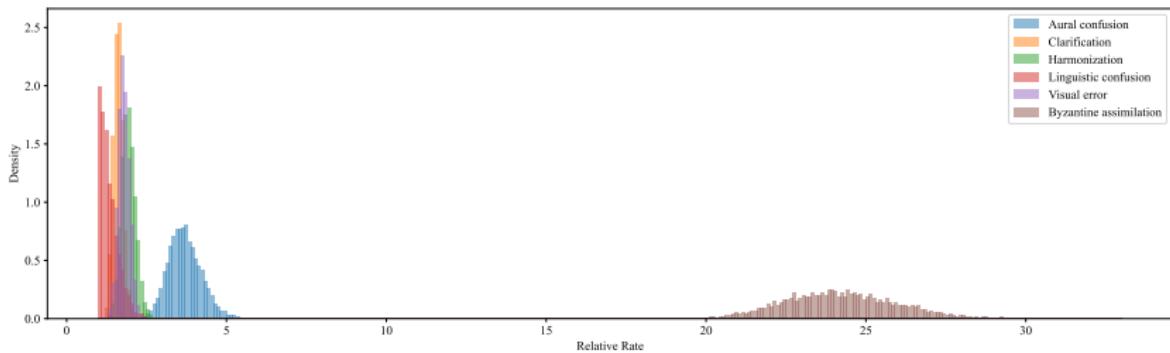
Subset of maximum clade credibility tree for ECM Mark data at UBS variation units (167 witnesses, 148 variation units, 20,000,000 iterations) with local clock model



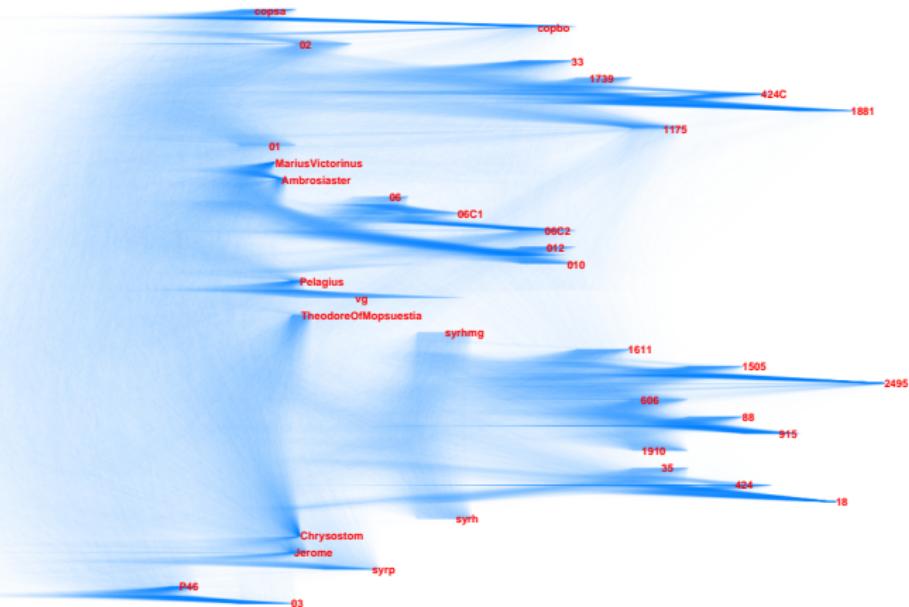
Subset of maximum clade credibility tree for ECM Mark data at UBS variation units (167 witnesses, 148 variation units, 20,000,000 iterations) with local clock model



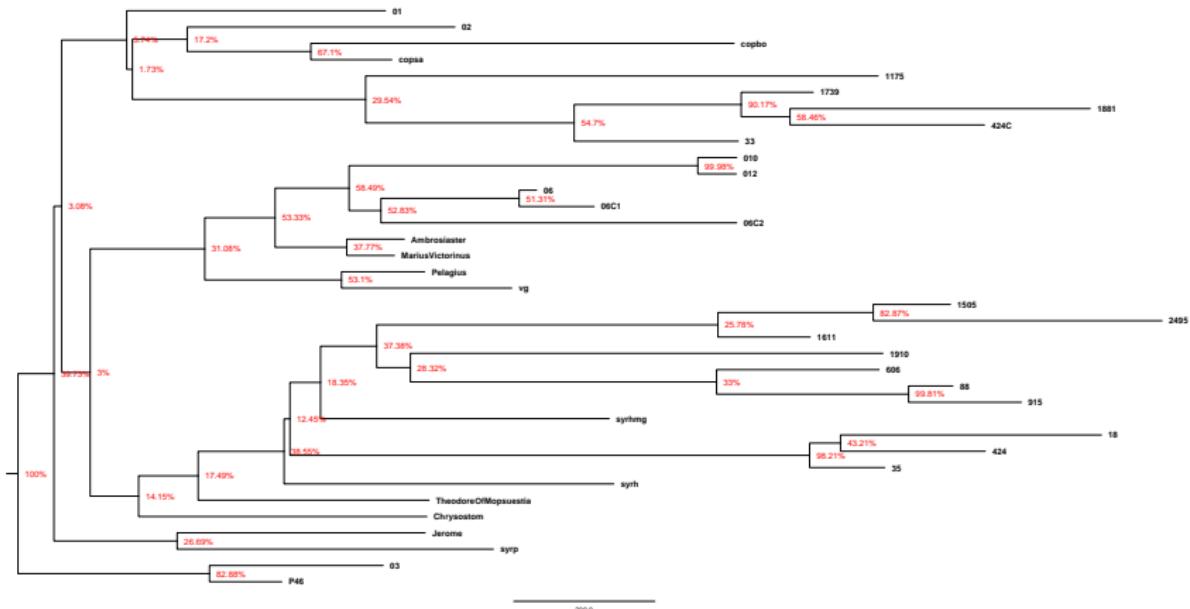
Subset of maximum clade credibility tree for ECM Mark data at UBS variation units (167 witnesses, 148 variation units, 20,000,000 iterations) with local clock model



