

HEBREW UNION COLLEGE ANNUAL
VOLUME LXXVI

An Exploratory Study on the
Use of a Phylogenetic Algorithm
in the Reconstruction of
Stemmata of Halachic Texts

AVISHAI YORAV

Jerusalem

TAL DAGAN

Heinrich-Heine Universität, Düsseldorf

DAN GRAUR

Tel Aviv University and University of Houston

Edward A. Goldman, Editor
Richard S. Sarason, Associate Editor

Editorial Board

David Ellenson, Reuven Firestone, Nili Fox,
Alyssa Gray, Samuel Greengus,
Adam Kamesar, Barry Kogan, Michael Meyer,
Stanley Nash, David Sperling,
Dvora Weisberg

HEBREW
UNION COLLEGE
ANNUAL

Volume LXXVI

Cincinnati

2005

©2006 by
Hebrew Union College - Jewish Institute of Religion
Library of Congress Catalogue Card Number 25-12620
ISSN 360-9049

*Design and composition by Kelby and Teresa Bowers
Printed in the United States of America*

*The publication of this volume
of the Hebrew Union College Annual
was subventioned by*

*The Henry Englander-Eli Mayer Publication Fund
established in their honor by
Mrs. Esther Straus Englander and Mrs. Jessie Straus Mayer*

Submissions

We welcome for consideration scholarly essays in Jewish and Cognate Studies, Ancient and Modern: Bible, Rabbinics, Language and Literature, History, Philosophy, Religion. Please address your submission inquiries to The Editor at annual@huc.edu.

Authors submitting manuscripts for publication are asked to do the following:

- 1 For manuscript formatting, follow *The University of Chicago Manual of Style* in general and *The SBL Handbook of Style* specifically.
- 2 All manuscripts, including notes, should be continuously paginated, double-spaced and employ generous margins all around.
- 3 Every manuscript must be accompanied by an English abstract of 200 words maximum.

Previous Volumes

The *Annual* office can supply vols. XLV–LXXV excluding vols. LV, LXIII, and LXVI (out of stock): vols. XLV–LII at \$15.00, vols. LIII–LVIII at \$20.00, vols. LIX–LXXXII at \$30.00, and vols. LXXXIII–LXXXV at \$40.00. Printed copies of individual articles are available for a fee of \$5.00. For more information email the *Annual* office at: annual@huc.edu. The American Jewish Periodicals Center at HUC-JIR, 3101 Clifton Avenue, Cincinnati, Ohio 45220 can supply microfilm copies of vols. XXXVIII–XLIV. ProQuest Company, 300 North Zeeb Road, Ann Arbor, Michigan 48106 can supply 16 mm, 35 mm, 105 mm microfiche, as well as photocopies.

Supplements

Yosef Hayim Yerushalmi. *The Lisbon Massacre of 1506 and the Royal Image in the Shebet Yehudah*. 1976.

Mark E. Cohen. *Sumerian Hymnology: the Eršemma*. 1981.

William C. Gwaltney, Jr. *The Pennsylvania Old Assyrian Texts*. 1982.

Kenneth R. Stow. “*The 1007 Anonymous*” and *Papal Sovereignty. Jewish Perceptions of the Papacy and Papal Policy in the High Middle Ages*. 1985.

Martin A. Cohen. *The Canonization of a Myth: Portugal’s “Jewish Problem” and the Assembly of Tomar 1629*. 2002.

Hebrew Union College Annual
3101 Clifton Avenue, Cincinnati, Ohio 45220

Contents

- 1 *Ezekiel as the Voice of the Exiles and Constructor of Exilic Ideology*
Dalit Rom-Shiloni, *Hebrew Union College-Jewish Institute of Religion*
- 47 *Siqoriqin Forfeited Land*
Moshe Gil, *Tel Aviv University*
- 63 *The Earliest Texts of the Birkat Haminin*
Uri Ehrlich, *Ben Gurion University of the Negev*
Ruth Langer, *Boston College*
- 113 *The Blemished First-Born Animal*
A Case Study in Tannaitic Sources
Tzvi Novick, *Yale University*
- 133 *Justification by Deed*
The Conclusion of Sanhedrin-Makkot and Paul's Rejection of Law
Devora Steinmetz, *Jewish Theological Seminary*
- 189 *The Economics of Jewish Childhood in Late Antiquity*
Amram Tropper, *Ben-Gurion University*
- 235 *The Constitution of the Burial Society of the*
Bucharest Sephardic Community, April 30, 1850
Isaac Jerusalmi, *Hebrew Union College-Jewish Institute of Religion*
- 259 *Abraham Firkovich and the Dedication Inscription of the Aleppo Codex*
Yosef Ofer, *Bar-Ilan University*
- 273 *An Exploratory Study on the Use of a Phylogenetic Algorithm*
in the Reconstruction of Stemmata of Halachic Texts
Avishai Yorav, *Jerusalem*
Tal Dagan, *Heinrich-Heine Universität, Düsseldorf*
Dan Graur, *Tel Aviv University and University of Houston*

289 *Jewish Biblical Theology: Whence? Why? And Whither?*
Ziony Zevit, *University of Judaism*

א „ותשם בפוך עיניה ותיטב את ראשה” מלכים ב ט 30
Yossi Leshem, *Hebrew Union College-Jewish Institute of Religion*

יא רב ששת לעומת רב נחמן
שתי שיטות פרשניות למקורות תנאיים
Barak S. Cohen, *Bar-Ilan University*

ג „עלייה וירידה”
מפתח לפתרון חלום יעקב
Yitzhak Peleg, *Beit Berl College, Israel*

An Exploratory Study on the Use of a Phylogenetic Algorithm in the Reconstruction of Stemmata of Halachic Texts

