From a Computer-Assisted Stemma to a Phylogenetic Tree: The Medieval Dutch *Martijn Trilogy* by Jacob van Maerlant

Sofie Moors[1,2,3], and Joey McCollum[4]

[1] Institute for the Study of Literature in the Low Countries (ISLN), University of Antwerp, Antwerp, Belgium
[2] Antwerp Center for Digital Humanities and Literary Criticism (ACDC), University of Antwerp, Antwerp, Belgium
[3] Research Foundation Flanders (FWO), Brussels, Belgium
[4] Institute for Religion and Critical Inquiry, Australian Catholic University, Melbourne, Australia

Abstract

This paper examines how computer-assisted stemmatology can bridge the methodological divide between traditional and new philological approaches to medieval textual criticism, using Jacob van Maerlant's *Martijn Trilogy* as a test case. With 18 surviving witnesses—including eight recently discovered manuscripts—this Middle Dutch text provides extensive evidence for understanding scribal variation and manuscript transmission. We reconstruct the 1918 stemma by Verdam and Leendertz and test more traditional stemmatic algorithms and Bayesian phylogenetic methods on datasets of varying inclusivity. We show that computational approaches with more traditional constraints on the classification of readings fail to reconcile conflicting genealogical developments in a single tree, while a Bayesian phylogenetic analysis that models these developments in terms of other scribal and demographic factors can overcome their conflicts and offer a result that measures the reliability of proposed relationships between witnesses. The results reveal that newly discovered single-quire manuscripts cluster with previously dismissed "worthless" fragments, such as E and Z, supporting recent evidence for rapid textual circulation during the second half of the 14th century. This clustering can explain the extensive variation in later witnesses. Ultimately, the Bayesian analysis confirms, refines, and expands on Verdam and Leendertz's reconstruction of the transmission history using traditional methods. We show that even when such an analysis is independent of Verdam and Leendertz's judgments about original readings, its results strongly agree with Verdam and Leendertz's 1918 stemma, which suggests that Bayesian phylogenetics captures many of the relationships they proposed between the manuscripts of the *Martijn Trilogy*.

Keywords: stemmatology, Bayesian Phylogenetics, digital philology, Jacob van Maerlant, medieval literature, Middle Dutch, manuscript transmission

1 Introduction

Traditionally, the primary goal of stemmatology in medieval philology is to establish a critical text edition. This involves reconstructing "the most recent witness from which all extant witnesses of a text" are derived, the so-called *archetype* [30, pp. 4, 221]. This objective is often criticized by proponents of new philology [1, p. 68][27, p. 28]. A new philologist rejects the notion of variation as mere "error" and questions the necessity or existence of a single archetype. As a result,

Sofie Moors, and Joey McCollum. "From a Computer-Assisted Stemma to a Phylogenetic Tree: The Medieval Dutch *Martijn Trilogy* by Jacob van Maerlant." In: *Computational Humanities Research 2025*, ed. by Taylor Arnold, Margherita Fantoli, and Ruben Ros. Vol. 3. Anthology of Computers and the Humanities. 2025, 172–193. https://doi.org/10.63744/vBRBHo3Hn4fX.

the traditional concept of a stemma might be considered less relevant or even useless within this framework [1, p. 71]. What if instead we use a stemmatic analysis not to reconstruct an *archetype* but to gain an understanding of the manuscript tradition itself [27, p. 27][18]? Stemmatic analysis could help us approximate the order of copying and discover where witnesses converge and diverge [1, p. 71][2, p. 232]. In that sense, stemmatology becomes a form of text analysis that can be of great help to medieval philologists "whether of the old school or the new" [1, p. 71].[1]

In this paper, we will try to uncover the history of the *Martijn Trilogy*, a poem written by the renowned medieval Flemish author Jacob van Maerlant (b. ca. 1230–1235, d. ca. 1288–1300) [24]. This short text (1820 verses total) has been subject to extensive copying: with 18 text witnesses surviving, the *Martijn Trilogy* is exceptionally well preserved for this type of Middle Dutch text, providing us with extensive evidence of scribal variation. Yet until now, scholars have studied the transmission primarily to gain insight into Maerlant's original text and its thirteenth-century origins [e.g., 39; 36]. However, this wide transmission also allows us to examine how copyists dealt with a text. To what extent can a stemmatic analysis really help us to study and understand this manuscript tradition of the *Martijn Trilogy*?

2 Out with the Old, in with the New?

Tara L. Andrews calls the difference between old and new philology a "subtle but crucial shift in purpose":

> The older practice of philology, whose methods are taken from traditional classical philology, emphasizes the "ideal" text whose authority supersedes that of any surviving witnesses [...]. Conversely, the emphasis of new philology is on the "real" text as it has been preserved, received, annotated, and used. [1, p. 64].

Andrews suggests "digital philology" as a potential way to navigate and transcend the differences between old and new philology, "allowing the scholar to adopt the best of both approaches as suits the nature and heritage of each individual text" [1, pp. 61–62]. She proposes a five-step workflow for text criticism [1, pp. 66–69], consisting of:

1. a transcription;
2. a collation;
3. an analysis, such as a stemmatological or phylogenetic analysis of the relationships between variants and witnesses;
4. an edition;
5. and finally a full text publication of all text witnesses.[2]

Can these steps help us transcend the distinction between old and new philology? How do we select the best from both approaches, depending on the "nature and heritage" of the *Martijn Trilogy*? There is extensive research on the *Martijn Trilogy* from the approach of old philology. Verdam and Leendertz [37] documented most of the scribal variants in a comprehensive *variant apparatus* and *stemma codicum* (Figure 1 and Figure 2). However, their analyses primarily aimed to determine which variants could be attributed to Jacob van Maerlant himself, reconstructing the "ideal author's original". Fragment Z exemplifies this approach. According to Verdam and Leendertz, this fragment was "worthless" for text criticism, leading to its exclusion from their apparatus [32, p. 57][37, p. xxxv].

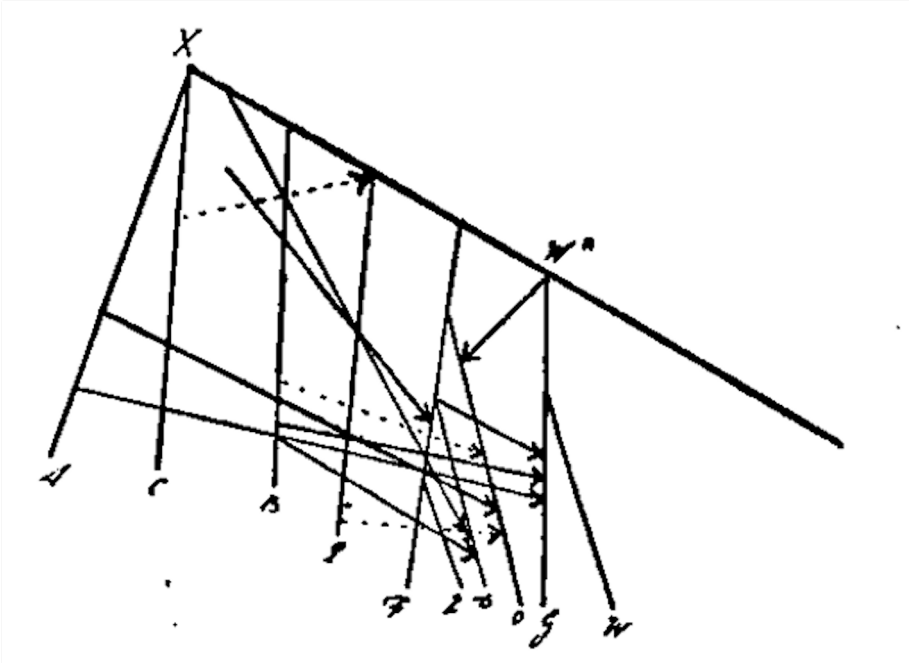


Figure 1: Stemma by Verdam and Leendertz [37] on the basis of ten text witnesses.

Since this last edition over a hundred years ago, eight new witnesses have surfaced (Ant, K, L, Y, Br, D2, Ge, and H), making a revision long overdue [26]. What if we used modern digital tools to create an updated stemma from a new philological perspective, shifting attention from the author to the scribes [1, p. 64][28, p. 98]? In doing so, could we still build upon the foundation laid by Verdam and Leendertz? Or, in Andrews's words, "Is the so-called 'new philology' better suited for the digital age than the 'old' methods that have their root in classical philology, and does the 'old' way have a future?" [1, p. 61].

3 Digital Philology

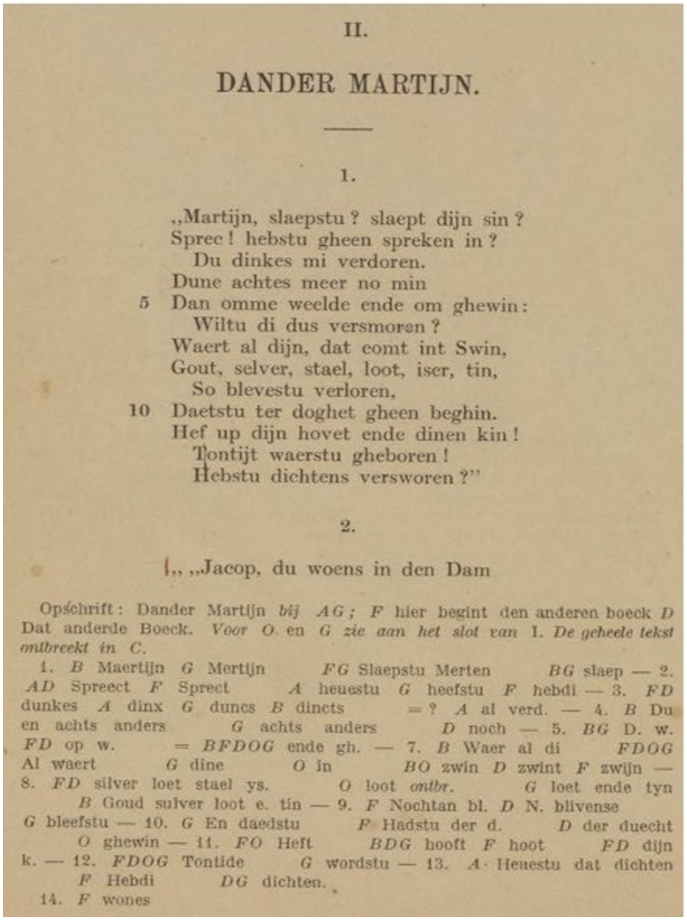
3.1 Transcription

This paper focuses on the third step in digital philology, which is "analysis", but this requires the first two steps of "transcription" and "collation". For the first step, we could use the diplomatic, digital XML-transcriptions of 17 textual witnesses that were published by Moors, Kestemont and Sleiderink [26]. The transcription of an eighteenth textual witness, H, was recently added by Moors [25]. The corpus amounts to over 16,000 verses.

