

# Quantitative Stemmataology and OpenRecensio: A Test Case on the Bible Historiale's Book of Exodus

Valentina Modolo<sup>1,2</sup>

P

<sup>1</sup> École nationale des chartes - PSL

<sup>2</sup> Scuola Normale Superiore di Pisa

## 1 Introduction

The translations of the Bible occupy a central position within the medieval European literary heritage. In recent decades, philological and codicological research has produced major contributions to the study of their transmission and documentary history. Despite these advances, comprehensive critical editions are still lacking. The scale of the corpus and the number of witnesses (often several dozen for a single book) make traditional *recensio* demanding and the fundamental reference for the field remains Berger's pioneering survey [1].

The case of the *Bible historiale* (BH) is emblematic. Conceived as a composite work and progressively intertwined with the *Bible du XIIIe siècle*, its textual stratigraphy and redactional phases complicate attempts at a stable reconstruction. Although partial editions exist for individual manuscripts, Michel's edition of the *Genèse* remains the only large-scale intervention so far [10], leaving open several stemmatological questions. Moreover, the BH exemplifies what Pasquali defined as an "open recensio" [13], where contamination, interpolation, and rewriting render purely mechanical methods insufficient.

This complexity renders computational approaches particularly relevant. Computer-assisted stemmatology has a well-established scholarly history [14, ch. 5], as demonstrated by recent applications to stratified and contaminated traditions [3]. This contribution applies these approaches to the transmission of *Exodus* in the BH, beginning with a subset of the primitive witnesses. The long-term aim is to extend the analysis to the seventy-four manuscripts transmitting *Exodus* and assess to what extent computational modelling can support stemmatic reconstruction in biblical traditions.

## 2 Corpus, Methodology, and Encoding

The corpus comprises eight witnesses: five *manuscripts primitifs* of the first redaction (Maz312, BnF 152, BR 987, JenEl 9596, Bein 129), two witnesses of the 1297 *Seconde Édition* (BnF 155, BL19DIII), and Arsenal 5059 (1317) as an early representative of the *Bible historiale complétée* (see 1). The *Historia Scholastica* [11] and the *Vulgata* [2] were included as external comparanda. Manuscript images were processed in eScriptorium [8] using CATMuS Medieval [Large] (v1.0.0 [5]; CER 98.5%), followed by manual post-correction. The output is natively exportable to ALTO-XML, ensuring interoperability with TEI-based workflows. Relevant variant locations were subsequently encoded in TEI P5 following Camps and Cafiero [4], preserving philological control before the inferential stage — a step that general-purpose software such as PHYLIP does not address intrinsically [14, ch. 5.5]. Each locus is classified by transformation type (lexical, syntactic, semantically driven replacement, macro-structural); purely graphic alternations were excluded. Variant locations were then analysed with Maximum Parsimony (PS [12]), Neighbor-Joining (NJ; [15]), and Unweighted Neighbor-Joining (UNJ [7]).

Table 1: Composition of the corpus. Dates follow Komada [9].

| Sigla            | Manuscript                           | Redaction                 | Date      |
|------------------|--------------------------------------|---------------------------|-----------|
| <b>Maz312</b>    | Paris, Bibl. Mazarine, 312           | <i>Première Éd.</i> ms.   | 1330-1360 |
| <b>BnF152</b>    | Paris, BnF, fr. 152                  | <i>Première Éd.</i>       | 1347      |
| <b>BR987</b>     | Bruxelles, KBR, ms. II 987           | <i>Première Éd.</i>       | 1350      |
| <b>JenEl9596</b> | Jena, ThÜLB, El. f. 95-96            | <i>Première Éd.</i>       | 1465-1473 |
| <b>Bein129</b>   |                                      | <i>Première Éd.</i>       | 1460 c.   |
| <b>BnF155</b>    | Paris, BnF, fr. 155                  | <i>Seconde Éd.</i>        | 1310-1315 |
| <b>BL19DIII</b>  | London, British Lib., Royal 19 D III | <i>Bibles à prologues</i> | 1411      |
| <b>Ars5059</b>   | Paris, Bibl. de l'Arsenal, ms. 5059  | <i>BH complétée</i>       | 1317      |

For instance, at Col. 1155A (*Tulit Moyses scriptum in lamina aurea nomen Domini tetragrammaton*), the witnesses diverge in the rendering of *lamina*: the *Historia Scholastica* reads *lamina*; BnF 155 and Ars5059 read *tombe*; Maz312, BnF152, BR987, JenEl9596, Bein129, and BL19DIII read *lame*. This semantic lexical variant is encoded in TEI P5 as a nested parallel-segmentation structure:

```
<p xml:id="Col.1155A" part="NV">
  non mie en
  <app type="semlex:nonsens" xml:id="a_511">
    <rdg wit="#Maz312 #BnF152 #BR987 #JenEl9596 #Bein129 #BL19DIII #hs">
      <app type="translation" xml:id="a_512">
        <rdg wit="#hs">lamina</rdg>
        <rdg wit="#Maz312 #BnF152 #BR987 #JenEl9596 #Bein129 #BL19DIII">lame</rdg>
      </app>
    </rdg>
    <rdg wit="#BnF155 #Ars5059">tombe</rdg>
  </app>
  la ou il fust
</p>
```

This encoding imposes no base text, is compatible with an open *recensio*, and enables decomposition into minimal operations essential for quantitative stemmatology.