AVISHAI YORAV

Jerusalem

TAL DAGAN

Heinrich-Heine Universität, Düsseldorf

DAN GRAUR

Tel Aviv University and University of Houston

Until recently, philological-historical analyses of ancient texts were based almost exclusively on the know-how of experienced scholars. The main disadvantage of this approach is the lack of a methodological means for resolving conflicting conclusions reached by different researchers. Additional problems arise when the amount of data that needs to be considered is too large to be dealt with manually, or when the evidence precludes a simple description of the ancestry of the various versions. This article describes the analysis of several halachic texts through the use of a phylogenetic algorithm called maximum parsimony, which was originally designed for the study of biological data. Hundreds of textual variants occurring in various manuscript witnesses of three halachic texts were used in conjunction with the maximum parsimony procedure to derive phylogenetic trees from encoded data. Group patterns produced by this algorithm were found to be fairly similar to some results obtained through conventional philological-historical research conducted on the same texts. In a significant number of cases, however, particular stemmatic statements were not supported by maximum parsimony. Our conclusion is that phylogenetic methodology may be useful to historical-philologists in reconstructing the stemmata of text traditions or the ancestry of halachic manuscripts. Computerized phylogenetic algorithms are expected to outperform the traditional manual approach especially as far as long documents are concerned, as well as in cases in which a proliferation of text witnesses needs to be considered.

The scientific tradition in the humanities differs in several ways from that in the natural and exact sciences. This is not only true of the divergent character of the subject matter, but also regarding the techniques employed for data organization and the methodology for deriving inferences. This essay explores the interdisciplinary use of a methodological tool from evolutionary biology in

philological-historical research.¹ In particular, we shall study the efficiency with which the history of ancient halachic text witnesses can be reconstructed correctly by maximum parsimony.

One of the principal tasks of historical philology as far as investigating texts is concerned is the arrangement of text witnesses in stemma form. Until recently, the standard methodology for stemmatic reconstruction was mainly based on the insights, know-how, and intuition of experienced scholars. The main disadvantage of this approach is the lack of a methodological means for resolving conflicting conclusions reached by different researchers. Additional problems arise when the amount of data to be considered is too large to be dealt with manually, or when inconsistencies due to contamination preclude the unequivocal description of the ancestry of the various text witnesses. The accuracy of inferred hierarchical clusterings of text witnesses increases with the number of paragraphs that are common to the textual variants under study. However, the number of possible combinations (stemmata) rises as well, and as a result so does the difficulty to reach properly integrated inferences. Manual resolution is only possible when the number of text witnesses is small, but becomes successively more difficult as these proliferate. A typical drawback of human intuition is the tendency to ignore evidence pertaining to negative relationships, that is, the assessment that links between two elements are *not found*. Phylogenetic algorithms can evaluate negative information as easily as positive information.

Here, we attempt to complement the traditional methodology with a computerized algorithm capable of clustering what might otherwise be indefinite or uninformative data. For convenience, we assume that the conventional analyses represent the “truth.” The reliability of the algorithm is, then, evaluated according to the level of congruence exhibited between its output and the “truth.” In general, we seek algorithmic assistance to human deduction. Under the reasonable assumption that it is impossible to program all the plethora of considerations necessary for philological-historical analysis, stemmatic reconstruction cannot be automated in its entirety. In general, we seek to draw on quantitative concepts, rather than be “restricted” by them.

Since the late 1960s, sophisticated methods for the reconstruction of the evolutionary history of biological species on the basis of molecular data (DNA and proteins) have been developed. Many of these methods have been implemented as computer software and are publicly available. In fact, this field of study, which is known as *molecular phylogenetics*,² has matured into a well established

1 This essay constitutes a second stage in the development of a methodology for the analysis of hierarchical relationships among halachic text witnesses. The first treatment of the subject was: Avishai Yorav, *Serial Norm for Comparison of Textual Variants* (Heb.) (August 2003) <http://www.daat.ac.il/daat/toshba/mechkar/norma-2.htm>.

2 For an introduction to the subject of molecular phylogenetics, see: Walter M. Fitch and Emman-

discipline mathematically, algorithmically, and biologically. Some of these reconstruction methods, that is, molecular phylogenetics, utilize textual information written either in the four-letter alphabet of DNA sequences or the 20-letter alphabet of proteins. In principle, therefore, these methods may be adapted for use on texts other than molecular sequences. There are many similarities between the development of text traditions and the evolution of biological entities. In both cases, information is transmitted by “descent with modification.” In living organisms, the information is contained in the genetic material, transmitted by replication of the genetic material, and modified by mutation. In text traditions, the information is the text, transmitted by copying, and modified by error or deliberate change. The similarities between biological and textual evolution have been noted in the 1970s.³ Beginning with the late 1990s, molecular phylogenetic methodology has been used to reconstruct the stemmata of text traditions and ancestor-descendant status of literary texts.⁴

uel Margoliash, “Construction of phylogenetic trees,” *Science*, 155 (1967) 279–284; Dan Graur and Wen-Hsiung Li, *Fundamentals of Molecular Evolution* (2nd ed.; Sunderland, Mass.: Sinauer Associates, 2000); Joseph Felsenstein, *Inferring Phylogenies* (Sunderland, Mass.: Sinauer Associates, 2003); Simon Whelan, Pietro Lió, and Nick Goldman, “Molecular Phylogenetics: State-of-the-art Methods for Looking into the Past,” *Trends in Genetics* 17 (2001) 262–72; David L. Swofford. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods)* (Sunderland, Mass.: Sinauer Associates, 2001); Roderic D. M. Page and Edward C. Holmes, *Molecular Evolution: A Phylogenetic Approach* (Oxford: Blackwell Science, 1998); Willi Hennig, *Phylogenetic Systematics* (Urbana: Univ. of Illinois Press, 1966); Ian J. Kitching, Peter L. Forey, Christopher Humphries, and David M. Williams, *Cladistics: The Theory and Practice of Parsimony Analysis*, Systematics Association, Publication 11 (Oxford: Oxford Univ. Press, 1998); Naruya Saitou and Masatoshi Nei, “The Neighbor Joining Method: A New Method for Reconstructing Phylogenetic Trees,” *Molecular Biology and Evolution* 4 (1987) 406–25.