This XML-data incorporates features such as converting capital letters to small letters, removing punctuation, and expanding abbreviations. We also automatically lemmatized words using the GaLAHaD tool [13] to handle spelling variation. Given the lack of consensus among scholars on the types of variants that reveal genealogical relationships, these digital data seem particularly useful. Some researchers argue, for example, that spelling variation is never significant, while others believe it indicates geographical and thus genealogical proximity [38, p. 192]. The capacity of the digital medium to analyze large amounts of data theoretically allows us to not have to limit ourselves to "relevant variants" [4, p. 62]. Instead, we can compare *every* textual variant, which in turn could result in unexpected findings. As Andrews argues, scholars must indeed attempt "to

[1] The *Handbook of Stemmatology* edited by Philipp Roelli provides an overview of both traditional and modern digital, computerized methods for stemmatology [30].

[2] A similar workflow is described by Peter Robinson [29, p. 639].



II.

DANDER MARTIJN.

1.

„Martijn, slaepstu? slaept dijn sin?
Spree! hebstu gheen spreken in?
Du dinkes mi verdoren.
Dune achtes meer no min
Dan omme weelde ende om ghewin:
Wiltu di dus versmoren?
Waert al dijn, dat comt int Swin,
Gout, selver, stael, loot, iser, tin,
So blevestu verloren,
Daetstu ter doghet gheen beghin.
Hef up dijn hovet ende dinen kin!
Tontijt waerstu gheboren!
Hebstu dichtens versworen?"

2.

„Jacop, du woens in den Dam

Opschrift: Dander Martijn *bij AG; F* hier begint den anderen boeck *D* Dat anderde Boeck. *Voor O* en *G zie aan het slot van* 1. *De geheele tekst ontbreekt in C.*
1. *B* Maertijn *G* Mertijn *FG* Slaepstu Merten *BG* slaep — 2. *AD* Spreect *F* Sprect *A* heuestu *G* heefstu *F* hebdi — 3. *FD* dunkes *A* dinx *G* dunes *B* dincts = ? *A* al verd. — 4. *B* Du en achts anders *G* achts anders *D* noch — 5. *BG* D. w. *FD* op w. = *BFDOG* ende gh. — 7. *B* Waer al di *FDOG* Al waert *G* dine *O* in *BO* zwin *D* zwint *F* zwijn — 8. *FD* silver loet stael ys. *O* loot *ontbr.* *G* loet ende tyn *B* Goud sulver loot e. tin — 9. *F* Nochtan bl. *D* N. blivense *G* bleefstu — 10. *G* En daedstu *F* Hadstu der d. *D* der duecht *O* ghewin — 11. *FO* Heft *BDG* hooft *F* hoot *FD* dijn k. — 12. *FDOG* Tontide *G* wordstu — 13. *A* Heuestu dat dichten *F* Hebdi *DG* dichten.
14. *F* wones

Figure 2: The first stanza of M2 in the standard edition by Verdam and Leendertz [37], with variants listed below. Source: Delpher, https://resolver.kb.nl/resolve?urn=mmkb02a:000032275:00007

use *all* the information available to us; to find out if some of those commas are significant after all" [1, pp. 72–73]. However, as we will demonstrate, this ideal proves challenging to implement in practice.

3.2 Collation

The second step before the analysis is defined by Andrews as "collation" of all the witnesses. She argues that automatic collation with a tool such as *CollateX* [11] makes the process less error-prone and time-consuming. At a verse level, collation was facilitated by the unique numbers added to the XML-transcriptions, based on the edition by Verdam and Leendertz [37]. At the word level, automatic collation was performed using *CollateX*. Once the texts have been fully transcribed and collated, the process of analysis may begin.

4 The Analysis

4.1 The Old Stemma

As a baseline and reference point, we first reconstructed the 1918 stemma by Verdam and Leendertz using the online editor Edotor (Figure 3). To make further comparative research possible, we published this reconstruction on Open Stemmata, an organization founded in 2020 by Jean-Baptiste

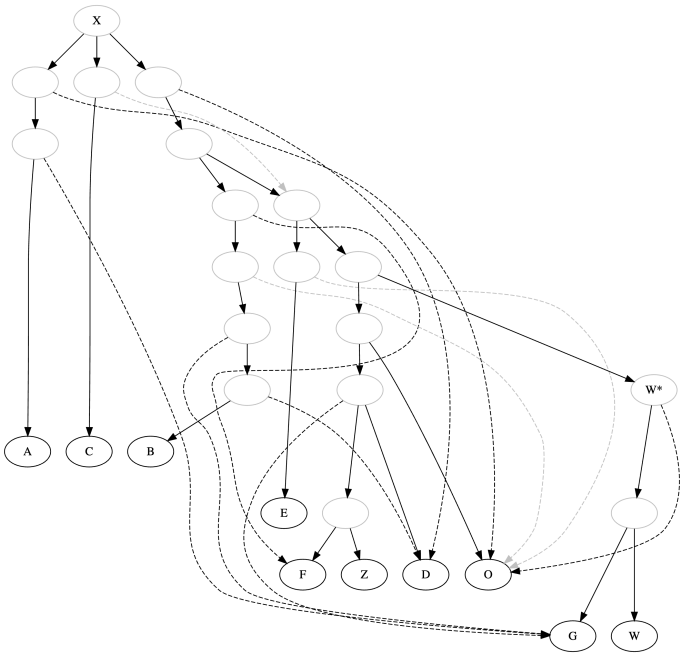


Figure 3: Reconstruction of the stemma by Verdam and Leendertz [37] using the online editor Edotor.

Camps and Gustavo Riva that aims to create an Open Source database of textual genealogies.[3]

The reconstructed stemma reveals considerable complexity. Verdam and Leendertz themselves acknowledged the challenges they faced, asking for indulgence in advance, as they could not always succeed in finding the right way in the "labyrinth of crisscrossing, eye-confusing paths" [37, p. xxii]. The 20th-century stemma consists of three main branches from which the entire extant tradition derives (Figure 4):

1. a branch represented by manuscript A;
2. a branch represented by manuscript C; and
3. a branch represented by all the other text witnesses B–G(W).

Verdam and Leendertz [37, p. xxxv] noted that groups A and C are "the older text witnesses" and have been "out of circulation" in remote monasteries. These manuscripts show no influence of other witnesses. Regarding the final group B–G(W), they argue that while all manuscripts are mutually similar, B is distinguished by its age and "greater independence", since it too was not influenced by other manuscripts. The aggregation of B with A and C is therefore of great importance, they conclude [37, pp. xxv, xxxvi].

They consider the remaining manuscripts, located at the intersection of E and W*, to represent a more strongly deviated redaction known as "the Vulgata" (E, F, Z, D, O, G, and W) [37, p. xxxv]. A "vulgate text" is a textual form "that reached the widest distribution at a time, possibly long after the archetype, when interest in the text experienced an upsurge for one reason or another and many copies were made" [30]. During periods of high textual interest, scribes often compared multiple exemplars to create what they considered a better, more correct text [30, p. 225]. This

[3] `https://github.com/OpenStemmata/database/tree/main/data/dum/VerdamLeendertz_1918_MartijnTrilogy`.

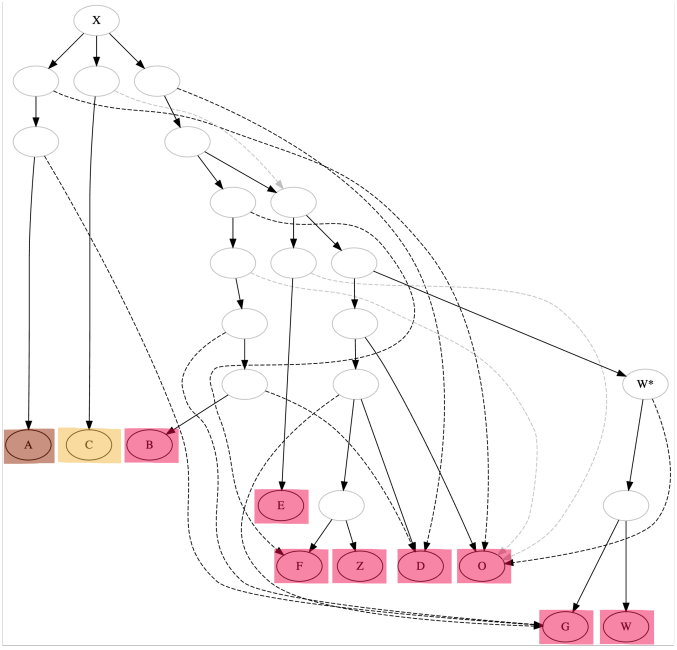


Figure 4: Reconstruction of the stemma by Verdam and Leendertz [37]. Color-coding indicates the three main branches represented by witnesses A (brown), C (yellow) and B–G(W) (pink).

practice, known as “contamination”, involves “the copying of readings from more than one exemplar, resulting in complex and often hard-to-detect relationships between textual witnesses within the transmission of a text” [17, p. 254].

Verdam and Leendertz concluded that Vulgata scribes indeed followed different exemplars, creating what they termed “manifold interferences” [37, p. xxxvi]. They consequently deemed these witnesses of “less significance” due to these complex relationships and because they represent “the youngest redaction” [37, p. xxxvi]. However, the scholars also recognized their value as “means of control”, since they sometimes preserved elements from older redactions (indicated by the dotted lines). For instance, the witnesses show influence of class C and had access to material from classes A and B, with manuscript O possibly using B as an exemplar [37, pp. xxx, xxxvi, xxxix]. According to the researchers, only manuscript E, also the oldest, preserved the Vulgate redaction in relatively “pure” form, while G and W, on the other hand, “can hardly claim an independent position” [37, p. xxxvi].