### 3 Results and Discussion

All three methods converge on four stable clusters:

{Maz312, BnF152, Bein129}, {BR987, JenEl9596}, {BnF155, Ars5059}, {BL19DIII}.

The contrast between the primitive cluster of the *Première* and the revised cluster of the *Seconde Édition* is clear. BnF155 is consistently followed by Ars5059, corresponding to the  $\beta$ - $\gamma$  families;

BL19DIII matches the  $\alpha$  branch of the *Bibles à prologues*.

The methods diverge in the internal configuration of the *Première Édition*. UNJ most closely reflects the qualitative analysis of the tradition: Maz312 and Bein129 form a subgroup opposed to BR987–JenEl9596, with BnF152 slightly more external. NJ is close, but separates Bein129 from Maz312–BnF152. MP, in contrast, clusters JenEl9596 with Bein129 against the nucleus made up of Maz312, BnF152, and BR987. Maz312 and BR987 are regarded as the most conservative witnesses of the primitive recension, and their position will require further investigation on the broader corpus. Bootstrap proportions [6] confirm the stability of the main clusters.

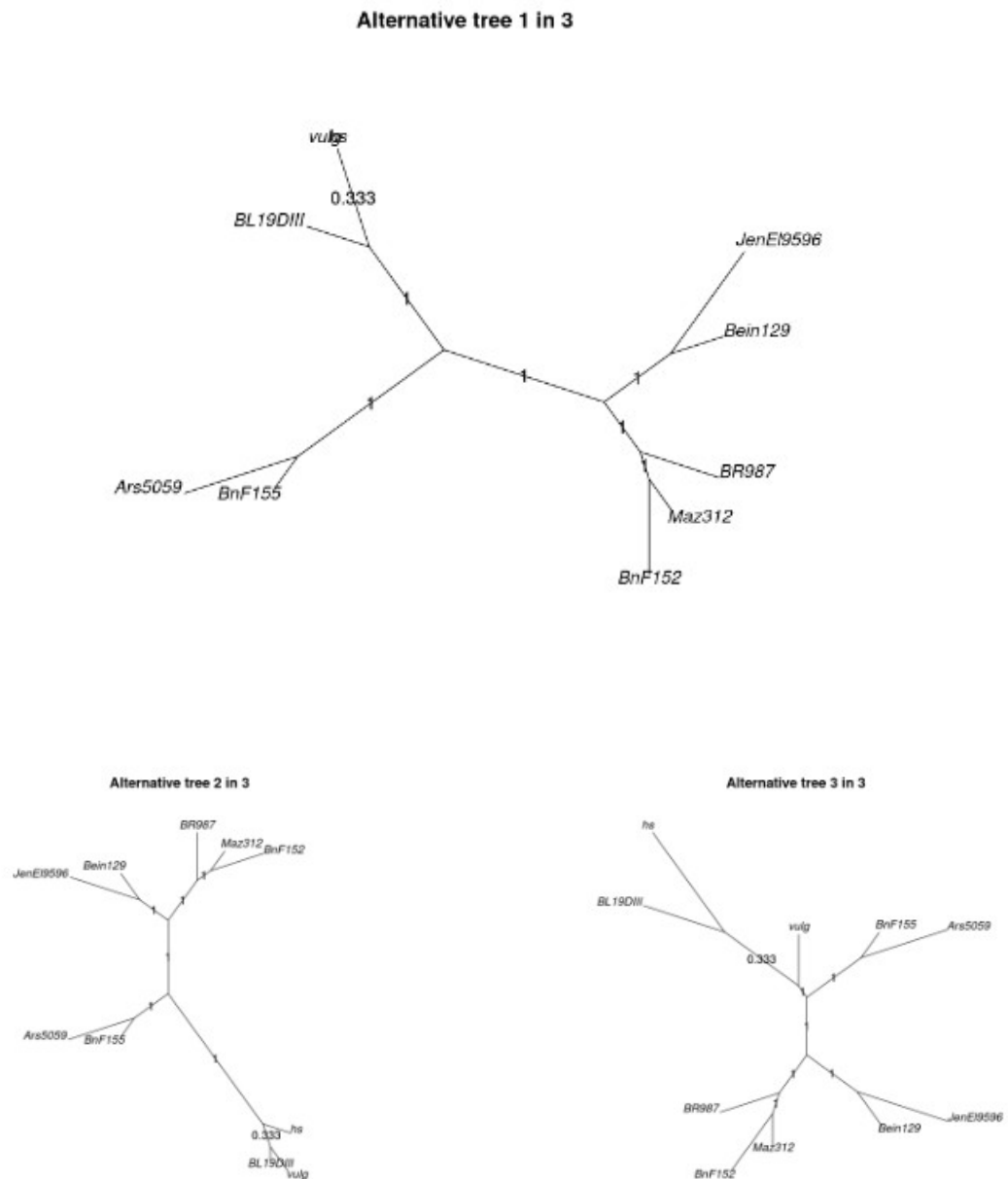


Figure 1: Three equally parsimonious MP trees. Edge values = bootstrap proportions [6]. Only the placement of **hs**, **BL19DIII**, and **vulg** differs.

