

Quantitative Stemmatology and Open *Recensio*: A Test Case on the Bible historiale's Book of Exodus

This paper examines the contribution of quantitative stemmatology to a medieval tradition marked by a Pasqualian open *recensio*. Eight witnesses of the *Bible historiale* transmitting *Exodus* were encoded in TEI with an apparatus recording lexical, syntactic, morphological and macro-structural variation. From this encoding, genealogical variant locations were extracted and modelled in R. The experiment evaluates how computational methods interpret a stratified and contaminated tradition and compares the results with Bénédicte Michel's qualitative hypotheses for the *Genèse*. Preliminary analyses reveal stable clusters crossing redactional phases and show where modelling supports philological *iudicium* and where the tradition resists mechanical reconstruction.

Introduction

The translations of the Bible occupy a central position within the medieval French literary heritage. In recent decades, philological and codicological research has produced major contributions to the study of their transmission and documentary history (Berger, 1884, Burgio, 2004, Lagomarsini, 2024, Nobel, 2006, Revol, 2019, Sneddon, 1978). Despite these advances, comprehensive critical editions are still lacking, especially in the Romance tradition. The scale of the corpus and the number of witnesses (often several dozen for a single book) make traditional *recensio* demanding.

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 (Burgio et al., 1995, Cattai, 1998, Roy, 1979) Michel's edition of the *Genèse* remains the only large-scale intervention so far, and it leaves open several stemmatological questions (Michel, 2004). Moreover, the BH exemplifies what Pasquali defined as an "open *recensio*" (Pasquali, 1952), where contamination, interpolation and rewriting render purely mechanical methods insufficient.

This complexity renders computational approaches particularly relevant. Recent research has shown the value of quantitative stemmatology for analysing stratified and contaminated traditions (Buzzoni et al., 2016).

This article applies these approaches to the transmission of *Exodus* in the *Bible historiale*, 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 Romance biblical traditions. It extracts a table of variant locations from a TEI encoded collation, using the R Stemmatology package (Camps and Cafiero, 2018) and then proceeds to compare various stemmatological methods, especially based on traditional neo-Lachmannian approaches, to classic algorithms borrowed from phylogenetics, such as Maximum Parsimony and Neighbor Joining (Roelli, 2020).

Material and Methods

Corpus

To address the stemmatic analysis of the *Bible historiale*, we selected a corpus of eight witnesses representing the two principal editorial stages of the text. The corpus includes five manuscripts traditionally described as *manuscripts primitifs* of the first redaction (Maz312, BnF 152, BR 987, JenEl 9596, Bein 129), and two primitive witnesses of the 1297 *Seconde Édition* (BnF 155 and BL19DIII). Arsenal 5059 (1317) was added as an early representative of the *Bible historiale complétée* (Appendix A).

To assemble the corpus, OCR and post-correction were performed using eScriptorium (Kiessling et al., 2019) and the CatMuS model (Clérice et al., 2024). Two additional sources, the *Historia Scholastica* (Migne et al., 1855) and the *Vulgata* (vul, 1592), were included as external *comparanda*.

Variant encoding and data model

Our encoding strategy follows Camps and Cafiero (2018) for identifying genealogical variant locations. Each locus is classified according to the type of transformation – lexical, syntactic, semantically driven replacement, or macro-structural modification – while purely graphic alternations and abbreviation-based variation were excluded, as they do not carry genealogical signal.

Consider, for example, the locus at Col. 1155A in the *Historia Scholastica* (*Tulit Moyses scriptum in lamina aurea nomen Domini tetragrammaton*), where the French witnesses diverge in the rendering of *lamina*:

- **hs:** *lamina*
- **PA:** *tombe*
- **MBHKJP1:** *lame*

These readings reflect distinct interpretative and translational choices, and the locus is classified as a semantic lexical variant. The TEI apparatus represents this as a nested variant structure using parallel-segmentation:

```
<p xml:id="Col. 1155A" part="NV">
  non mie en
  <app type="semlex:nonsens" xml:id="a_511">
    <rdg wit="#K #M #B #H #J #P1 #hs">
      <app type="translation" xml:id="a_512">
        <rdg wit="#hs">lamina</rdg>
        <rdg wit="#K #M #B #H #J">lame</rdg>
      </app>
    </rdg>
  </app>
```

```

    </rdg>
    <rdg wit="#P #A">tombe</rdg>
  </app>
  la ou il fust
</p>

<cit>
  <quote xml:lang="la">Tulit Moyses scriptum in lamina aurea nomen Domini
    tetragrammaton</quote>
</cit>

```

This encoding imposes no base text and remains compatible with an open *recensio*. It also enables decomposition into minimal operations (replacement, insertion, deletion), essential for quantitative stemmatology.