- 3 Norman I. Platnick and Howard Donald Cameron, “Cladistic Methods in Textual, Linguistic, and Phylogenetic Analysis,” *Systematic Zoology* 26 (1977) 380–85.
- 4 Adrian C. Barbrook, and others, “The Phylogeny of *The Canterbury Tales*,” *Nature* 394 (1998) 839; Christopher J. Howe, and others, “Manuscript Evolution,” *Trends in Genetics* 17 (2001) 147–52; Matthew Spencer and Christopher J. Howe, “Estimating Distances Between Manuscripts Based on Copying Errors,” *Literary and Linguistic Computing* 16, 4 (2001) 467–84; Linne R. Mooney, and others, “Stemmatic Analysis of John Lydgate’s Verse Chronicle ‘*The Kings of England Sithen William the Conqueror*,’” *Revue d’Histoire des Texts* 31 (2003) 277–99; Arthur R. Lee, III, “Numerical Taxonomy Revisited: John Griffith, Cladistic Analysis and St. Augustine’s *Quaestiones in Heptateuchum*,” *StPatr* 20 (1989) 24–32; Peter M. W. Robinson, “Computer-Assisted Stemmatic Analysis and ‘Best-Text’ Historical Editing,” in Pieter van Reenen and Margot van Mulken, eds., *Studies in Stemmatology* (Amsterdam: Benjamins, 1996); Peter M. W. Robinson and Robert J. O’Hara, “Cladistic Analysis of an Old Norse Manuscript Tradition,” in Susan Hockey and Nancy Ide, eds., *Research in Humanities Computing* (Oxford: Oxford Univ. Press, 1996) 4, 115–37; Stephen C. Carlson, “A Phylogenetic Approach to N.T. Textual Criticism” (December 1, 2002) <http://www.mindspring.com/~scarlson/tc/>. An interesting forerunner to the use of phylogenetic methodology in the humanities is the field of glottochronology, which deals with the evolution of languages. For an early short review of the subject, see Robert B. Lees, “The Basis of Glottochronology,” *Language*, 29 (1953) 113–27. The last few years have also witnessed the emergence of a “post-modernistic”

So far, however, no Hebrew or Aramaic manuscripts have been subjected to phylogenetic analysis.

Our paper has two goals. The first is to apply the phylogenetic methodology to rabbinic texts. The second is to compare deductions derived from traditional studies with those derived from quantitative phylogenetic analyses. This comparison may give us some clues on the extent with which computerized phylogenetic algorithms, which were obviously designed for distinctively unrelated purposes, may be of assistance to philological historians interested in rabbinic texts.

Phylogenetic trees for the different versions of a manuscript were inferred using maximum parsimony. Maximum parsimony is a method of evolutionary inference in which the phylogenetic tree with the least number of changes is chosen as the most likely one.⁵ How evolutionary change is measured depends on the particular variant of the parsimony employed. In this study, we used the simplest parsimony method, called Wagner parsimony.⁶ Our data was coded into binary character states: 1 for presence, 0 for absence. The algorithm can also process a modest amount of “no data” input. The Wagner parsimony criterion assumes that each character can be either inserted (0 changes to 1) or deleted (1 changes to 0).⁷ We count the number of changes during the “evolution” of a manuscript for any possible tree. The most parsimonious tree is the one that requires the smallest number of changes.

The Wagner maximum parsimony algorithm (as well as other variants of maximum parsimony) yields unrooted phylogenetic trees. In such a tree, the distances between items and their relative arrangement on it represent an estimate of the affinities among the items. The algorithm identifies pairs of related items (sister taxa), and associates them with other items or clusters. In this study we opted for unscaled trees, that is, trees in which the lengths of the branches do not provide a quantitative estimate of the number of changes along the branch. The affinity between items or groups of items (degree of similarity) is graded by the number of branches separating them. Since the tree is unrooted, the direc-

approach. Followers of this trend are no longer concerned with the “original text,” but rather treat each text witness as a distinct entity. Steering clear of any ideological discussion on this issue, we will only comment that our present study deals with situations in which it is safe to assume that the text witnesses have, in fact, “evolved” from an ancestral *urtext*.

- 5 For the reconstruction of the trees, we used the MIX software within the PHYLIP package. Joseph Felsenstein, “Parsimony in Systematics: Biological and Statistical Issues,” *Annual Review of Ecology and Systematics* 14 (1983) 313–33.
- 6 Joseph Felsenstein, “PHYLIP – Phylogeny Inference Package (Version 3.2)” *Cladistics* 5 (1989) 164–66. Newer versions, such as the one used by us, can be obtained from Joseph Felsenstein at the Department of Genetics, University of Washington, Seattle (<http://evolution.genetics.washington.edu/phylip.html>).
- 7 The more common variant (character state) was assigned a value of zero; the other was assigned a value of one.

tion of an affinity is unspecified. That is, neighboring items may not necessarily have a common ancestor or be descended from one other. To determine the direction of a tree, that is, to produce a real historical stemma-like diagram, a particular item or group of items should be identified as the root of the phylogenetic tree.⁸ Following the selection of a root, a directional (or rooted) tree is produced. This tree represents one evolutionary path from the root, via intermediate (mostly hypothetical) items, to the texts under study, which appear as terminal nodes.⁹ Specifying the true root requires knowledge that is extraneous to the texts under study (for example, the ages of the manuscripts or testimony pertaining to the order in which the most ancient versions within a group have been copied). Alternatively, a trial-and-error system may be used, in which the researcher designates a particular item or group of items as a root, and then makes an expert decision on the likelihood of the resulting rooted tree based on external considerations. In the absence of any ability to root the tree unequivocally, unrooted trees may be used to test certain hypotheses concerning the relationships among manuscripts.

The number of possible trees that can be built depends on the number of textual variants under consideration. This number increases quite rapidly with the number of text witnesses,¹⁰ and may become even larger in datasets with missing characters or indefinable character states. We note that the resulting tree may contain errors; however, by using such statistical methods as bootstrapping, one can ascertain which branches are reliable and which are not.