UNJ proves the most promising method, recovering all major groupings supported by the qualitative analysis of the tradition. The affinity between JenEl9596 and Bein129 — the two latest copies of

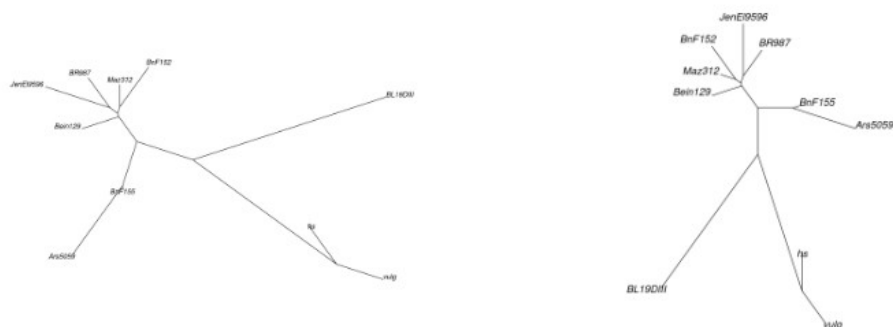


Figure 2: Unrooted NJ (left) and UNJ (right) trees based on ML distances.

the *Première Édition*, captured together by MP and partially by NJ — warrants further investigation, and may find compatibility with Michel’s identification of both witnesses as contaminated from the same source in the Genèse [10]. Future work will extend the sample to 45 witnesses.

### 3 Results and Discussion

The source data and code are available at

<https://github.com/bh-anon-repo/stemmatology-exodus>.

### References

- [1] Berger, S. (1884). *La Bible française au Moyen Âge*. Vol. 1. Imprimerie nationale.
- [2] *Biblia Sacra Vulgata*. (1592). Editio Romana. Editio Clementina promulgated by Clement VIII. Roma: Typographia Apostolica Vaticana.
- [3] Buzzoni, M. et al. (2016). Open versus closed recensions (Pasquali): Pros and Cons of Some Methods for Computer-Assisted Stemmatology. *Digital Scholarship in the Humanities* 31(3), 652–669.
- [4] Camps, J.-C. and Cafiero, F. (2018). Stemmatology: an R package for the computer- assisted analysis of textual traditions. *Proceedings of the Second Workshop on Corpus-Based Research in the Humanities (CRH-2)*, 65–74.
- [5] Clérice, T. et al. (2024). CATMuS Medieval: A Multilingual Large-Scale Cross- Century dataset in Latin Script for Handwritten Text Recognition and Beyond. *2024 International Conference on Document Analysis and Recognition (ICDAR)*, 174–194.
- [6] Felsenstein, J. (1985) Confidence limits on Phylogenies: An Approach using the Bootstrap. *Evolution* 39(4), 783–791.
- [7] Gascuel, O. (1996). Concerning the NJ algorithm and its unweighted version, UNJ. *Mathematical Hierarchies and Biology* 37, 149–170.
- [8] Kiessling, B. et al. (2019). eScriptorium: An Open Source Platform for Historical Document Analysis. *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, 19–19. doi: 10.1109/ICDARW.2019.10032
- [9] Komada, Akiko. (2004). *Les Illustrations de la Bible historique: les manuscrits réalisés dans le Nord*. PhD thesis. Paris 4.
- [10] Michel, B. (2004). *La “Bible historique” de Guiart des Moulins : édition critique de la Genèse d’après le manuscrit Bruxelles II 1987*. Supervisor: Jean-Marie Fritz. PhD thesis. Université de Dijon.
- [11] Migne, J. P. et al. (1855). *Adami Scoti Canonici regularis ordinis praemonstratensis Opera omnia... accedunt Magistri Petri Comestoris Historia Scholastica, Sermones... Godefridi Viterbiensis*

*Chronicon...* Vol. 198.

[12] Nixon, K. C. (1999). The Parsimony Ratchet, A New Method for Rapid Parsimony Analysis. *Cladistics* 15(4), 407–414.

[13] Pasquali, G. (1952). *Storia della tradizione e critica del testo*. Felice Le Monnier, 1952.

[14] Roelli, P. (2020). *Handbook of Stemmatology: History, Methodology, Digital Approaches*. De Gruyter.

[15] Saitou, N. and Nei, M. (1987). The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution* 4(4), 406–425.