Data Extraction and Representation

All variant locations encoded in the *Exodus* apparatus were analysed with three computer-assisted stemmatological methods: Maximum Parsimony (PS; Nixon (1999)), Neighbor-Joining (NJ; Saitou and Nei (1987)), and Unweighted Neighbor-Joining (UNJ; Gascuel (1997)).

Maximum Parsimony yields three nearly identical trees, all converging on four stable clusters:

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

The contrast between the primitive cluster of the *Première Édition* (*Maz312*, *Bein129*, *BnF152*, *BR987*, *JenEl9596*) and the revised cluster of the *Seconde Édition* (*BnF155*) is clear. *BnF155*—the earliest witness of the Second Edition (ca. 1310–1315)—is consistently followed by *Ars5059*, in line with Michel’s reconstruction of the β – γ families. *BL19DIII* forms its own group, matching its placement in the *alpha* branch of the *Bibles à prologues*.

Where the results differ from Michel is in the internal configuration of the *Première Édition*. Michel associates *BR987* with *JenEl9596* and *Bein129* with *Maz312*–*BnF152*; here, however, *JenEl9596* clusters with *Bein129*, while *BR987* aligns with *Maz312*–*BnF152*. This likely reflects quantitative affinity among later copies (*JenEl9596*, *Bein129*) rather than a genuine genealogical shift.

NJ and UNJ reproduce the same four clusters, but diverge in the internal organisation of the primitive branch. UNJ fully recovers Michel’s structure: *Maz312* and *Bein129* form a subgroup opposed to *BR987*–*JenEl9596*, with *BnF152* only slightly more external than *Maz312* and *Bein129*. NJ is very close to this configuration but separates *Bein129* from the *Maz312*–*BnF152* pair, introducing a different internal distance while preserving all major groupings. UNJ therefore offers the closest quantitative counterpart to Michel’s stemmatic model.

The UNJ topology shows the expected affiliations among the witnesses but, as an unrooted tree, offers no indication about the root of the tradition. Independently of these results, *Maz312* and *BR987* are regarded in the philological literature as the most conservative witnesses of the primitive recension. Their position will require further investigation on the broader *Exodus* corpus.