Application of a method from such an unrelated discipline as molecular evolution highlights a series of assumptions of conventional philological methodology, some of which are at odds with those used in maximum parsimony. For example, when studying copies of a text, one must allow for multiple identical errors or corrections of independent origin. This process of babelization¹¹ in the evolution of texts is analogous to the occurrence of homoplasies (reversals, parallelisms, and convergences) in DNA or proteins. Due to these processes,

8 It is important to note that the root must be uniquely specified. Therefore, a stemma with two or more elementary branches cannot be produced. Hybrid products in either evolutionary biology or philology cannot be analyzed with a methodology designed to produce binary graphs.

9 Inferred (or hypothetical) items on the diagram appear as internal nodes. The actual texts under study appear as terminal nodes and are indicated in this study by Hebrew letters.

10 The number of bifurcating unrooted trees for n textual variants is very large, $(2n-5)!/[2^{n-3}(n-3)!]$. The number of rooted trees is even larger, $(2n-3)!/[2^{n-3}(n-3)!]$. Given the staggering number of possible trees, only about 25 textual variants can be dealt with by exact maximum parsimony algorithms. When the number of variants exceeds 25, tree reconstruction requires heuristic approximations.

11 Babelization or “babylonization” has been shown to be the cause of similarities between a tenth-century text from Italy, a fourteenth-century text from Spain, and a seventeenth-century text from Germany, independently of any common textual substrate. See, Avishai Yorav, *Serial Norm for Comparison of Textual Variants*.

similar character states may not always constitute evidence of a genealogical link between text-witnesses. Methods of phylogenetic reconstruction, including maximum parsimony, are ill equipped to deal with homoplasious or babelized character states. Moreover, the one-root-only and the binary-bifurcation-only rules may be too stringent for dealing with the evolution of texts.

Complete reconstruction of relationships among text witnesses consists of two steps: the grouping phase (that is, construction of an unrooted phylogenetic tree) and determination of directional ancestral-descendant relationships (that is, rooting the unrooted tree). Phylogenetic methodology can only assist in the first step of the process. It is important to emphasize that the phylogenetic analysis is purely morphological and does not deal with the different meanings of the textual variants. Moreover, phylogenetic analysis can only be used to study stemmatological relationships within a set of textual variants. It cannot be used to aid in the resolution of questions pertaining to version traditions prior to the consolidation of the definite version of a text.

In this study, three halachic texts composed in late-antiquity or early medieval times and transcribed during the Middle Ages were investigated. The surviving text witnesses had been studied and documented extensively using traditional methods. Meticulous encoding of the textual variants was carried out according to a standardized scale.¹² Trivial textual variants at level 1 of the scale¹³ were not used. Verification studies were not conducted; our basic premise was that the published text witnesses are errorless. The text witnesses were positioned as “taxa” and the textual variants as “characters.”

FIRST CASE STUDY

SEFER HILKHOT HARIF ^cAL MASSEKHET PESAH RISHON

The editor of *Sefer hilkhot harif ^cal massekhet pesah rishon*¹⁴ studied 14 manuscripts and a number of early printed editions. His conclusion was that all the printed editions were derived from the Constantinople edition (7) and, therefore, it is the only printed edition that needs to be considered in a stemma. Moreover, one of the manuscripts (7) is thought to have been derived from the printed editions and may, therefore, be excluded from further consideration.

12 See Yorav, *Serial Norm*. A detailed list of all the textual variants used in this study will be presented in a separate report. In the course of encoding, we were compelled to make several assumptions regarding the nature of the texts. Indeed, the encoding method and the hypotheses intrinsic to it may have added unrealistic elements to the study. For example, the assumption that variations within sentences or paragraphs are independent of one another may be problematic.

13 Level-1 variants include minor variations, such as alternations between complete and partial spellings or the use of very similar letters.

14 R. Hillel Hyman, *Sefer hilkhot harif ^cal massekhet pesah rishon* (Jerusalem: Jewish Theological Seminary, 1990).

The editor suggests that the text witnesses can be clustered into three groups¹⁵ (hypothesis 1). The first group consists of ה, ט, י, and כ (hypothesis 2), the second of ל, מ, נ, ס, and ט (hypothesis 3), and the third of ע, פ, צ, ק, and ר (hypothesis 4). He further conjectures that manuscript נ was copied from ס (hypothesis 5), and ל and מ are to some extent related to the first group (hypotheses 6 and 7). ט is the most recent of the text-witnesses and was influenced to some degree by all the groups (hypothesis 8). ע is somewhat closely related to the second group (hypothesis 9). פ, צ and ק are said to more closely related to one another than any of them is related to either ע or ר (hypotheses 10 and 11, respectively). Of these eleven hypotheses, some are untestable by maximum parsimony (for example, hypothesis 8); others, such as hypotheses 3 and 6, are contradictory in the phylogenetical sense (that is, they cannot both be true). Moreover, some hypotheses, for example, hypothesis 2 may be partially correct and partially wrong. Finally, some hypotheses, for example, hypothesis 1, may only be tested against the rooted tree.

We applied the maximum parsimony algorithm to fourteen text witnesses consisting of thirteen manuscripts (ט was excluded) and one printed edition (ט). The textual variants in the first three chapters, consisting of approximately half of the text, were used. A total of 611 non-trivial variations was found. The resulting unrooted phylogenetic tree is shown in *Figure 1*. A rooted tree with ט as root is shown in *Figure 2*.

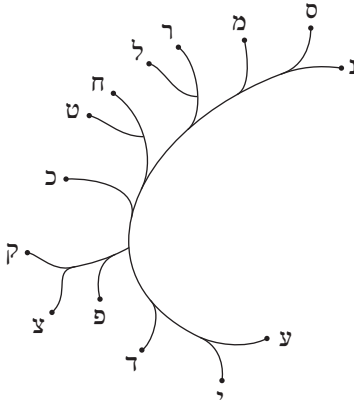


Figure 1. Unscaled unrooted phylogenetic tree for 14 text witnesses of *Sefer hilkhot harif'al massekhet pesah' rishon*.

