Available online at www.sciencedirect.com
SCIENCE DIRECT®
Stud. Hist. Phil. Sci. 37 (2006) 237–268
Studies in History and Philosophy of Science
www.elsevier.com/locate/shpsa

Copying and conflation in Geoffrey Chaucer's *Treatise on the astrolabe*: a stemmatic analysis using phylogenetic software

Catherine Eagleton [a], Matthew Spencer [b]

[a] *Department of Coins and Medals, The British Museum, London WC1B 3DG, UK*
[b] *Department of Mathematics and Statistics, Dalhousie University, Halifax, NS B3H 3J5, Canada*

Received 8 February 2005; received in revised form 17 August 2005

Abstract

Chaucer's *Treatise on the astrolabe* is one of the earliest English-language works on an astronomical instrument. It draws on earlier sources, including a work on the astrolabe attributed in the Middle Ages to Messahalla, but reorders and reworks these sources to produce a description of the parts of, and the use of, the planispheric astrolabe. In their turn, fifteenth-century scribes sometimes drew on more than one source when producing a new copy of Chaucer's text. Conflation of this kind means that working out the relationship between the extant manuscript copies of the text can be complex and confusing. In this paper we reconsider the relationship between copies of Chaucer's *Treatise on the astrolabe* by using new techniques derived from evolutionary biology to clarify the copying and conflation of this important early scientific text. We propose a stemma for the extant manuscripts, place fragmentary copies of the text on it in relation to the more complete versions, and argue that the conflation of Chaucer's text may have been motivated by a concern amongst scribes for the completeness of what always was an incomplete work.

Keywords: Astrolabe; Geoffrey Chaucer; Manuscripts; Stemmatics; Phylogenetics

E-mail addresses: katie@eagleton.net (C. Eagleton), matts@mathstat.dal.ca (M. Spencer).

0039-3681/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsa.2005.08.020

1. Introduction

Chaucer's *Treatise on the astrolabe* is one of the earliest English-language works on an astronomical instrument. It draws on earlier sources, including a work on the astrolabe attributed in the Middle Ages to Messahalla, but reorders and reworks these sources to produce a description of the parts of, and the use of, the planispheric astrolabe.[1] The text of the *Astrolabe* divides into two main sections, which for convenience are usually called part I and part II by modern editors. Part I describes the parts of the astrolabe, naming them and explaining their functions, and part II details how to use the astrolabe to observe and to calculate.

In his introduction to the text, Chaucer explains that he will write a five-part treatise, including a set of tables. Michael Seymour argues that the text was completed, but that the tables dropped from the scribal tradition 'being troublesome to copy',[2] and he identifies sections II16–25 and II26–40 as the planned third and fourth sections of Chaucer's work. Other modern scholars believe that the text was probably never finished, and the planned third, fourth and fifth parts, along with the tables, were never part of the scribal tradition of the *Treatise on the astrolabe*. This view is supported by the *explicit* naming Chaucer in London, Royal College of Physicians, MS 358, which reads 'explicit secundum Galfredi Chauciers', at the end of part II, indicating that this scribe, at least, thought that this was the end of the work.[3]

This *explicit* naming Chaucer was written by a scribe linked to the Augustinian Priory at Thurgarton, Nottinghamshire. Only section II36 survives,[4] the earlier sections having presumably been contained in a quire which later became detached, but it ends with the explicit given above, confirming Chaucer as the author of the text of the *Astrolabe*. Roughly contemporary with this, John Lydgate indicated Chaucer's authorship of the work in his poem *Fall of princes* (written ca. 1435).[5] Other attributions—in marginal notes and colophons—to Chaucer tend to have been added to the manuscripts after they were copied, sometimes several centuries later.[6]

The popularity of the *Astrolabe* with medieval copyists and readers is indicated by the number of copies that survive—the current total is thirty-four manuscripts, more surviving copies than any other of Chaucer's works, except for the *Canterbury tales*.[7] In addition, the physical format of *Astrolabe* manuscripts varies greatly—some were clearly professionally produced; others are likely to have been copied by an individual for their own use or interest. Whilst acknowledging that some of the surviving manuscripts are large, Seymour