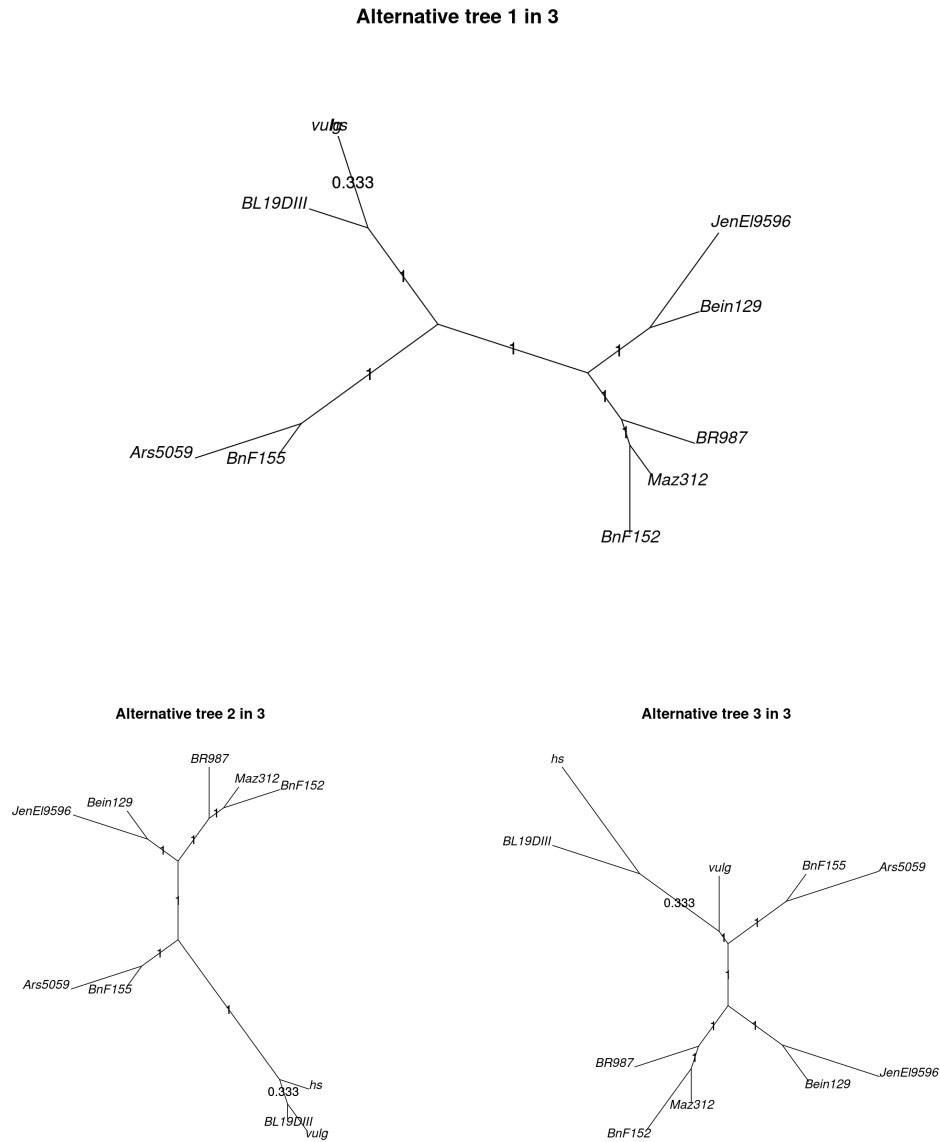


Figure 1: Three equally parsimonious Maximum Parsimony unrooted trees for the *Exodus* corpus; numerical values on the edges represent Felsenstein's bootstrap proportions (Felsenstein, 1985). As demonstrated by these values, the only difference concerns the exact relationship of **hs**, **BL19DIII**, and **vulg**

Results

The comparison between alternative topologies shows that not all computational approaches yield an equally reliable reconstruction. Maximum Parsimony confirms the main split but not the internal structure, while NJ approximates the expected subgroupings with some dispersion. UNJ proves the most promising: it clearly recovers the groups identified by Michel for the *Genèse*, including the two primitive subgroups (Maz312–BnF152–Bein129 and BR987–JenEl9596) and the revised pair (BnF155–Ars5059). In this layout, Maz312 and BR987 emerge as the most conservative witnesses, indicating the likely zone for rooting the stemma. Overall, quantitative mod-

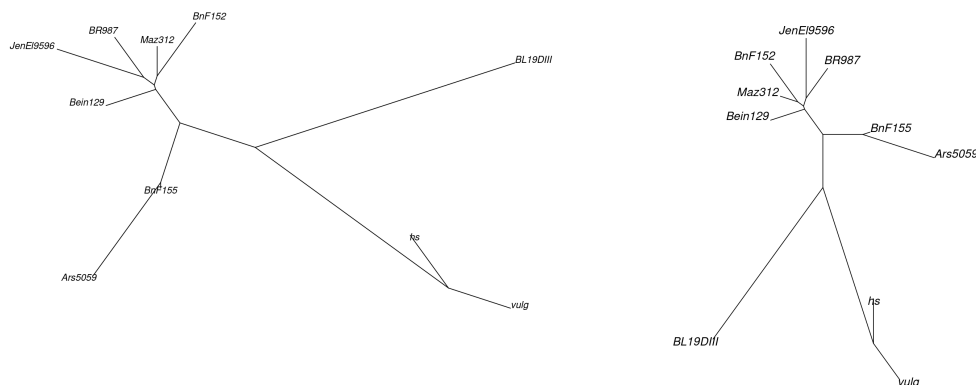


Figure 2: Unrooted NJ (left) and UNJ (right) trees based on ML distances.

elling captures the major structures of the *Bible historiale*, though finer interpretation remains dependent on philological *iudicium*.

Discussion

Our analyses show that quantitative procedures provide reliable indicators for modelling medieval French biblical traditions. In this case study, the results refine our understanding of the *Bible historiale* and broadly support Michel’s intuition. Future work will extend the sample (45 witnesses) and incorporate additional phylogenetic methods, including Bayesian approaches, with the aim of implementing the expanded corpus for the DH Benelux 2026 Conference.

Data and code availability

The source data and code used in this paper is available at <https://github.com/bh-anon-repo/stemmatology-exodus>.

Acknowledgements

Removed for anonymity.

References

- Biblia Sacra Vulgata*. Typographia Apostolica Vaticana, Roma, 1592. Editio Clementina promulgated by Clement VIII.
- Samuel Berger. *La Bible française au moyen âge*, volume 1. Imprimerie nationale, 1884.
- Eugenio Burgio. I volgarizzamenti oitanici della bibbia nel xiii secolo (un bilancio sullo stato delle ricerche). *Critica del testo*. VII, 1, 2004, pages 1000–1040, 2004.