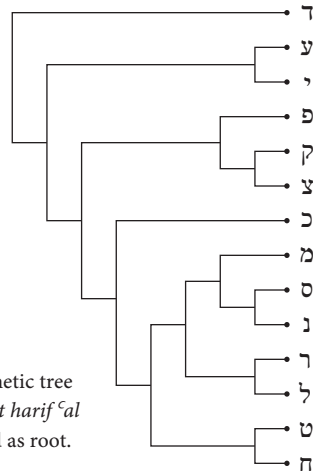


Figure 2. Unscaled rooted phylogenetic tree for 14 text witnesses of *Sefer hilkhot harif'al massekhet pesah' rishon*. ט was used as root.

¹⁵ Hyman, *Hilkhot harif*, 11–14.

Let us first test several hypotheses against the unrooted phylogenetic tree. Hypothesis 2 is only very partially supported by the phylogenetic analysis, with η and υ emerging as sister taxa. Hypothesis 3 is partially supported by the tree, if γ is excluded and δ is included. Hypothesis 4 is partially supported by the tree, if γ is included and ν and δ are excluded. Hypothesis 5 is strongly supported by the tree. In the matter of the weak hypothesis 9, no support was garnered for a link between ν and the group containing η , υ , ν , and δ , although we recovered a close relationship between ν and ν .

An inspection of the rooted tree in *Figure 2* indicates that it is impossible to divide the tree into three monophyletic groups of any kind, let alone into the particular tripartite division envisioned by R.Hillel Hyman. Moreover, the choice of the root does not matter, as the unrooted tree in *Figure 1* also is impervious to tripartition. We must, therefore, conclude that in the strict sense of the word, our results are incongruent with those of the editor.

SECOND CASE STUDY

THE ARAMAIC TEXT OF THE TARGUM OF JOB

The editor of the *Targum of Job*¹⁶ examined fourteen manuscripts and two early printed editions. In his opinion, one of the manuscripts (δ) reflects two variant traditions. The first of these is found in chapters 1–13 (1 δ); the second in chapters 14–42 (2 δ). In total, there were 17 text witnesses. The editor identifies four basic groups, with certain links between them.¹⁷ A simplified rendition of this stemma is shown in *Figure 3*.

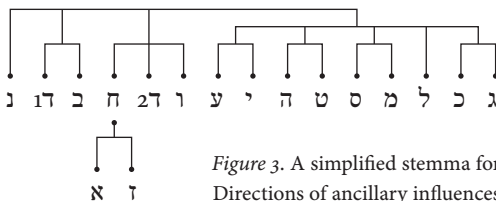


Figure 3. A simplified stemma for 17 text witnesses of *Targum of Job*. Directions of ancillary influences, degrees of certainty, and “group” influences are not shown.

Maximum parsimony was applied to the textual variants of chapters 1–6 and 14–19 (approximately 30% of the text). A total of 788 variations was found. The unrooted tree is shown in *Figure 4*. We have considered several possible rootings for the tree, for example, \aleph and γ .¹⁸ In the end, we opted for a rooted tree that maximizes the iconographic similarity to the traditional stemma (*Figure 5*).

¹⁶ David M. Stec, *The Text of the Targum of Job* (Leiden: Brill, 1994).

¹⁷ Stec, *Targum of Job*, 84.

¹⁸ γ is noteworthy in that it is the only manuscript from North Africa; the others are from Spain, Germany, or Italy. \aleph is the most recent of the text witnesses.

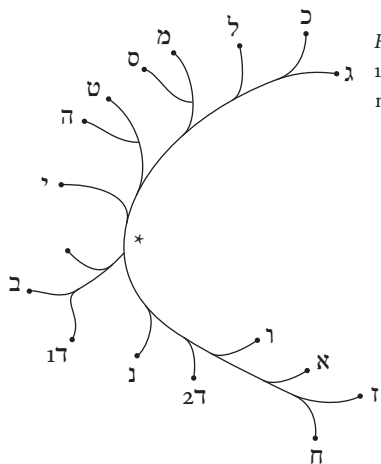


Figure 4. Unscaled unrooted phylogenetic tree for 17 text witnesses of *Targum of Job*. An asterisk marks the position of the root used in Figure 5.

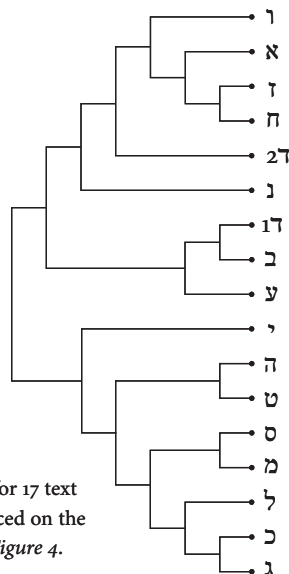


Figure 5. Unscaled rooted phylogenetic tree for 17 text witnesses of *Targum of Job*. The root was placed on the internal branch marked with an asterisk in Figure 4.

A comparison of the stemma with the maximum parsimony rooted reconstruction reveals a truly extraordinary degree of congruence. Four basal groups are recovered. The first groups consisting of א, ב, ג, ד, ה, ו, ז, ח, ט, י it is almost identical with the group proposed by David Stec. The only differences between the phylogenetically inferred group and the one in the stemma concerns the position of י, which in the maximum parsimony tree clusters with the correctly identified groups consisting of ב and יא. In fact, even for the position of י it is not possible to claim total error, because י is suspected to have been affected by multidirectional influences. The internal relationships within the basal א-ב-ג-ד-ה-ו-ז-ח-ט-י group were also identified correctly, with א clustering with ב, ג clustering with ד, ה clustering with ו, and ז clustering with ח. The algorithmic reconstruction confirms the grouping of ב with יא, and strongly supports the editor's decision to separate י into two independent text witnesses, יא and יב. The algorithmic reconstruction also confirms the clustering of י, ז, ח, ט, and יב. In the stemma, י appears as a group by itself. The phylogenetic analysis indicates a distant affinity with the י-ז-ח-ט-יב cluster.