Their analysis clearly reflected old philological principles. Recent scholarship, however, has revealed new insights that challenge their conclusions. For example, Verdam and Leendertz distinguished manuscript B from other Vulgate witnesses partly based on its presumed early date, but Erik Kwakkel has recently redated this witness to 1450–1500—more than half a century later than they had assumed (personal communication). Other recent research by Moors [25] has confirmed that such an upsurge in textual production did occur at the end of the 14th century. In addition to the already known fragments E and Z, the newly discovered manuscripts Ge, L, Ant, Y, Br and H were all distributed as “single-quire manuscripts”—easily producible booklets that facilitated rapid distribution. Could this rapid late-14th-century distribution explain the extensive variation among witnesses, and might these manuscripts be more significant than previously thought for understanding the complex transmission of the *Martijn Trilogy*?

4.2 The Computer-Assisted Stemma

Computational procedures and tools for doing stemmatics are appearing at a rapid pace [30, p. 4]. These approaches can be broadly categorized into two groups: those that build on established text-critical methodologies (such as Lachmannian methods) and those that draw inspiration from disciplines outside traditional textual criticism, including bioinformatic algorithms such as (Bayesian) phylogenetics [8, p. 66][15, p. 339]. In the former approaches, the stemma is built piece by piece through the isolation of larger and larger families of witnesses based on shared indicative errors (textual changes that are one-directional and unlikely to have occurred coincidentally). The latter approaches treat a stemma as a hypothesis about the relationships of witnesses, and they search through millions of these hypotheses to find ones that best explain the observed data of the surviving tradition. With the help of computers, these newer approaches address the construction of a stemma within a "big data" paradigm, incorporating most or all textual variations (including those that do not correspond to traditional indicative errors). In this paradigm, paratextual and historical information about the transmission process can also be included as evidence to be explained by the stemma. Given our investigation into which philological approach—old or new—is most appropriate, we have experimented with tools from both categories.[4]

For the more traditional methodology, we used *Stemmatology: an R package for the computer-assisted analysis of textual traditions* by Jean-Baptiste Camps and Florian Cafiero [8]. They implement a method based on traditional philological principles in a user-friendly package, offering an algorithmic and easily computable transposition of the "common error" method [8, p. 66]. The package identifies conflicts between variant locations through pairwise genealogical comparisons and visualizes these conflicts as a network graph where problematic variants can be identified and eliminated. It also detects potential contamination by systematically removing witnesses to see which ones significantly reduce overall conflict. Finally, it constructs a stemma using a disagreement-based algorithm that focuses on disagreements where at least two witnesses oppose at least two others. The algorithm proceeds step-by-step to assess manuscript groups and eliminate derivative copies until a genealogical tree emerges—all while maintaining strong user interaction to guide the computational analysis with scholarly expertise.

Camps and Cafiero's package [8, p. 66] does not prescribe a particular data model for representing textual variations, recognizing that both the basic unit of variation and the types of variation to incorporate may vary considerably across different contexts and projects. In other words, they offer complete flexibility: "The exact meaning to give to variant location and variant is defined by the user, according to his or her approach and the nature of the materials being analysed" [8, pp. 66–67]. To remain independent of all these choices, they adopt a very simple and abstract presentation, where each column stands for a witness, and each line for a variant location. Each variant gets a numeric code (0, 1, $\ldots n$).

While this represents a significant advantage, it also poses the greatest challenge. In an ideal, new philologist world, stemmata would be generated with different input parameters—with and without spelling variation, with and without punctuation, and so forth. Using the digital data and the automated collation tool *CollateX*, this approach should be feasible. *CollateX* has an option to store output as a TEI XML variant apparatus, displaying witnesses that cluster together.

Unfortunately, in practice, at least with our corpus, this approach is not so straightforward. Due to the extensive spelling and orthographic variation inherent in Middle Dutch, the variant apparatus contains an extreme amount of variation and fails to show consistency between witnesses where such consistency clearly exists. Even after normalization (removing punctuation, expanding abbreviations, and converting capital letters to small letters) and automatic lemmatization with the

[4] The basic workflow used by Katarzyna Anna Kapitan [18] is similar to ours, but her study employs different phylogenetic software (PHYLIP package) and takes a material-philological and transmission-historical approach as its starting point.

```
<rdg wit="#A">vvaphene</rdg>
<rdg wit="#B #L #O #Y">waphene</rdg>
<rdg wit="#C">uuapene</rdg>
<rdg wit="#D #G">wapen</rdg>
<rdg wit="#F">wapene</rdg>
```

```
<rdg wit="#A">waphenen</rdg>
<rdg wit="#B #O">Wafene</rdg>
<rdg wit="#C">-</rdg>
<rdg wit="#D #F #G">wapen</rdg>
<rdg wit="#L">wafen</rdg>
<rdg wit="#Y">waphenen+hij</rdg>
```

Figure 5: Excerpts from the TEI XML variant apparatus generated by *CollateX* showing the first word of the *Martijn Trilogy*. The left lines show the original corpus; the right lines show the lemmatized version (GaLAHaD) [13]. For manuscript C, GaLAHaD could not identify a lemma.

tool GaLAHaD, the problem persists [13].[5] In this example—which represents an extreme case—this actually creates an additional variant (Figure 5). According to Andrews [1, p. 68], using digital tools should have "minimized the temptation to curtail or 'normalize' the data prematurely and have avoided the need to assess the significance of any piece of evidence". In reality, however, some intervention still appears necessary before the "analysis" step can begin.

Based on our "approach and the nature" of the material we want to analyze, we decided not to work with the automatic generated variant apparatus but with two manually generated input files representing the "old" and "new" philologist approach.[6] For the "old" approach, we focused on variants identified by Verdam and Leendertz [37] as "important" for reconstructing their stemma. Their stemma, in fact, precedes nearly 15 pages of running text listing specific verses they believed contained meaningful variants. For these locations, we have added the variants from the eight newly discovered witnesses (Y, Ge, D2, K, Br, L, H, Ant). The data frame was created in Excel and then converted to CSV format. This final version included 116 variation units in total (Figure 6).

For example, row 21 in Figure 6 represents verse 200 from the *Eerste Martijn*. The variation occurs in the rhyming word (Figure 7): witnesses A, B, C and G have "heet", while D, F and O have "wreet". The witnesses D, F and O receive a 2 instead of a 1, because they share this variant reading. The witnesses Ge, D2, K, W, Br, E, Z, H and Ant are marked "NA" because this verse is not present in these manuscripts, while 0 indicates a textual omission.

To validate our approach, we first tested the data frame using only the witnesses known by Verdam and Leendertz in 1918, thus ignoring the columns representing the new witnesses. This allowed us to determine whether the computational method would reproduce the stemma that they constructed using the same "meaningful" variants. However, the analysis of witnesses A, B, C, D, E, F, G, O, W and Z required lowering the conflict threshold to 0.01 to generate any usable results. At higher thresholds, the abundance of conflicting variants prevented the algorithm from identifying clear genealogical relationships. This high level of textual conflict suggests that even the variants Verdam and Leendertz considered most reliable for stemmatic analysis contain significant contradictions, highlighting the inherent complexity of this manuscript tradition.

The computational stemma (Figure 8) shows both convergences and divergences with the 1918 analysis. While it successfully identifies the separation of manuscripts A, B and C and recognizes smaller groupings like F–Z, G–W and O–E, it produces a fundamentally different architecture. Instead of three main branches, the algorithm generates a binary structure with branches "GWDFZ" and "OEBCA".

[5] Another possible implementation could be a Levenshtein metric to ignore these differences, but Levenshtein distance does not distinguish between allographs—different graphic forms of the same phoneme or letter, such as *u* versus *v*—and therefore does not offer a straightforward solution for orthographic variation.

[6] All the data are available on the companion GitHub repository: `https://github.com/SofieMoors/martijncollation`.