[1] Eisner (2002), pp. 19–22.
[2] Seymour (1995), pp. 101–102.
[3] Eagleton (2003).
[4] The sections are numbered for editorial convenience. Sections II1–40 are often considered to be the core of the second part, with II41–46 labelled 'supplementary conclusions'. However, Edgar Laird explains that sections II37–39 might also be considered as 'free floating fragments' which were added to the core of II1–36: see Laird (1999). Seymour (1995), p. 103, claims that, although the situation is somewhat uncertain, Chaucer probably did not write the supplementary conclusions.
[5] Lydgate, *Fall of princes*, 1.293–295, quoted by Eisner (2002), p. 22.
[6] See Eagleton (2003).
[7] Eisner (2002), considers thirty-two manuscripts, to which we add London, Royal College of Physicians, MS 358 (see Eagleton, 2003) and London, British Library, MS Sloane 446 (included by Seymour, 1995, and Reidy, 1988, pp. 1193–1195, among the manuscripts of the *Treatise on the astrolabe*, but discussed in detail by neither).

states that 'The format of the original work seems to have been a small illustrated booklet suitable for a literate youngster's physical and intellectual grasp'.[8] However, it is clear that the *Treatise on the astrolabe* quickly had a wider readership than Lewis—Chaucer's son, to whom the work is addressed—and Sigmund Eisner outlines the ways in which Lewis also serves as a rhetorical device: his tender years give Chaucer a reason for writing in English.[9]

As much as the physical nature of the manuscripts varies, so their content changes. Several scholars have considered the relationship between manuscripts, grouping similar versions of the text together, and usually distinguishing two main groups—called α and β.[10] The α group includes ten manuscripts, and the β group eight, leaving sixteen copies that do not fall clearly into either group. Some of these are conflated manuscripts deriving from more than one ancestor, while others show features of both groups, but no direct evidence of having more than one ancestor.

These conflated copies, along with the number of surviving manuscripts, and the variation between the versions of the text contained in them, means that working out the relationship between the many copies is extremely difficult. With the exception of Derek Price, who prepared an unpublished stemma showing about half of the surviving manuscripts,[11] most work on the relationship of the manuscripts has been descriptive, relying on grouping of the copies, and analysing which the closest relative of a particular copy might be, but stopping short of giving a detailed stemma showing the relationships between the manuscripts and the development of the *Treatise on the astrolabe*. In addition, all scholars have excluded some manuscripts from consideration, either because they are too fragmentary, too abbreviated, or too conflated to be easily considered by traditional methods.

One feature identified by Eisner, and used to place copies of the *Treatise on the astrolabe* into the α or β groups is the spelling of the word 'zenith' where it appears, which in β group manuscripts tends to have been replaced with some form of the word 'signet'.[12] In addition, Eisner uses the fact that some copies have additions, deletions and reordering of sections to group the surviving copies. Manuscripts in the β group tend to order the sections of part II as II1–12, 19–21, 13–18, 22–35, and then add some or all of sections II36–45 and II41ab, 42ab, 43a (the so-called 'spurious conclusions'). β group manuscripts also tend not to include II38–40, but do have sections after II41, whereas α group manuscripts, if they have sections after II36 at all, have only II37–40, and none of the sections from II41 onwards. The third method used by Eisner to classify and group the manuscript copies is the number of shared variants a pair of manuscripts has. Table 1 summarises the results of the study so far of the relationships between copies.[13]

Added to these manuscripts is a fragment not known to Eisner which preserves only section II36; the place of this copy in relation to the other copies has not previously been studied. It is an early copy of the *Treatise on the astrolabe*, perhaps dating from the second

[8] Seymour (1995), p. 105.

[9] Eisner (1985). See also Laird (1999).

[10] Eisner (2002), pp. 46–47, gives a concise summary of the work of P. Pintelon, Derek de Solla Price, John Reidy and Michael Seymour on how the manuscripts relate to each other. See also Skeat (1872), Pintelon (1940), Reidy (1988), and Seymour (1995).

[11] Price (N.d.).

[12] Eisner (2002), pp. 47–48.

[13] Manuscript sigils and discussion of the relationships between them is summarised following Eisner (2002), pp. 40–81. In addition, we have examined all manuscripts except for Ph (studied as a photostat copy), and Bu, T, and N (which are currently in private collections).