THIRD CASE STUDY
TRACTATE *PESAḤIM* CHAPTERS 3 AND 4

The fact that chapters 3 and 4 of *tractate Pesahim* in Talmud Bavli have been annotated and edited independently by two researchers in two academic institutions on different continents,¹⁹ allows us to pursue two types of comparative research. The basic assumption (with very rare exceptions) is that two chapters derived from the same manuscript represent a single integral textual unit. Therefore, if the phylogenetic algorithm works without error and if the various text witnesses do not contain hybrid parts, then the phylogenetic tree based on one chapter is expected to be identical to the phylogenetic tree based on the text of the other chapter. In other words, the text-witness groupings should be identical in both trees. Our first comparison will, therefore, contrast the inferred tree based on the textual variants of chapter 3 with the inferred tree based on the textual variants of chapter 4. We will, thus, be able to identify which parts are congruent between the two trees, that is, have been reliably reconstructed, and which are not. Our second sets of comparisons will contrast the trees derived by maximum parsimony with those derived by traditional means, whose technical details are oftentimes unspecified and frequently unspecifiable.²⁰ We note that in this type of analysis it is possible to evaluate (or grade) the “relative worth” of different scholars working in the conventional philological-historical tradition. We shall abstain from using maximum parsimony to raise personal philological criticisms. Rather, we shall use these comparisons to discuss the consistencies of the general conclusions of the editors.

The editor of chapter 3 of *Pesahim*²¹ examined ten manuscripts, a printed edition (רנציה), and seven Genizah fragments. The editor provides a fairly detailed account of the methods employed in his stemma arrangement.²² He distinguishes between “significant differences,” “widespread significant differences,” and other differences (but does not specify the criteria for the classification). He lists 128 variations between the text witnesses. His stemma consists of two main branches and an indeterminate group. Simply put, the typical representative of branch I is רט' 125, רט' 134, רט' 109, and רנציה belong to this branch. רט' 109, and רנציה represent later additions. Branch II is epitomized by רט' 1608.

19 It is important to note that the studies were indeed independent; there was no cooperation between the researchers, and at times competition was implied.

20 For example, Hyman, *Hilkhot harif*, 14, states: “Ultimately I must admit that an important part of the arrangement [of the text witnesses] into families was based on a sense born of hundreds of hours of work and experience . . .”

21 Shmuel Yosef Wald, *Talmud Bavli, Tractate Psachim, Third Chapter* (Jewish Theological Seminary; Jerusalem, 2000).

22 Wald, *Psachim*, 318–26.

Other descendants of branch II are 6 מִינְכָן, שְׁשׁוֹן, גְּנִיזָה 1, אֲוִקְסְפּוֹרֵד, and 95 מִינְכָן. 1623 סִמ' and אֲוִקְסְפּוֹרֵד are somewhat more distantly related to branch II. 1623 סִמ' and קוֹלוֹמְבִיָּה, which are Yemenite manuscripts, are said to be closely related to each other, but their relationship to the other two groups was not determined. *Figure 6* presents a schematic and simplified summary of the stemma. The stemma rests on a very elaborate system of considerations. After delineating which text witnesses are viewed as distinctively representative of the two branches, and claiming that the other text witnesses are “heterogeneous,” the editor presents 20 two-dimensional tables²³ quantifying the degree of agreement between the various “heterogeneous” text witnesses and the representatives of the two branches. Such meticulous statistical processing is unusual in philological-historical editing and analysis. Nevertheless, decisions based on traditional analyses are sometimes arbitrary.²⁴ The placement of וְנִצִּיָּה and 109 וְט' is especially problematic. Shmuel Yosef Wald appended them to branch I in the graphical stemma,²⁵ but his detailed tables²⁶ appear to indicate a branch-II affiliation.²⁷ No explanation for this contradiction is provided. The above clearly demonstrates that “general impressions” may influence experts more than tabulations of concrete quantities.

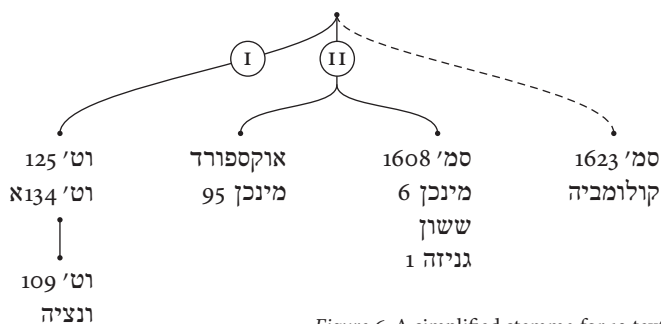


Figure 6. A simplified stemma for 12 text witnesses of chapter 3 of *Pesahim*.

23 Wald, *Psachim*, 319–25.

24 All the analyses start with a dichotomic division, whereby two text witnesses are identified as “opposite poles.” Subsequently, all the other text witnesses are compared separately to the two opposites to determine how similar or dissimilar they are to the two poles. 125 וְט' was chosen as main representative of branch I and 1608 סִמ' was selected as main representative of branch II. Surprisingly, no reason is provided for this fundamental assertion, other than claiming that contradictions between the text variants were found in 95–98% of the cases.

25 Wald, *Psachim*, 283.

26 Wald, *Psachim*, 323, 325.

27 In the tables, there is a 39% match of 109 וְט' to branch I and a 56% match to branch II. וְנִצִּיָּה has a 30% match to branch I and a 67% match to branch II.

For chapter 3, we performed a maximum parsimony analysis on twelve text witnesses consisting of ten manuscripts, a printed edition, and a Genizah fragment (גניזה 1). One of the manuscripts (וט' 134א) encompasses only approximately 75% of the text. גניזה 1, which is the largest of the Genizah fragments available for analysis, contains only about 12% of the text. Hence our analysis contained a considerable fraction of missing data. We identified 353 non-trivial variations. The unrooted maximum parsimony tree is shown in *Figure 7*. The first and foremost observation is that Wald's tripartite division of the text witnesses cannot be recovered from the unrooted tree no matter which node or internal branch is chosen as root. Moreover, of the many groups and subgroups identified in the stemma in *Figure 6*, the only clustering that is recovered by the maximum parsimony phylogenetic algorithm is that of the Yemenite manuscripts קולומביה and סמ' 1623. 6 מינכן and 1608 סמ' were found to be closely related, as in the stemma, but without either ששון or גניזה 1. At this point it is not clear whether or not the lack of congruence between the stemma and the phylogenetic tree for chapter 3 is solely due to the fragmentary nature of וט' 134א and גניזה 1.