	F	D	G	O	B	Y	C	Ge	D2	K	W	A	Br	E	L	Z	H	Ant
1	2	2	2	2	3	2	2	NA	NA	NA	NA	1	NA	NA	2	NA	NA	NA
2	2	2	1	1	1	1	1	NA	NA	NA	NA	1	NA	NA	1	NA	NA	NA
3	0	0	1	0	1	0	1	NA	NA	NA	NA	1	NA	NA	0	NA	NA	NA
4	2	2	1	2	1	2	1	NA	NA	NA	NA	1	NA	NA	2	NA	NA	NA
5	2	2	1	2	1	2	2	NA	NA	NA	NA	1	NA	NA	1	NA	NA	NA
6	0	0	1	0	1	0	1	NA	NA	NA	NA	1	NA	NA	0	NA	NA	NA
7	1	1	2	1	3	2	1	NA	NA	NA	NA	2	NA	NA	1	NA	NA	NA
8	2	2	2	2	2	2	3	NA	NA	NA	NA	1	NA	NA	1	NA	NA	NA
9	1	2	2	1	1	1	1	NA	NA	NA	NA	1	NA	NA	2	NA	NA	NA
10	2	2	3	2	1	2	2	NA	NA	NA	NA	1	NA	NA	2	NA	NA	NA
11	2	2	1	1	1	2	1	NA	NA	NA	NA	1	NA	NA	2	NA	NA	NA
12	2	2	2	1	2	2	2	NA	NA	NA	NA	1	NA	NA	2	NA	NA	NA
13	1	2	2	3	1	1	1	NA	NA	NA	NA	1	NA	NA	1	NA	NA	NA
14	2	2	2	2	2	2	1	NA	NA	NA	NA	1	NA	NA	2	NA	NA	NA
15	2	3	2	2	1	2	1	NA	NA	NA	NA	1	NA	NA	2	NA	NA	NA
16	0	0	0	0	0	0	1	NA	NA	NA	NA	1	NA	NA	0	NA	NA	NA
17	1	1	1	2	2	1	1	NA	NA	NA	NA	NA	NA	NA	1	NA	NA	NA
18	2	2	3	4	1	2	3	NA	NA	NA	NA	1	NA	NA	2	NA	NA	NA
19	2	2	2	1	1	NA	1	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA
20	0	0	1	1	1	NA	1	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA
21	2	2	1	2	1	NA	1	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA
22	2	2	2	2	2	NA	1	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA
23	2	2	2	2	2	NA	1	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA
24	2	2	2	2	2	NA	1	NA	NA	NA	NA	1	NA	2	NA	NA	NA	NA
25	2	2	2	2	1	NA	1	NA	NA	NA	NA	1	NA	2	NA	NA	NA	NA
26	2	2	2	2	2	NA	1	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA
27	2	1	2	2	1	NA	2	NA	NA	NA	NA	1	NA	2	NA	NA	NA	NA
28	0	0	1	1	1	NA	1	NA	NA	NA	NA	1	NA	1	NA	NA	NA	NA
29	2	2	2	2	1	NA	2	NA	NA	NA	NA	1	NA	2	NA	2	NA	NA
30	2	2	2	2	2	NA	1	NA	NA	NA	NA	1	NA	2	NA	2	NA	NA
31	2	2	2	2	1	NA	1	NA	NA	NA	NA	1	NA	2	NA	2	NA	NA
32	2	1	1	1	1	NA	1	NA	NA	NA	NA	1	NA	1	NA	2	NA	NA
33	1	3	1	2	4	NA	1	NA	NA	NA	NA	1	NA	2	NA	1	NA	NA
34	3	2	NA	2	1	NA	1	NA	NA	NA	NA	1	NA	2	NA	2	NA	NA
35	2	1	NA	2	1	NA	2	NA	NA	NA	NA	2	NA	2	NA	2	NA	NA
36	2	2	1	2	2	NA	2	NA	NA	NA	NA	1	NA	2	NA	1	NA	NA
37	3	3	3	1	1	NA	2	NA	NA	NA	NA	1	NA	2	NA	3	NA	NA
38	0	0	0	0	2	NA	1	NA	NA	NA	NA	1	0	0	NA	0	NA	NA
39	2	2	2	2	2	NA	1	NA	NA	NA	NA	1	2	2	NA	2	NA	NA
40	2	2	1	2	1	NA	1	NA	NA	NA	NA	1	1	1	NA	1	NA	NA
41	1	1	1	2	1	NA	1	NA	NA	NA	NA	1	1	2	NA	1	NA	NA
42	2	1	2	1	1	NA	1	NA	NA	NA	NA	1	2	2	NA	2	NA	NA
43	1	1	3	2	1	NA	1	NA	NA	NA	NA	2	1	1	NA	2	NA	NA
44	1	2	1	2	2	NA	1	NA	NA	NA	NA	1	1	1	NA	1	NA	NA
45	1	1	1	2	1	NA	1	NA	NA	NA	NA	1	1	2	NA	NA	NA	NA
46	1	1	1	2	2	NA	1	NA	NA	NA	NA	1	1	1	NA	NA	NA	NA
47	1	2	1	2	2	NA	1	NA	NA	NA	NA	1	1	1	NA	NA	NA	NA
48	1	2	2	1	NA	NA	1	NA	NA	NA	NA	1	2	NA	NA	NA	NA	NA
49	2	2	2	2	2	NA	2	NA	NA	NA	NA	1	2	NA	NA	NA	2	NA
50	2	1	2	2	1	NA	2	NA	NA	NA	NA	1	2	NA	NA	NA	2	NA

Figure 6: Input data frame for Camps and Cafiero's R package based on variants listed by Verdam and Leendertz [37] with witnesses as columns and variant locations as rows.

M1_16_200_A	dat	hi	–	den	sondare	es	so	heet
M1_16_200_B	dat	hi	es	den	sondare	–	so	heet
M1_16_200_C	dat	hi	–	den	zondaer	is	zo	heet
M1_16_200_D	dat	hi	–	den	sondaer	is	soe	wreet
M1_16_200_F	datti	–	–	den	sondare	es	so	wreet
M1_16_200_G	dat	hi	–	den	sondare	es	soe	heet
M1_16_200_O	dat	hij	–	den	zondare	es	zo	wreet

Figure 7: Collation of verse 200 from the *Eerste Martijn* showing word-level variation across witnesses A, B, C, D, F, G and O.

Our primary objective, however, was to construct an updated stemma that included the newly discovered witnesses. This task proved significantly more complex. Despite lowering the threshold to 0.0001, the algorithm continued to encounter irreconcilable conflicts. As mentioned above, the 1918 stemma was constructed primarily according to old philological principles. When ev-

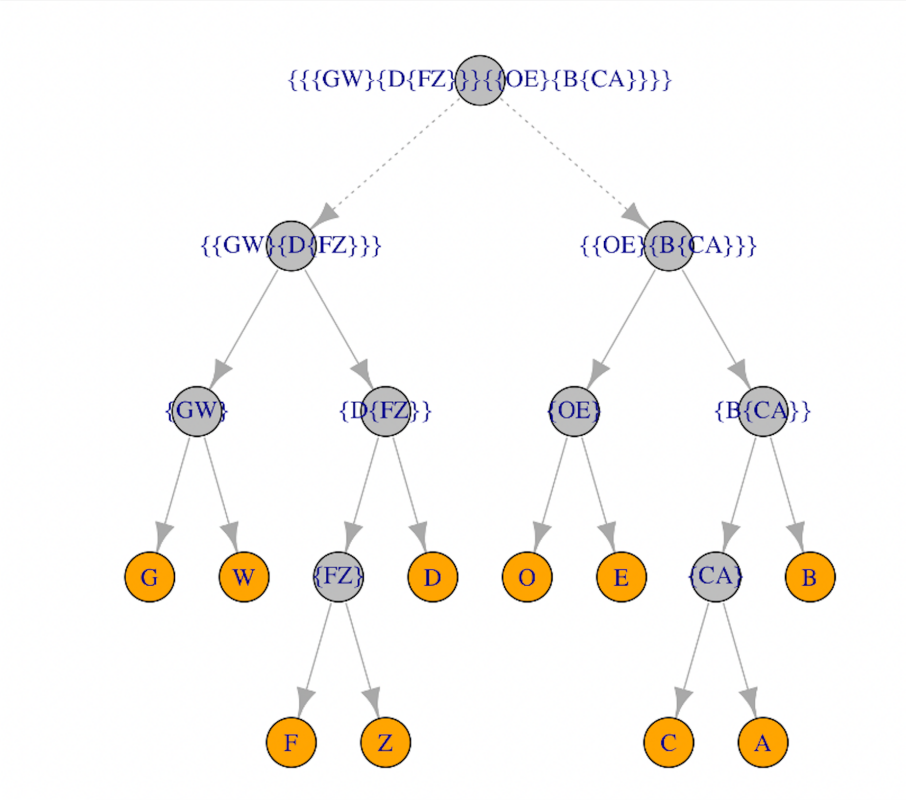


Figure 8: Stemmatic output from Camps and Cafiero's R-package, based on "meaningful variants" defined by Verdam and Leendertz [37].

idence contradicted their assumptions, Verdam and Leendertz quickly dismissed it as "too uncertain" or "coincidental" [37, p. xxxii]. To test whether the algorithm's failure stemmed from insufficient data or rather from these restrictive editorial choices, we expanded our dataset using more inclusive criteria. Following the methodology established by Bosmans and Sleiderink [31] and Moors [23], we created an extended variant apparatus that captures a broader range of textual differences. This approach prioritizes variations like omissions, additions, word replacements, and altered word order, while excluding features like dialectal and spelling differences, semantically neutral inflectional variants, gender alternations, and cliticization effects.

The expanded dataset contains 435 variation units, nearly four times the original size. However, this inclusivity revealed the fundamental challenge: rather than resolving the computational conflicts, additional data introduced even more "anomalous variant locations". This suggests that the manuscript tradition's complexity cannot be solved simply by gathering more evidence—the relationships themselves may be genuinely contradictory, reflecting the contaminated transmission that Verdam and Leendertz struggled to accommodate within their genealogical framework.

4.3 The Bayesian Phylogenetic Tree Distribution

We therefore turned to phylogenetic methods that could better handle conflicting data. Recent research by Joey McCollum and Robert Turnbull [22] demonstrates the advantages of Bayesian phylogenetics in this setting. This approach from biology combines the strengths of classical stemmatic analysis with robust and highly customizable models.

We prepared our collation data in a multi-step process. We converted our Excel-based CSV collation files to TEI XML format using a custom Python script. Then, using the `teiphy` Python

package [21], we converted the TEI XML collation to an XML input file for the BEAST 2 Bayesian phylogenetic software [6].

Bayesian phylogenetics addresses the real-world complexities of manuscript transmission in ways that simpler methods cannot. It parametrizes the lengths of branches so that more changes are expected over longer branches. Using clock models, it can also incorporate dating information for witnesses and their hypothetical ancestors, and branches can be assigned constant or variable rates of change to accommodate variance in scribal accuracy. These models are detailed in a biological setting in Felsenstein [14], and their adaptation for textual criticism is described in McCollum and Turnbull [22]. To explore their feasibility and power in reconstructing the transmission history of the *Martijn Trilogy*, we have implemented several phylogenetic models in our analysis. Our design choices are detailed in the appendix.

We ran multiple BEAST 2 analyses with these configurations. For the smaller dataset with 116 variation units and the larger dataset with 435 variation units, we ran one analysis with a uniform prior on readings at the root and one analysis with a non-uniform prior based on the judgments of Verdam and Leendertz [37], for a total of four analyses. We sampled 100,000,000 stemmatic hypotheses in each analysis. Of these samples, we logged one out of every 1,000, resulting in 100,000 logged states. We discarded the first half as "burn-in" to ensure that the most stable 50,000 states from each analysis were represented in the output.

To visualize the posterior distribution of trees from our samples, we used DensiTree plots. These plots draw the trees in the posterior distribution over one another so that more frequently attested topological features appear denser [5]. For reasons of space, our DensiTrees are rotated so that the root is at the left and later witnesses appear on the right. The time separating ancestors from their descendants is measured along the horizontal axis (despite the branches being drawn diagonally). Because various dates for reconstructed ancestors are sampled in the analysis, the early (left) ends of branches tend to spread over increasing date ranges. Similarly, for witnesses whose texts are fragmentary or potentially mixed, their varying placements in sampled stemmata are depicted as the convergence of branches from different directions. Each DensiTree also includes a "Placeholder" witness. This witness has no text, but it is included to overcome current limitations of how BEAST 2 determines the timespan of the reconstructed history; without a witness dated to the present, BEAST 2 would incorrectly treat the copying date of the latest extant witness as the present, and the shortened timeframe for transmission would fit poorly with estimated copying and loss rates for manuscripts.[7] DensiTrees for our four analyses are depicted in Figures 9–12.