Table 1
Manuscript copies and the first printed edition of the *Treatise on the astrolabe*

Group α		
Ab (Aberdeen, University Library, MS 123)	I4–23, II1–23, II25–36 (losses probably due to damage to the manuscript after it was copied)	Perhaps related to Bl2 and N, but less close to them than to a common ancestor
Bl1 (Oxford, Bodleian Library, MS Bodley 619) **Ad2** (London, British Library, MS Additional 29250)	I1–23, II1–40	Pintelon groups Bl1 with Br; Eisner argues that this is incorrect and agrees with Price that Ad2 was perhaps copied from Bl1
Br (Brussels, Bibliothèque Royale Albert 1er, MS 4862-69) **Rl1** (Oxford, Bodleian Library, MS Rawlinson D913)	I1–23, II1–40	Rl1 shows some links to Pl and Dd2
Dd1 (Cambridge University Library, MS Dd.3.53) **M1** (Oxford, Bodleian Library, MS E Museo 54)	I1–23, II1–40[14]	Probably copied by the same scribe; Dd1 probably a corrected version of their common source
Hg (Aberystwyth, National Library of Wales, MS Peniarth 359) **R1** (Cambridge, Trinity College, MS R.15.18)	I1–23, II1–33, II35–36, II34	Show some influence from **β** manuscripts; Price related R1 to Sl2 but Eisner argues it is closer to Hg
R2 (Cambridge, Trinity College, MS R.14.52)	II37, II40, II39, II38	Over half of variant readings are unique, but some shared with Dd2, Pl, Rl1 or Dd1
Group β		
A2 (Oxford, Bodleian Library, MS Ashmole 360) **Ad1** (London, British Library, MS Additional 23002) **Dg** (Oxford, Bodleian Library, MS Digby 72) **Eg** (London, British Library, MS 2622)	I1–14 (Dg includes I15–23 from an **α** source), II1–2, II2a, II3–12, II19–21, II13–18, II22–35, II41–45, II41ab, II42a (omitted in A2), II43a, II42b, II36–37	A2 and Dg share many variants and perhaps had an exclusive common ancestor; Ad1 perhaps copied from Eg
J (Cambridge, St. John's College, MS E.2)	I1–23, II1–12, II19–21, II13–18, II22–35, II41–44, II41a, II41b, II42a, II42b, II43a, II36–38	Closest to A2, Ad1, Dg, Eg and M2 but omits section II45
M2 (Oxford, Bodleian Library, MS E Museo 116)	II31–35, II41–45, II41a, II41b, II42a, II43a, II42b, II36–37	Only sections after II31 present; somewhat related to Dg, Eg, Ad1; also shares some variants with Rl2 and Ph where all have text
Ph (London, Institute of Electrical Engineers, MS Thompson 1) **Rl2** (Oxford, Bodleian Library, MS Rawlinson D.3)	I1–23, II1–12, II19–21, II13–18, II22–35	Price suggests that Ph is the direct ancestor of Rl2; this subgroup of the **β** group probably developed independently from the Dg, A2, Eg and Ad1 subgroup