- Eugenio Burgio et al. Ricerche sulla tradizione manoscritta delle vite antico-francesi di giuda e di pilato. i. le redazioni in prosa della vita di pilato. *Annali di Ca' Foscari*, 34:97–137, 1995.
- Marina Buzzoni, Eugenio Burgio, Martina Modena, and Samuela Simion. Open versus closed recensions (Pasquali): Pros and cons of some methods for computer-assisted stemmatology. *Digital Scholarship in the Humanities*, 31(3):652–669, 2016.
- Jean-Baptiste Camps and Florian Cafiero. Stemmatology: an R package for the computer-assisted analysis of textual traditions. In *Corpus-Based Research in the Humanities CRH-2*, pages 65–74, 2018.
- Tiziana Cattai. Les “Évangiles” dans la Bible historique de Guyart des Moulins. édition critique du ms. Paris, BnF, fr. 155. Master’s thesis, Università Ca’ Foscari Venezia, 1998. Supervisors: M. Eusebi and E. Burgio.
- Thibault Clérice, Ariane Pinche, Malamatenia Vlachou-Efstathiou, Alix Chagué, Jean-Baptiste Camps, Matthias Gille Levenson, Olivier Brisville-Fertin, Federico Boschetti, Franz Fischer, Michael Gervers, et al. Catmus medieval: A multilingual large-scale cross-century dataset in latin script for handwritten text recognition and beyond. In *International Conference on Document Analysis and Recognition*, pages 174–194. Springer, 2024.
- Joseph Felsenstein. Confidence limits on phylogenies: an approach using the bootstrap. *evolution*, 39(4):783–791, 1985.
- Eléonore Fournié. Les manuscrits de la bible historique. présentation et catalogue raisonné d’une œuvre médiévale. *L’Atelier du Centre de recherches historiques. Revue électronique du CRH*, (03.2), 2009.
- Olivier Gascuel. Concerning the nj algorithm and its unweighted version, UNJ. In *Mathematical Hierarchies and Biology: DIMACS Workshop, November 13-15, 1996*, volume 37, page 149. American Mathematical Soc., 1997.
- Benjamin Kiessling, Robin Tissot, Peter Stokes, and Daniel Stökl Ben Ezra. escriptorium: an open source platform for historical document analysis. In *2019 international conference on document analysis and recognition workshops (icdarw)*, volume 2, pages 19–19. IEEE, 2019.
- Akiko Komada. *Les illustrations de la Bible historique: les manuscrits réalisés dans le Nord*. PhD thesis, Paris 4, 2000.
- Claudio Lagomarsini. *La Bible française du XIIIe siècle. Édition critique des livres de Ruth, Judith et Esther*. Number 664 in Textes littéraires français. Droz, Genève, 2024.
- Marina Michel. *La “Bible historique” de Guiart des Moulins : édition critique de la Genèse d’après le manuscrit Bruxelles II 1987*. PhD thesis, Université de Dijon, 2004. Supervisor: Jean-Marie Fritz.
- Jacques Paul Migne et al. *Adami Scoti Canonici regularis ordinis praemonstratensis Opera omnia... accedunt Magistri Petri Comestoris Historia Scholastica, Sermones... Godefridi Viterbiensis Chronicon...*, volume 198. Migne, 1855.
- Kevin C Nixon. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics*, 15(4):407–414, 1999.

- Pierre Nobel. *La Bible d'Acre: Genèse et Exode: édition critique d'après les manuscrits BNF nouv. acq. fr. 1404 et Arsenal 5211*. Presses Univ. Franche-Comté, 2006.
- Giorgio Pasquali. *Storia della tradizione e critica del testo*. Felice Le Monnier, 1952.
- Thierry Revol. Traduire un texte sacré. analyses syntaxiques de la bible anglo-normande. *Linx. Revue des linguistes de l'université Paris X Nanterre*, (78), 2019.
- Philipp Roelli. *Handbook of stemmatology: History, methodology, digital approaches*. De Gruyter, 2020.
- P. Roy. *An Edition of the Gospel Harmony in British Museum: Royal 19 D III*. PhD thesis, University of Ottawa, 1979. Supervisor: R. St. Jacques.
- Naruya Saitou and Masatoshi Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, 4(4):406–425, 1987.
- Clive Roderick Sneddon. *A critical edition of the four gospels in the thirteenth-century Old French translation of the Bible*. PhD thesis, University of Oxford, 1978.

References

A. Manuscripts of the *Bible historiale* considered in this study

For a full description of these manuscripts, see (Berger, 1884, Fournié, 2009, Komada, 2000).

- | | |
|------------------|--|
| Maz312 | Paris, Bibliothèque Mazarine, ms. 312 (Première Édition). |
| BnF152 | Paris, Bibliothèque nationale de France, fr. 152 (Première Édition). |
| BR987 | Bruxelles, KBR, ms. II. 987 (Première Édition). |
| JenEl9596 | Jena, ThULB, El. 9596 (Première Édition). |
| Bein129 | New Haven, Beinecke Library, ms. 129 (Première Édition). |
| BnF155 | Paris, Bibliothèque nationale de France, fr. 155 (Seconde Édition, 1297). |
| Ars5059 | Paris, Bibliothèque de l'Arsenal, ms. 5059 (Bible historiale complétée, 1317). |
| BL19DIII | London, British Library, Royal 19 D III (Seconde Édition, Bible à prologues, 1411–1412). |