All four DensiTrees generally confirm Verdam and Leendertz's reconstruction of the earliest history of transmission. One branch in the archetypal split generally consists of the three multi-quire manuscripts A, B, and C or of A and C alone. These classifications resemble with the classical stemma of Figure 1, which depicts A and C as direct descendants of the tradition's archetype. The main distinction is that our sampled stemmata betray some uncertainty about how A, B, and C relate to one another.

The DensiTrees disagree on some of the details of the rest of the tradition, but their common classifications are informative. The multi-quire manuscripts G and W are consistently classified as siblings as in Figure 1, though the analyses with smaller and larger datasets disagree as to their precise placement. In all four DensiTrees, the multi-quire manuscript O consistently appears to preserve a text from an early stage of the later tradition—a point that may support its classification as an extensively mixed witness in Figure 1—though it is unclear from the sampled stemmata which manuscripts are its closest relatives. In all four DensiTrees, the later part of the tradition in which G, W, and O were produced seems to reflect a rise in the *Martijn Trilogy*'s popularity starting in the fourteenth century. Our picture of this period is substantially improved by the inclusion of new manuscripts and previously disregarded fragments in our larger dataset: with the popularization of

[7] This distinction is important for the tree prior described in the appendix.

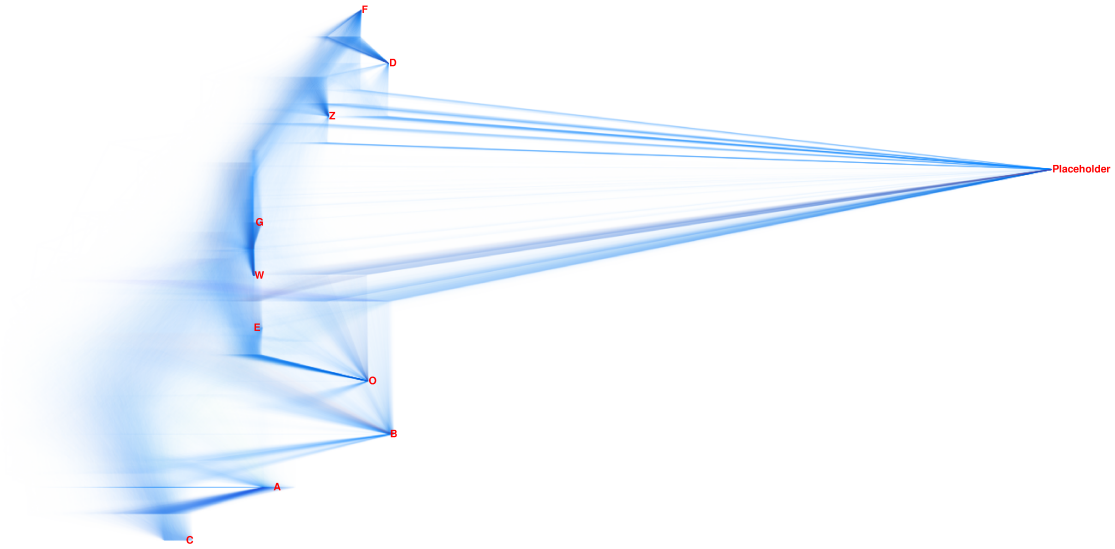


Figure 9: DensiTree plot for the smaller dataset, with prior probabilities assigned according to Verdam and Leendertz [37].

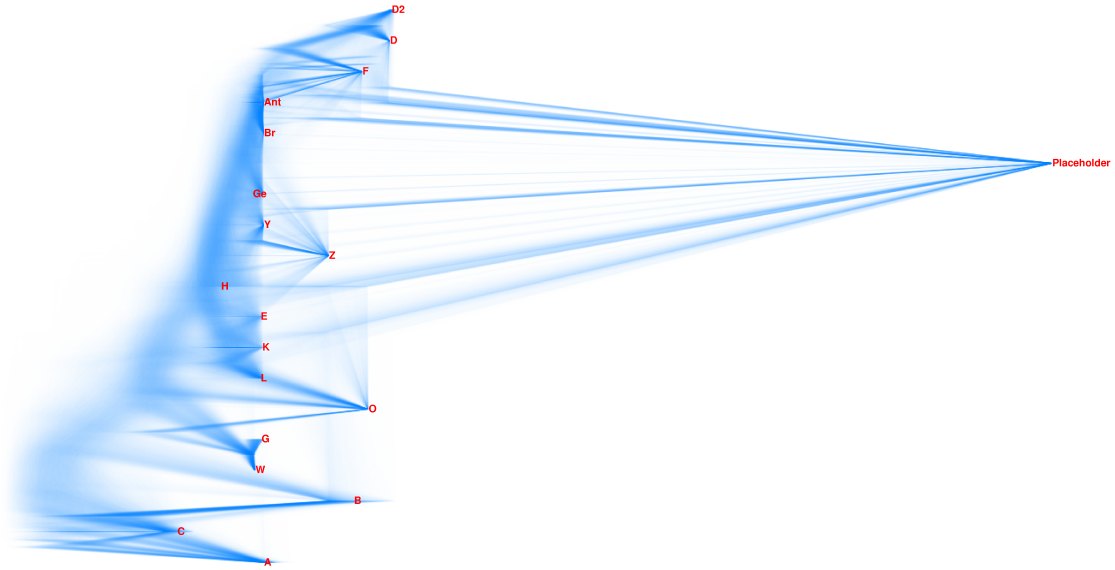


Figure 10: DensiTree plot for the larger dataset, with prior probabilities assigned according to Verdam and Leendertz [37].

the booklet format, single-quire *quinio* manuscripts (E, L, H, Y, and Z) appear to have flourished first, followed by single-quire *senio* manuscripts (Ant, Br, Ge). The text at the end of this period was evidently copied in another multi-quire manuscript (F) and served as a source for the printed books at the end of the tradition (D and D2).

Other paratextual and historical evidence supports and refines this general transmissional picture. The close relationship of G and W coheres with their known shared locality: as Erik Kwakkel discovered, both manuscripts were copied by the Speculum Scribe (named after his copy of the second part of the Middle Dutch *Spiegel historiael*) in the Carthusian monastery of Herne [20, p. 97].

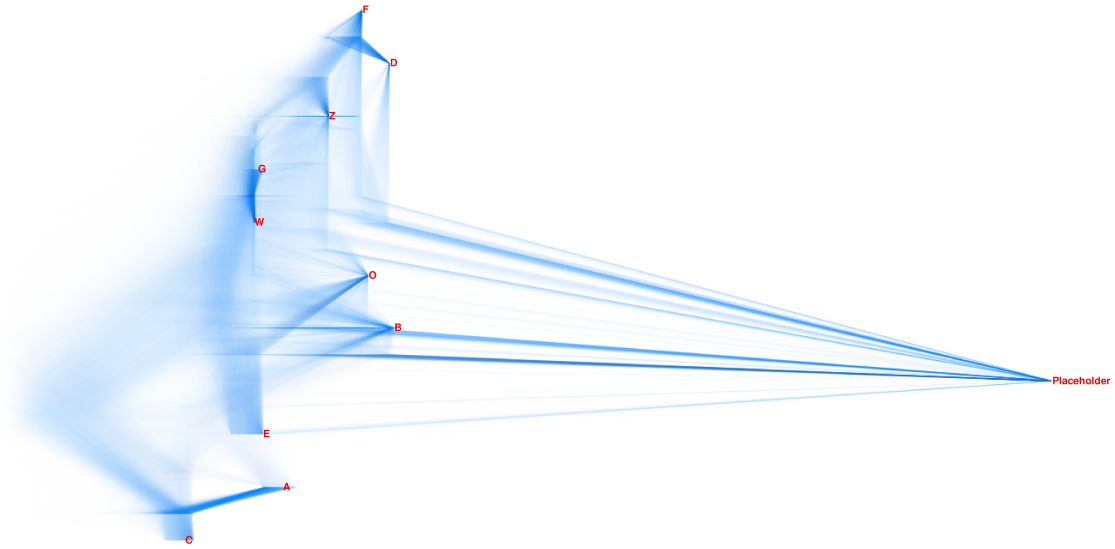


Figure 11: DensiTree plot for the smaller dataset, with equal prior probabilities for all variant readings.

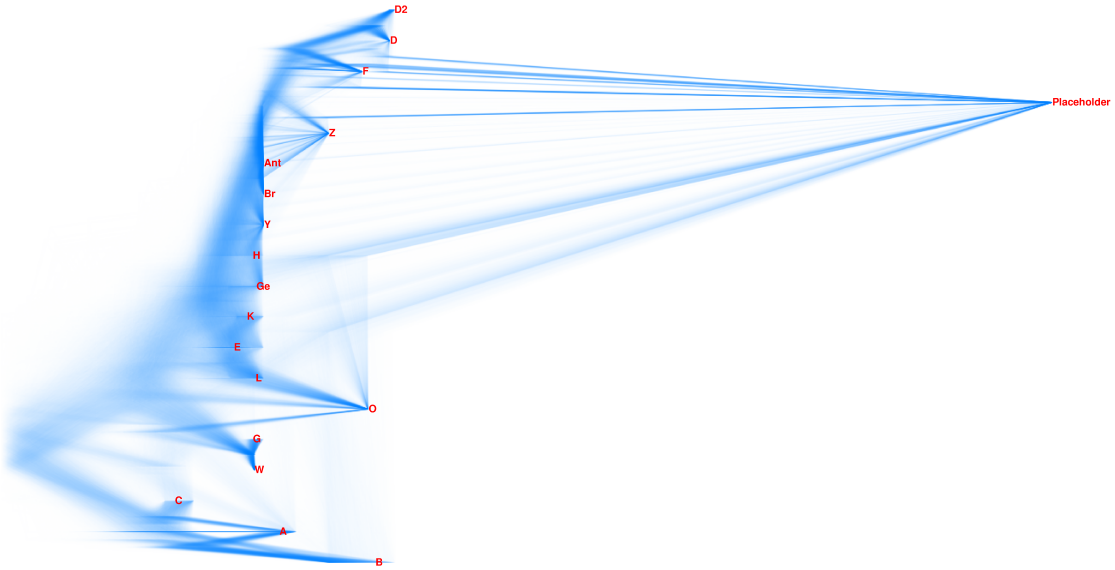


Figure 12: DensiTree plot for the larger dataset, with equal prior probabilities for all variant readings.