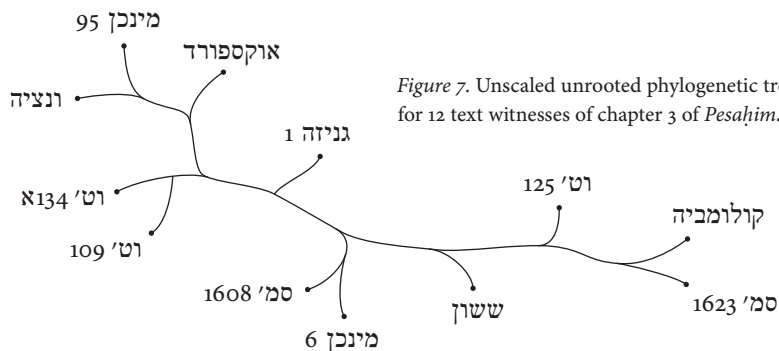


Figure 7. Unscaled unrooted phylogenetic tree for 12 text witnesses of chapter 3 of *Pesahim*.

The editor of chapter 4²⁸ examined 12 large text-witnesses and 11 additional Genizah fragments.²⁹ He identifies two main branches. Branch I is most distinctively represented by וט' 125; branch II by 6 מינכן. Aaron Amit does not present a graphic stemma, but claims that 1608 סמ' (in 75% of the cases) and וט' 109 (in 60% of the cases) agree with branch II. The remaining text-witnesses are claimed to be heterogeneous. Manuscript וט' 134א was identified as composed of two unequal parts, each of which originated from a different text witness. There are only very minor variations between קולומביה and סמ' 1623. The editor does not say anything explicit regarding the arrangement of the other

28 Aaron Amit, "Makom sheNahagu, Talmud Bavli, Tractate Pesachim, Fourth Chapter" (masters thesis; Bar-Ilan Univ., Ramat Gan, 1995).

29 Not including גניזה 1 (mentioned in the previous section) as it lacks text from the fourth chapter.

text witnesses, nor does he offer verification for his contentions, except for directing the reader in general terms to an elaborate apparatus of textual variant comparisons in the thesis.

The maximum parsimony algorithm was run on thirteen text witnesses of chapter 4 (134 'וט was divided into א134 וט and ב134 וט). The Genizah fragments were excluded from the analysis. There were 586 non-trivial variations throughout the chapter. The unrooted maximum parsimony tree is shown in *Figure 8*.

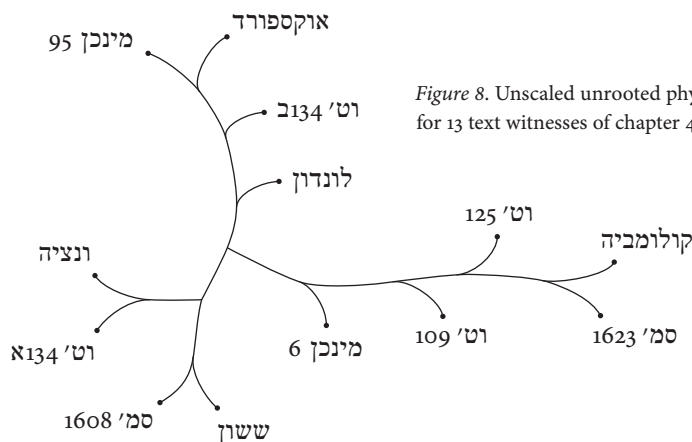


Figure 8. Unscaled unrooted phylogenetic tree for 13 text witnesses of chapter 4 of *Pesahim*.

Aaron Amit provided almost no strong arguments in favor of or against certain groupings. As a result it is difficult to contrast his claims with those obtained through maximum parsimony. All that can be said is that the grouping of קרלומביה with 1623 'סמ is confirmed. On the other hand, the phylogenetic tree does not support a link between 6 מינכן and 1608 'סמ or 109 'וט.

In the main, the conclusions reached by the editor of chapter 4 of *Pesahim* agree with those of the editor of chapter 3.³⁰ Unfortunately, Amit does not list his considerations, other than discussing the differences between 125 'וט and 6 מינכן and providing a series of examples of textual variations between the two branches.³¹ Of the 10 examples from category I, which he regards as most significant, he lists קרלומביה and 1623 'סמ as exhibiting strong similarities to branch I, and 1608 'סמ, רנציה, אוקספורד, 134 'וט, and שרון to branch II, with no clear resolution regarding 109 'וט and 95 מינכן. If we accept this sample as representative – in practice, the examples were merely meant to illustrate the basic division between the two branches rather than arrange the text witnesses – then there

³⁰ Both Wald and Amit select 125 'וט as the main representative of branch I. The editor of chapter 3 prefers 1608 'סמ to represent branch II, while the editor of chapter 4 prefers 6 מינכן. The two text-witnesses are 85% similar according to Wald, and 75% similar according to Amit.

³¹ Amit, "Makom sheNahagu," 19–37.

are several discrepancies between the conclusions of the two editors. According to the editor of chapter 3, א¹³⁴ רט' belongs unequivocally to branch I, whereas Amit assigns it to branch II. According to the editor of chapter 3, קולומביה and סמ' 1623 are hybrids with a slight tendency toward branch II. The editor of chapter 4 agrees with the hybrid status, but sees a greater similarity to branch I. Wald assigns 109 רט' and ונציה to branch I (although, as we have previously observed, in his tables he places them closer to branch II). According to Amit they are closer to branch II.

To compare the two traditionally derived stemmata with the results of the maximum parsimony algorithm, as well as assess the source of incongruity between the phylogenetic trees for chapters 3 and 4, we combined the data from the two chapters into a single database. The eleven text witnesses that contain both chapter 3 and chapter 4 were used in the combined analysis.³² א¹³⁴ רט' was identified as a composite of two separate documents,³³ one from the thirteenth century and one from the fifteenth. For the combined analysis, we only used the later document that contains both chapters. This “pruning” of fragmentary and hybrid text witnesses is likely to improve our chances of reconstructing a less ambiguous tree than the previous ones. In total, 940 non-trivial textual variations were used.