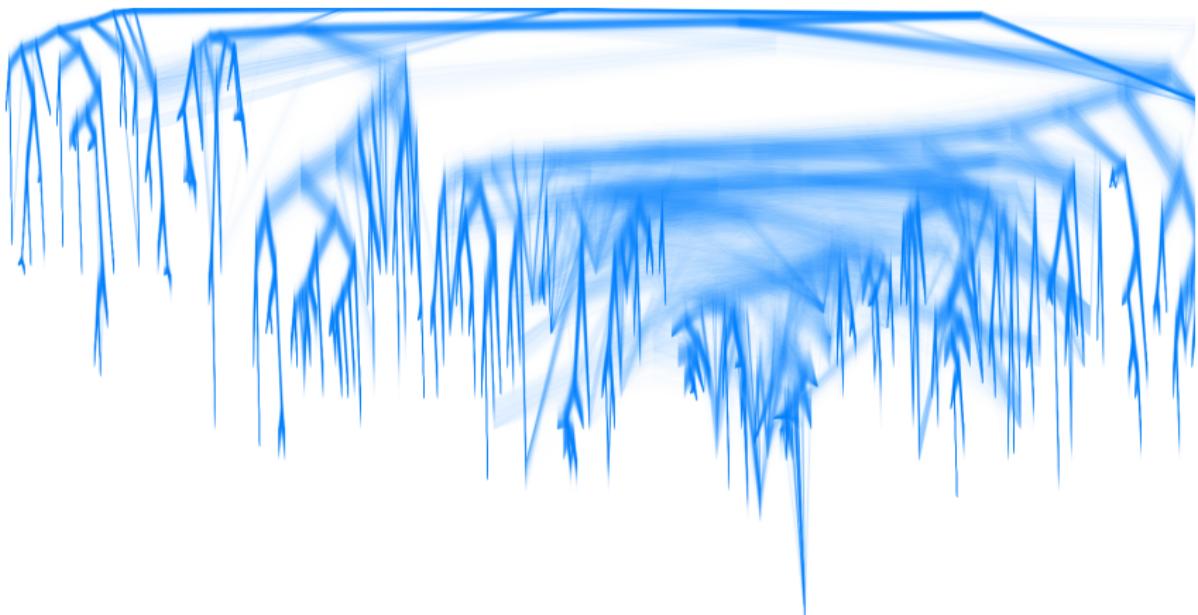
Transcriptional rate posteriors for ECM Mark data at UBS variation units (167 witnesses, 148 variation units, 20,000,000 iterations) with local clock model



Posterior distribution of stemmata for subset of UBS Ephesians data at UBS variation units (36 witnesses, 42 variation units, 20,000,000 iterations) with local clock model



Maximum clade credibility tree for subset of UBS Ephesians data at UBS variation units (36 witnesses, 42 variation units, 20,000,000 iterations) with local clock model



Posterior distribution of stemmata for IGNTP Ephesians data at all variation units (193 witnesses, 915 variation units, 20,000,000 iterations) with strict clock model
(Details to come later!)

For more details:

- Joey McCollum and Robert Turnbull, “teiphy: A Python Package for Converting TEI XML to NEXUS and Other Formats,” *Journal of Open Source Software* 7.80 (2022): 4879, doi:10.21105/joss.04879
- Joey McCollum and Robert Turnbull, “Using Bayesian Phylogenetics to Infer Manuscript Transmission History,” *Digital Scholarship in the Humanities* (in revision)
- Joey McCollum, “Bayesian Stemmatology and Its Application to the Text of Ephesians” (PhD diss., Australian Catholic University, expected 2025)
- The datasets and outputs for this presentation are freely available at <https://github.com/jjmccollum/sbl-2023-talk>