The single-quire *quinio* manuscript E and the manuscript K containing an excerpt of the *Martijn Trilogy* are both Flemish, and all of the later fragments in their clade are from Brabant. Their provenance and their place in the sampled stemmata suggest that the popular tradition transmitted in single-quire booklets emerged in Flanders and extended to Brabant. Given the popular text's increasing sphere of influence and the acceleration of scribal activity in the fourteenth century, it is unsurprising that the multi-quire manuscript F and the printed books D and D2 (all three of which are also from Brabant) feature many new variant readings.

The sampled stemmata show less resolution in the relationships among the witnesses within

these groups. Mixture of sources (which the current configuration of BEAST 2 cannot model) is a possible reason, but the more likely reason is that most of these witnesses are fragmentary: Ant is only extant in around 5% of variation units, Br in around 38%, E in around 20%, Ge in around 17%, H in around 41%, K in around 6%, L in around 52%, W (a copy of just the third part of the *Martijn Trilogy*, presumably because it features the most religious emphases) in around 18%, Y in around 22%, and Z in around 27%. Uncertainty about their contents produces uncertainty about their placement in the stemma. In this respect, these witnesses are not as informative to a reconstruction of transmission history as we would like, though as we have shown above, their physical and historical details confirm and refine this reconstruction at crucial points.

For the nearly complete manuscripts whose placement varies among the sampled stemmata, mixture remains a reasonable explanation. In all of our analyses, B is variously placed as a sibling of C or an early sibling of the later tradition (in agreement with the classical stemma of Figure 1 in the latter case). Under Kwakkel's proposed redating for this manuscript after the peak of the popular tradition (noted in Section 4.2), it is entirely possible that B mixes an early text like that of A and C with elements of the popular text. Similarly, O, an equally late multi-quire manuscript, is variously placed as a sibling to G and W, as a sibling to E, K, and L, and as a transitional witness between both groups in the sampled stemmata. Its text could be derived from a progressive mixture of sources from both of these branches (among others), as the stemma in Figure 1 suggests. As Bart Besamusca has noted, it seems that someone of high economic and intellectual status commissioned the copying of this manuscript for personal study [3, p. 17], so its copyist likely had both the means and the motivation to consult multiple manuscripts to produce an authoritative text. Finally, in the analysis with the larger dataset and non-uniform priors on root readings, just over 25% of sampled stemmata place F as a sibling with Ant, Br, or the common ancestor of Ant and Br rather than with the common ancestor of the printed books D and D2. This may indicate that F inherited some of its early readings by mixture. This would agree with the hypothesis of mixture suggested by Figure 1, though not with respect to the source of this mixture.

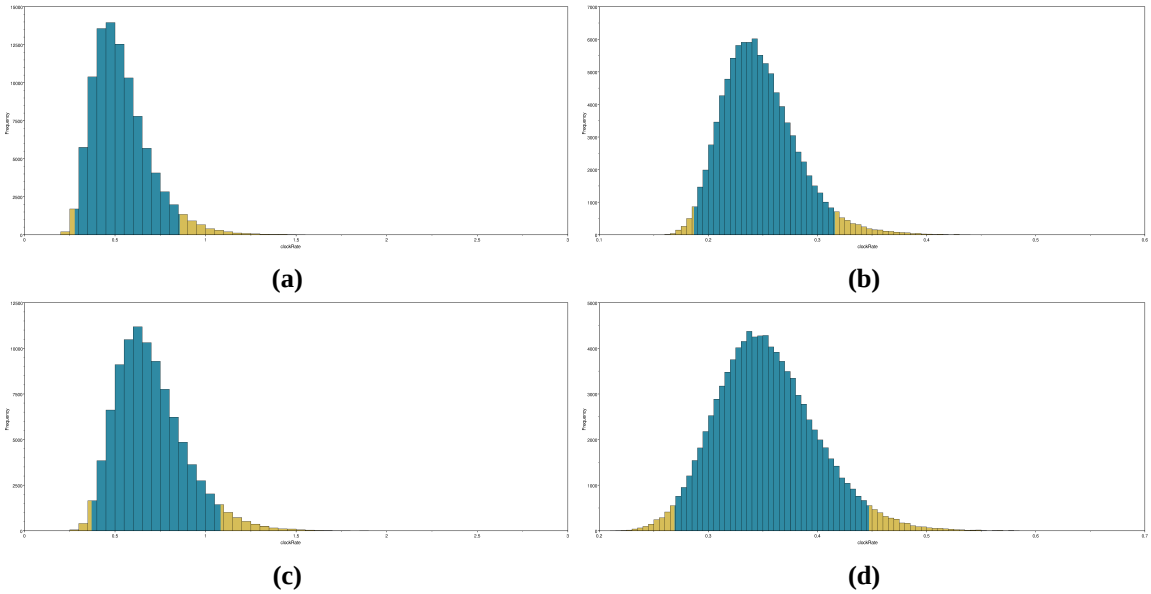
Even in the analyses where our priors on root readings favor many of A's readings in agreement with Verdam and Leendertz's, A is frequently grouped with C rather than set apart as a singularly authoritative witness. This can be explained by multiple factors. First, C is the earliest manuscript of the *Martijn Trilogy* currently known. Second, because lacunae are evaluated over all of their possible original states in Bayesian phylogenetics, C can still have an early place in a candidate stemma because it could have agreed with A in its missing portions. Verdam and Leendertz themselves expressed doubt over which of A and C to prefer, because the strength of C's age conflicted with the defect of its lack of the *Tweede Martijn* [37, p. xxxvi].

Their concerns and ultimate reversion to A bring the philosophical and methodological distinctions between old and new philology into sharp relief. Under the old philological approach, more complete witnesses are preferred because they better facilitate the production of a complete critical text. New philology, which does not prioritize this goal to the same extent, is not limited by lacunae in the same way. By virtue of how it handles uncertainty in data, Bayesian phylogenetics is well-suited to operate within the framework of new philology.

Another factor that explains C's consistent proximity to the root (in terms of its being few branches removed from the root) in the sampled stemmata is the strict clock model we assumed in our analyses. Under such a model, C is unlikely to have accumulated more scribal changes than A because it is an older witness and the branches leading to it cannot have a higher rate of change than those leading to A. The fact that A was copied during a time with a higher expected birth rate than the time when C was copied only makes matters worse. A promising direction for future research will be to experiment with an epoch-based uncorrelated random clock model, which can accommodate variation in rates of scribal freedom along different branches.

On a final related note, it is worth discussing how the Bayesian phylogenetic analyses have

estimated the expected number of textual changes made by the average scribe. Histograms of sampled values for our four analyses are depicted in Figure 13a.



(a) **(b)**

(c) **(d)**

Figure 13: Posterior distribution histogram of scribal change rate (expected changes per variation per copying event) for the smaller dataset, with prior probabilities assigned according to Verdam and Leendertz (a); the larger dataset, with prior probabilities assigned according to Verdam and Leendertz (b); the smaller dataset, with equal prior probabilities for all variant readings (c); and the larger dataset, with equal prior probabilities for all variant readings (d).

The means and standard deviations of these distributions are given in Table 1.

Analysis	Mean	Standard deviation
Smaller dataset, non-uniform root reading priors	0.535	0.160
Larger dataset, non-uniform root reading priors	0.247	0.034
Smaller dataset, uniform root reading priors	0.698	0.192
Larger dataset, uniform root reading priors	0.354	0.046

Table 1: Mean and standard deviation of estimated scribal change rate (expected changes per variation unit per copying event) by analysis.

The magnitude of the expected number of changes per variation unit is suspiciously high. Even restricting our attention to places where textual variation occurs, it seems unlikely that the average scribe would alter 25% of them (the lowest mean in Table 1), much less 70% of them (the highest mean in Table 1), at one time. These values are well above the simulation-based estimates scribal accuracy described by Spencer and Howe [34].

One potential explanation is that our assumed ranges for copying and loss rates were too strict. Under the current ranges, many of the branches in the sampled stemmata are so short that we would not expect even one copying event to occur over their duration. If we expanded our ranges of birth rates to accommodate much higher values, this would allow for lower scribal change rates. But substantially increasing the birth rate in the most active period of transmission would result in a survival rate well above the 7.5% estimate for medieval Dutch works given by Mike Kestemont et al. [19, p. 769], at least as long as we assume the death rates in line with the estimates of Eltjo Buringh [7, pp. 227–250]. If we allow for the reasonable possibility that manuscripts of the *Martijn Trilogy* were discarded more frequently during the peak period of the work's transmission, then higher birth and death rates could achieve the same survival rate and admit scribal change rates closer to their expected values.

Another solution is suggested by the observation that our stemmata have many short branches instead of a few long branches. Our current configuration with BEAST 2 does not allow branches to be collapsed together, so that an ancestor always has two descendants. But it is plausible that some hypothetical ancestors, especially those in the most active period of transmission, were copied in more than two surviving branches of transmission. It may even be that some of our extant manuscripts were ancestors to others.[8] Modeling such scenarios in our analysis would likely result in more conservative estimates of scribal change rates. This will require the adoption and development of Bayesian phylogenetic models better-suited to the idiosyncrasies of textual transmission.

5 Discussion

This research demonstrates both the potential and limitations of computational approaches to textual criticism. The inability of the traditional computational stemmatic method to reconstruct the manuscript tradition—even with expanded datasets and minimal thresholds—confirms the intuitions of new philologists who argue that many textual traditions cannot be reduced to simple genealogical trees. Meanwhile, our Bayesian phylogenetic analysis succeeded by taking textual data, paratextual details, and hidden parameters of the transmission process into account. The analysis confirmed and offered refinements to the history of the tradition that Verdam and Leendertz reconstructed in 1918 using traditional methods. As we have argued in the previous section, it sheds

[8] Traditionally, *codices descripti*, or surviving copies of surviving exemplars, are excluded from stemmata because they add no new information for reconstructing earlier states of the text. But in a Bayesian phylogenetic framework, they are still informative for the estimation of scribal change rates and should therefore be included in the analysis. This is yet another way that Bayesian phylogenetics complements the framework of new philology.

more light on later developments in the tradition, and it avoids the bias of Verdam and Leendertz regarding the earliest part of the tradition.