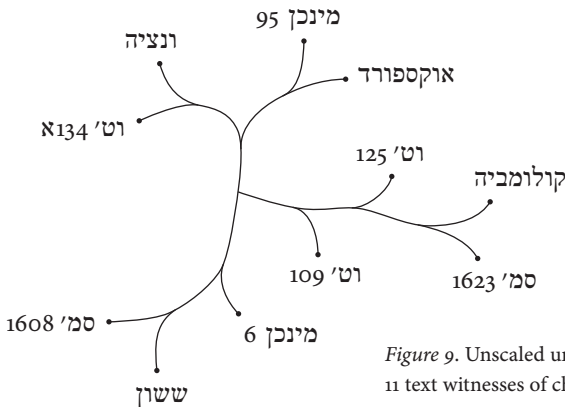


Figure 9. Unscaled unrooted phylogenetic tree for 11 text witnesses of chapters 3 and 4 of *Pesahim*.

The unrooted maximum parsimony tree for the combined data is shown in Figure 9. The first observation is that it is impossible to partition the tree into the two branches advocated by the traditional studies. The putative representatives

32 For example, the לינדרן manuscript (extant in chapter 4 alone) and גניזה 1 (only partially extant in chapter 3) were excluded.

33 Pages 5b–11a and pages 54b to the end are from the earlier document (Germany, thirteenth century?); pages 15a to 54b, except pages 45b–48a, were completed by a later scribe (Germany, fifteenth century?).

of branch I (125 'רט, א134 'רט, 109 'רט, and ונציה) are scattered throughout the tree, so that no matter where we place the root the monophyly of the two branches cannot be recovered. On the other hand, the maximum parsimony tree strongly supports a monophyletic clade consisting of the principal members of branch II (1608 'סמ, 6 מינכן, and ששון). א134 'רט and ונציה are clustered, but they have no relationship to branch II. Interestingly, they are intimately related to 125 'רט, which is the typical branch-I representative. The algorithm links א134 'רט to ונציה with no affinity to either branch I or branch II. This is in contradiction to the claim by the editor of chapter 3, who assigned both א134 'רט and ונציה to branch I. The editor of chapter 4 regards these text witnesses as "composites." The intermediate position of ונציה and א134 'רט supports this contention. The maximum parsimony algorithm's product places 109 'רט as a sister taxon of 125 'רט and the קרלומביה and 1623 'סמ clade. The editor of the third chapter assigned it (with some reservation) to branch I, while the editor of the fourth chapter assigned it to branch II (with reservations as well).

EVALUATION OF THE MAXIMUM PARSIMONY ALGORITHM IN RECONSTRUCTING STEMMATIC RELATIONSHIPS AMONG HALACHIC TEXT WITNESSES

An exhaustive evaluation of the relationships between the products of two methodologies from such distant disciplines is an impractical endeavor. The best we can do is to look at the degree of congruence between the methods and assess the usefulness of the maximum parsimony algorithm as an aid in the reconstruction of stemmatic relationships among halachic text witnesses. The facts, as outlined above, are that in some cases we obtain surprising degrees of congruence between the phylogenetic algorithm and traditional methodology. For example, in the case of the Aramaic *Targum of Job*, a truly extraordinary degree of congruence was observed. Basal and distal groups were identified correctly, and even differences could be explained by hybridization and multidirectional influences. In fact, in the case of *Targum of Job*, several new relationships (clades) have been discovered on which traditional methods had nothing to say. We believe that these relationships should be further investigated. Admittedly, the congruence between the two methods is far from perfect, and in some cases there was very little in common. We think, however, that despite the inconsistencies discovered between the methods, the computerized algorithm may be of significant aid to conventional research. Had the editors of the four texts seen the phylogenetic outputs, they might have concentrated their efforts and restricted themselves to examining a limited number of "suspect" groups. It is quite possible that the phylogenetic results would have

allowed them to settle doubts, refine groups, and possibly improve upon their final conclusions. Of course, binary trees derived through phylogenetic methodology can never be “true” in the absolute sense of the word, since it is universally accepted that all the text witnesses in our exploratory study have been subject to some degree of contamination.³⁴

Traditional philological-historical studies require the researcher to approach his subject free of prejudice, but not without prior knowledge. However, sometimes the number of alternatives that a philologist may be required to consider is simply too large, and the task of stemmatic reconstruction becomes an impossible one. For example, the editor of the *Targum of Job* considered 17 text witnesses. Thus, in theory, he should have considered 6, 190, 283, 353, 629, 375 unrooted arrangements and 191, 898, 783, 962, 510, 625 rooted ones.³⁵ Given these mind-boggling numbers, a computer program is likely to be of great value by focusing the attention of the researcher on particular subsets of text witnesses.

One can only consider what it would take to build a stemma for the *Maimonides Code (Mishneh Torah)*, which consists of approximately 800,000 words and for which there are several hundreds of text witnesses (mostly in fragmentary form). Can this task be even contemplated by using “manual” labor?

34 A similar problem exists in phylogenetic reconstruction of bacteria that have experienced horizontal gene transfer (textual contamination) in their past. See Hervé Philippe and Christophe J. Douady, “Horizontal Gene Transfer and Phylogenetics,” *Current Opinion in Microbiology* 6 (2003) 498–505. Several phylogenetic methods have been proposed in the literature to deal with such “contamination,” for example, Patricia Escobar-Páramo, and others, “Decreasing the Effects of Horizontal Gene Transfer on Bacterial Phylogeny: The *Escherichia Coli* Case Study,” *Molecular Phylogenetics and Evolution* 30 (2004) 243–50. It is not inconceivable that this class of methods may one day be adopted to identify multidirectional influences in text witnesses.

35 Graur and Li, *Fundamentals of Molecular Evolution*, 173.