In this paper, we used a real world tradition with all its intricate complexities. To determine whether "old" or "new" philological methods are better suited, future work could include evaluating both types of methods on an equally complex *artificial* tradition. However, the strong agreement between our Bayesian analysis using uniform priors, totally independent of Verdam and Leendertz's judgments about original readings, and the traditional stemma from 1918, suggests that Bayesian phylogenetics captures many of the relationships they proposed between the manuscripts of the *Martijn Trilogy*.

Acknowledgments

The authors are grateful to the anonymous referees for their generous comments and recommendations, and to Elli Bleeker, Ben Nagy, and Joris van Zundert for looking over the submitted draft of this paper and offering constructive insights on it. We also thank Jean-Baptiste Camps for his feedback on the methodological aspect of the paper.

This article is the result of a research project by Sofie Moors, funded by the Research Foundation – Flanders (FWO), project number 1182725N, under the supervision of Remco Sleiderink and Mike Kestemont. Joey McCollum's contribution to this work was supported by an Australian Government Research Training Program (RTP) Scholarship (doi.org/10.82133/C42F-K220) and by the University of Melbourne's Research Computing Services and the Petascale Campus Initiative.

References

[1] Andrews, Tara L. "The Third Way: Philology and Critical Edition in the Digital Age". In: *The Journal of the European Society for Textual Scholarship 10* (2013), pp. 61–76. DOI: 10.1163/9789401209021_006.

[2] Andrews, Tara L. "What We Talk About When We Talk About Collation". In: *Advances in Digital Scholarly Editing*, ed. by Peter Boot, Anna Cappellotto, Wout Dillen, Franz Fischer, Andreas Mertgens Aodhán Kelly, Anna-Maria Sichani, Elena Spadini, and Dirk van Hulle. Leiden: Sidestone Press, 2017, pp. 231–234.

[3] Besamusca, Bart. "Boeken uit Brugge. Over de materiële context van de Middelnederlandse literatuur". Lecture slides. Universiteit Utrecht, 2017. URL: https://research-portal.uu.nl/en/publications/boeken-uit-brugge-over-de-materi%C3%ABle-context-van-de-middelnederlan.

[4] Bleeker, Elli. *Mapping Invention in Writing: Digital Infrastructure and the Role of the Genetic Editor*. PhD thesis. University of Antwerp, 2017.

[5] Bouckaert, Remco R. "DensiTree: Making Sense of Sets of Phylogenetic Trees". In: *Bioinformatics 26*, no. 10 (2010), pp. 1372–1373. DOI: 10.1093/bioinformatics/btq110.

[6] Bouckaert, Remco R. et al. "BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis". In: *PLoS Computational Biology 15*, no. 4 (2019), pp. 1–28. DOI: 10.1371/journal.pcbi.1006650.

[7] Buringh, Eltjo. *Medieval Manuscript Production in the Latin West: Explorations with a Global Database*. Global Economics History Series 6. Leiden: Brill, 2011.

[8] Camps, Jean-Baptiste and Cafiero, Florian. "Stemmatology: An R package for the computer-assisted analysis of textual traditions". In: *Corpus-Based Research in the Humanities CRH-2* (2018), pp. 65–74. URL: https://hal.science/hal-01695903.

[9] Camps, Jean-Baptiste and Randon-Furling, Julien. “Lost Manuscripts and Extinct Texts: A Dynamic Model of Cultural Transmission”. In: *Proceedings of the Computational Humanities Research Conference*, ed. by Folgert Karsdorp, Alie Lassche, and Kristoffer Nielbo. Vol. 3290. CEUR Workshop Proceedings. 2022, pp. 198–214.

[10] Cisne, John L. “How Science Survived: Medieval Manuscripts’ ‘Demography’ and Classic Texts’ Extinction”. In: *Science 307*, no. 5713 (2005), pp. 1305–1307. DOI: 10.1126/science.1104718.

[11] Dekker, Ronald Haentjens and Middell, Gregor. “CollateX”. Software tool. 2019. URL: https://collatex.net/.

[12] Drummond, Alexei J., Ho, Simon Y. W, Phillips, Matthew J., and Rambaut, Andrew. “Relaxed Phylogenetics and Dating with Confidence”. In: *PLoS Biology 4*, no. 5 (2008), e88. DOI: 10.1371/journal.pbio.0040088.

[13] Dutch Language Institute. “GaLAHaD (Version 1.2.5: hug-tdn-1400-1600): A Computational Language Model for Middle Dutch”. Online service. 2025. URL: https://hdl.handle.net/10032/tm-a3-b2.

[14] Felsenstein, Joseph. *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates, 2004.

[15] Guillaumin, Jean-Baptiste. “Criticisms of digital methods ”. In: *Handbook of Stemmatology: History, Methodology, Digital Approaches*, ed. by Philipp Roelli. Berlin/Boston: De Gruyter, 2020, pp. 339–356.

[16] Hautphenne, Sophie and Patch, Brendan. “Birth-and-Death Processes in Python: The BirDePy Package”. In: *Journal of Statistical Software 111*, no. 5 (2024), pp. 1–54. DOI: 10.18637/jss.v111.i05.

[17] Heikkilä, Tuomas. “Dealing with open textual traditions”. In: *Handbook of Stemmatology: History, Methodology, Digital Approaches*, ed. by Philipp Roelli. Berlin/Boston: De Gruyter, 2020, pp. 254–272.

[18] Kapitan, Katarzyna Anna. “A digital perspective on the role of a stemma in material-philological transmission studies”. In: *arXiv preprint arXiv:2505.06938* (2025).

[19] Kestemont, Mike, Karsdorp, Folgert, Bruijn, Elisabeth de, Driscoll, Matthew, Kapitan, Katarzyna A., Macháin, Pádraig Ó, Sawyer, Daniel, Sleiderink, Remco, and Chao, Anne. “Forgotten books: The application of unseen species models to the survival of culture”. In: *Science 375*, no. 6582 (2022), pp. 765–769. DOI: 10.1126/science.abl7655.

[20] Kwakkel, Erik. *Die dietsche boeke die ons toebehoeren: de kartuizers van Herne en de productie van Middelnederlandse handschriften in de regio Brussel (1350-1400)*. Miscellenea Neerlandica. Leuven: Peeters Publisher, 2002.

[21] McCollum, Joey and Turnbull, Robert. “teiphy: A Python Package for Converting TEI XML Collations to NEXUS and Other Formats”. In: *Journal of Open Source Software 7*, no. 80 (2022), p. 4879. DOI: 10.21105/joss.04879.

[22] McCollum, Joey and Turnbull, Robert. “Using Bayesian phylogenetics to infer manuscript transmission history”. In: *Digital Scholarship in the Humanities 39*, no. 1 (2024), pp. 258–279. DOI: 10.1093/llc/fqad089.

[23] Moors, Sofie. “Een bijzondere Brabantse tekstgetuige van de *Martijntrilogie* van Jacob van Maerlant. De fragmenten Lyon, Bibliothèque municipale, ms 6848”. In: *Queeste 29*, no. 1 (2022), pp. 36–65. DOI: 10.5117/QUE2022.1.002.MOOR.

[24] Moors, Sofie. “Jacob van Maerlant: leven en werk van de auteur van de *Rijmbijbel*”. In: *De Rijmbijbel van Jacob van Maerlant. Het oudste geïllustreerde handschrift in het Nederlands*, ed. by Bram Caers and Jan Pauwels. Amsterdam: Amsterdam University Press, 2023, pp. 6–12.

[25] Moors, Sofie. “The Success of Maerlant. The Transmission of Middle Dutch Text Traditions in Single-Quire Manuscripts”. In review.

[26] Moors, Sofie, Kestemont, Mike, and Sleiderink, Remco. “The *Martijn Trilogy* Manuscripts: An Open Dataset for Analyzing Scribal Variation”. In: *Research Data Journal for the Humanities and Social Sciences 9*, no. 1 (2024), pp. 1–21. DOI: 10.1163/24523666-bja10047.

[27] Nury, Elisa. *Automated Collation and Digital Editions: from Theory to Practice*. Classical studies. King’s College London, 2018.

[28] Palumbo, Giovanni. “Criticism and controversy”. In: *Handbook of Stemmatology: History, Methodology, Digital Approaches*, ed. by Philipp Roelli. Berlin/Boston: De Gruyter, 2020, pp. 88–109.

[29] Robinson, Peter. “Four rules for the application of phylogenetics in the analysis of textual traditions”. In: *Digital Scholarship in the Humanities 31*, no. 3 (2016), pp. 637–651. DOI: 10.1093/llc/fqv065.

[30] Philipp Roelli, edited by. *Handbook of Stemmatology: History, Methodology, Digital Approaches*. Berlin/Boston: De Gruyter, 2020.

[31] Sleiderink, Remco and Bosmans, Glenn. “Maerlant in Yvelines. Een nieuw fragment van de Martijns van Jacob van Maerlant (Montigny-le-Bretonneux, Archives départementales des Yvelines, 1F 180)”. In: *Queeste 26*, no. 1 (2019), pp. 70–97.

[32] Sleiderink, Remco, Moors, Sofie, and De Schepper, Marcus. “Bandwerk. Een nieuw fragment van Wapene Martijn van Jacob van Maerlant (Antwerpen, Universiteitsbibliotheek, Bijzondere Collecties, mag-p 64.20)”. In: *Queeste 27*, no. 1 (2020), pp. 43–62. DOI: 10.5117/QUE2020.1.002.SLEI.

[33] Spencer, Matthew, Davidson, Elizabeth A, Barbrook, Adrian C, and Howe, Christopher J. “Phylogenetics of artificial manuscripts”. In: *Journal of Theoretical Biology 227*, no. 4 (2004), pp. 503–511. DOI: 10.1016/j.jtbi.2003.11.022.

[34] Spencer, Matthew and Howe, Christopher J. “How Accurate Were Scribes? A Mathematical Model”. In: *Literary and Linguistic Computing 17*, no. 3 (2002), pp. 311–322. DOI: 10.1093/llc/17.3.311.

[35] Stadler, Tanja, Kühnert, Denise, Bonhoeffer, Sebastian, and Drummond, Alexei J. “Birth-death Skyline Plot Reveals Temporal Changes of Epidemic Spread in HIV and Hepatitis C Virus (HCV)”. In: *Proceedings of the National Academy of Sciences 110*, no. 1 (2013), pp. 228–233. DOI: 10.1073/pnas.1207965110.

[36] van Driel, Joost. *Meesters van het woord. Middelnederlandse schrijvers en hun kunst*. Hilversum: Uitgeverij Verloren, 2012.

[37] van Maerlant, Jacob. *Jacob van Maerlant’s Strophische gedichten. Nieuwe bewerking der uitgave van Franck en Verdam*, ed. by J. Verdam and P. Leendertz. Leiden: A.W. Sijthoff’s, 1918.

[38] van Zundert, Joris. “Data representation”. In: *Handbook of Stemmatology: History, Methodology, Digital Approaches*, ed. by Philipp Roelli. Berlin/Boston: De Gruyter, 2020, pp. 175–207.

[39] Warnar, Geert. “The discovery of the dialogue in Dutch medieval literature. A discourse for meditation and disputation”. In: *Meditatio – Refashioning the Self: Theory and Practice in Late Medieval and Early Modern Intellectual Culture*, ed. by Karl Enenkel and Walter Melion. Leiden: Brill, 2010, pp. 69–88.

[40] Zuckerkandl, Emile and Pauling, Linus. “Evolutionary Divergence and Convergence in Proteins”. In: *Evolving Genes and Proteins*, ed. by Vernon Bryson and Henry J. Vogel. New York, NY: Academic Press, 1965, pp. 97–166.

A Bayesian Phylogenetic Analysis Design

This appendix details our model design choices for the BEAST 2 Bayesian phylogenetic analysis described in Section 4.3. While this section is more technical in nature, we have endeavored to present our model choices in an accessible manner, and we justify these choices based on previous studies that may interest readers with a humanities background. In addition to providing a template that others can use for their own Bayesian phylogenetics for textual traditions,[9] this section demonstrates the power and customizability of BEAST 2 in analyzing textual traditions.

A.1 Tree Prior

For our first design choice, we employed a birth-death-sampling tree prior informed by expected demographic variations in manuscript copying and loss rates. As its name suggests, a tree prior assigns a probability to the topology of a candidate stemma before the textual data of the collation is taken into consideration. Our tree prior, as formulated by Tanja Stadler et al. [35], evaluates the probability that the hypothetical witnesses in the stemma would be copied at their assigned dates and that the extant witnesses would survive to the present. It treats sampling as a third type of event alongside birth/copying and death/loss to accommodate the likely scenario that some surviving remnants of the tradition remain undiscovered. We use the version of this tree prior with serial sampling through time rather than contemporary sampling at the present, because in a stemma, extant witnesses are dated to when they were copied rather than to the present (as their texts are assumed not to evolve continuously after they have already been copied). Because BEAST 2 assumes that the latest witness is dated to the present, we added a placeholder witness with an empty text to our collation and dated it to 2000 to represent a present point of observation. This ensured that the underlying transmission process would be modeled over a duration close to its true duration.

We allowed the Bayesian phylogenetic analysis to estimate the birth, death, and sampling rates within different ranges. We assumed that the birth rate (hereafter denoted λ) varied over four segments of the work's transmission: (1) an initial period of moderate activity from the work's composition to 1350 CE, with λ between 0.01 and 0.03 (corresponding to a given manuscript being copied every 33 to 100 years on average); (2) an active period from 1350 to 1450 CE corresponding to the popularization of the work, with λ between 0.3 and 0.1 (corresponding to a given manuscript being copied every 10 to 33 years); (3) a return to normal copying from 1450 to 1500 CE, with λ in the same range as the first period; and (4) the age of print from 1500 CE to the present, in which λ plummets to the negligible value between 0 and 0.005 (corresponding to a given manuscript being copied at least every 200 years on average). Throughout the tradition, we assumed that the death rate (hereafter denoted μ) would fall in the narrow range of 0.004 to 0.005, corresponding to an expected per-manuscript lifespan of between 200 and 250 years. For single-use manuscripts, an estimated average lifespan of 400 to 500 years (corresponding to μ between 0.002 and 0.0025) is

[9] Our TEI XML collation files, the BEAST 2 XML input files we used for our analyses, and the auxiliary Python scripts we note in this section are all available at `https://github.com/SofieMoors/martijncollation`.

offered by Buringh [7, pp. 227, 239, 246, 250], but the same source notes that in palimpsests, the time between the first and second use is 250 years ($\mu = 0.004$) on average [7, pp. 238, 246]. We arrived at these ranges by simulating a heterogeneous birth-death process over the time periods defined above with different choices of λ and μ, using a Python script we wrote based on the BirDePy simulation package [16] and the manuscript transmission simulation software described by Camps and Randon-Furling [9]. We based our upper and lower bounds for λ and μ on the simulations that resulted in mean manuscript survival rates close to 7.5%, close to the general survival rate estimated by Cisne [10] and matching the survival rate of medieval Dutch works estimated by Kestemont et al. [19, p. 769]. For the serial sampling rate (hereafter denoted ψ), we used different rates for the same time periods defined for the birth rates. In period (1), we assumed a low range of 0 to 0.01 (corresponding a given manuscript giving birth to a preserved copy at least every 100 years on average); for periods (2) and (3), we assumed a wider range from 0 to 0.03 (corresponding a given manuscript giving birth to a preserved copy at least every 33 years on average); and for period (4), we assumed a minimal range of 0 to 0.005 (corresponding a given manuscript giving birth to a preserved copy at least every 200 years on average). The decreasing upper bound of these ranges reflects the lower probability of survival for the earliest manuscripts.

A.2 Clock Model

Our second design choice was to use an epoch-based clock model. This type of clock model allows the mean mutation rate of branches to vary at different periods of time, similar to the rates in a birth-death tree prior. This is crucial for our purposes because in textual transmission, the rate of evolution is coupled with the birth rate of manuscripts. In biology, the process of genetic mutation is typically treated as independent of the birth-death process, because "birth" events actually correspond to events in which one population splits into two independently evolving subpopulations. In textual transmission, by contrast, "birth" events correspond to copying events, and it is precisely at these events, both at the nodes in the stemma and along its branches, that scribal changes are introduced. To know the expected number of textual changes along a branch, we must know the expected number of intermediate copying events along that branch and the expected number of changes introduced by each scribe. For a branch spanning a duration of time t, the expected number of intermediate copying events should be close to the expected number of "birth" events, which for a birth rate λ is just λt. For the average number of changes introduced by each scribe, the simulation-based estimates derived by Matthew Spencer and Christopher J. Howe [34] are a potential starting point. But since our collation currently does not account for segments of the text where variation is not observed (i.e., constant sites), and since the model assumed by Spencer and Howe assumes a pure-birth process in which no exemplars are lost, we felt it would be best to estimate the average number of changes per scribe in a Bayesian fashion. For our prior, we assumed that the mean number of changes per variation unit per scribe followed a log-normal distribution with a log-mean of -6 and a log-standard deviation of 1.

One further note is in order regarding our choice of clock model. For this study, we have chosen a strict clock model [40], which assumes that the rate of scribal change is the same for all branches in the same epoch. It is also possible to account for variation in the accuracy of different scribes using an uncorrelated random clock model [12]. But since the epoch-based variations of clock models have so far only been tested with the strict clock model, and since we can reasonably expect different scribes' tendencies to average out along branches, we have opted to use this model at this time.[10]

[10] This is noted in `https://www.beast2.org/2025/02/01/epoch-clock-models.html` (last accessed 8 July 2025).

A.3 Variation Unit Rate Heterogeneity

Our third design choice was to allow for different rates of scribal change in different variation units. As Spencer et al. [33] have shown, different segments of text tend to "evolve" at different rates just as different genes do in biology. This coheres with our intuition that some passages of text more easily invited mechanical error (e.g., if they began or ended with the same sequence as a previous or subsequent passage) or clarifying glosses (e.g., if their meaning was obscure to most readers) than others. At the start of the Bayesian phylogenetic analysis, we assigned all variation units equal relative rates of 1. Throughout the analysis, we sampled different values under the constraint that all the rates had to have an average of 1 (so as not to increase or decrease the expected number of changes per variation unit calculated for each branch).

A.4 Paratextual Variation Units

Our fourth design choice was to include a variation unit for details of the material production of manuscripts. For both datasets, we treated the following manuscript categories as "variant readings": (1) multi-quire manuscripts, (2) single-quire *senio* (six bifolia) manuscripts, (3) single-quire *quinio* (five bifolia) manuscripts, and (4) printed books. Modeling changes between these categories with the same substitution model we use for variant readings allows us to test how influential an exemplar's design was on a copy's design and whether or not manuscripts produced in the same way tended to align textually, as well.

A.5 Root Reading Priors

Finally, we incorporated prior probabilities for readings at the root of the stemma. For each of the two datasets described in Section 4.2, we ran two analyses: one with equal prior probabilities for all readings in each variation unit, and one with prior probabilities informed by the editorial judgments of Verdam and Leendertz [37]. In the latter case, we assigned relative odds for one reading's originality over that of another reading, effectively quantifying Verdam and Leendertz's scholarly expertise about the intrinsic authority of readings, relevant scribal behavior, and textual history. Readings printed in their edition without qualification were assigned odds of 100 over all of their alternatives. Readings whose apparatus entries were marked with "=" were regarded as equally likely, and these readings were assigned odds of 100 over any remaining readings in their variation unit. Where a reading was marked in the apparatus with "=?" relative to another reading, we assigned the first reading odds of $\sqrt{10} \approx 3.162$ over the second reading to quantify a slight preference for the first reading, and we assigned the second reading odds of 100 over any remaining readings. As McCollum and Turnbull [22] argue, the assignment of prior probabilities to authorial readings is not only defensible but beneficial, because Bayesian phylogenetics can manage and leverage the varying degrees of certainty inherent in human judgments, and it can use overall trends of prior probabilities to determine which witnesses are most likely to be close to the root of the stemma.