Other manuscripts

A1 (Oxford, Bodleian Library, MS Ashmole 391) **Sl2** (London, British Library, MS Sloane 314)	II1–23, II1–36 (A1 lacks II4–24 through loss of a quire)	Do not have the ***β*** group section ordering but do show influence from a ***β*** source; marginal notes by Stevins show that he used Sl2 when compiling Sl1
A3 (Oxford, Bodleian Library, MS Ashmole 393)	I1–23, II1–3	Perhaps related to Bu and T; this copy is much abbreviated
Bl2 (Oxford, Bodleian Library, MS Bodley 68)	I1–23, II1–36	Bl2 not linked to Hg despite missing the same parts of the prologue and part II; Price suggests a close relation to R1 and Sl2 and a less close one to A3 and T; shares the most variants with N, and the next most with Ab
Bu (formerly Marquess of Bute, MS 13 (A19))	Bu: I1–17, I21–23, II2, II308, I17–21, II38–40, II2a, II3–11, II41–43, II41a, II41b, II42a, II43a, II42b	T with Dg, Eg, Ad1, A2 subgroup up to I14, after that point it has many unique variants from them; Bu rarely with either T or Ad1 subgroup in supplementary conclusions, indicating that it perhaps had more than one source
T (Tokyo, Takamiya, MS 9)	T: I1–23, II2–3, II38–40, II2a, II3, II4–35, II41–45, II41a, II41b, II42a, II43a, II42b, II36–37	
Cp (Cambridge, Corpus Christi College, MS 424)	II16–36	Contains only sections II16 to II36; most closely related to T; Price puts this manuscript with R1, A1 and Sl2
Dd2 (Cambridge University Library, MS Dd.12.51) **Pl** (Columbia University Library, MS 254)	I1–23, II1–43	Share some variant readings with Hv; the main exemplar of TH1 very close to this group
Hv (Harvard, Houghton Library, MS English 920)	I1–23, II1–37, II41ab, II43a, II42ab	Shares some variant readings with Pl and Dd2; exemplars likely to have been idiosyncratic and no longer extant; Price suggested it may represent a missing link between ***α*** and ***β*** groups
N (Northumberland, Alnwick Castle, MS 460)	I1–23, II1–13, II36, II35, II34, II33, II16–19, II26–32, II14, II35 (duplicate), II41–45, II41a, II41b, II42a, II43a, II42b, II36 (duplicate) II37	Probably two ancestor manuscripts: part I and most of part II (up to II14) from an ***α*** manuscript, perhaps similar to Ab; II35–36 (duplicates), II37, II41–45, II41ab, II42ab, II43a from a ***β*** manuscript
Pu (Aberystwyth, National Library of Wales, MS 3567B)	(as TH1) II1–12, II19–21, II13–18, II22–43, II41ab, II42a, II43a, II42b	Shares many variants with printed editions, particularly TH3 (1561 edition of the *Treatise on the astrolabe*, set directly from TH1)

(*continued on next page*)

Table 1 (*continued*)

Other manuscripts		
Sl1 (London, British Library, MS Sloane 261)	I1–23, II1–16, II18–34, II36–43, II41a, II41b, II42a, II43a, II42b	Use made by Stevins of Sl2 in preparing this manuscript; its main exemplar probably like TH1
TH1 (William Thynne, *The workes of Geffray Chaucer, newly printed/with dyuers workes whych were neuer in print before*, 1532 (STC 5068))	I1–23, II1–16, II18–34, II36–43, II41a, II41b, II42a, II43a, II42b	Thynne gathered many Chaucer MSS while preparing his edition; strongly related to Pl and Dd2 subgroup; weaker relationships with J, Rl2, Dd1, Ad1, Eg, Hv, M2, N, T
W (Aberystwyth, National Library of Wales, MS 3049D)	I1–23, II1–16, II18–34, II36–43, II41a, II41b, II42a, II43a, II42b	Derives from an early printed edition, perhaps TH2 (1542 edition of the *Treatise on the astrolabe*, set directly from TH1)

[14] Sections II35 and 36 have been displaced to the end in M1, whether by the scribe's use of a disordered exemplar, or his mishandling of the source.

quarter of the fifteenth century, and we have allocated it the sigil Rcp for the purposes of this study.[15]

There is, however, one manuscript which we will not consider in detail in this paper. London, British Library, MS Sloane 446 is an astronomical compilation from the mid sixteenth century, in which there are two short sections attributed to Chaucer. This copy is not discussed or listed among the manuscripts of the *Treatise on the astrolabe* by Eisner, but is allocated the sigil Sl3 by Reidy, and described as follows:

> This last [Sl3] contains Latin instructions for establishing the twelve houses, based on 2.36 and 37, and a Latin version (not a translation) of 2.23, both attributed to Chaucer; it includes a horoscope for April 21, 1547, and also has instructions for finding the height of the tower, slightly resembling those in 2.41, and probably based on them, but not attributed in the manuscript to Chaucer.[16]

Sl3 is a collection of material on astronomy and astrology, drawn from a range of sources. Much of the treatment of the astrolabe comes from Johannes Stöffler's 1524 *Astrolabium*, but there are two sections attributed to Chaucer which are, as Reidy states, related to sections of Chaucer's text, but which might be more properly described as being reworkings of those sections rather than translations into Latin. The section based on Chaucer's II23 has been altered such that the main example described is different, and the section based on II36–37 is much more astrological than are the corresponding sections of Chaucer. The section on measuring the height of a tower could as easily have come from Stöffler, or any number of other fifteenth- and sixteenth-century works on astronomical instruments, as from Chaucer, so we do not agree with Reidy that it is also related to the *Treatise on the astrolabe*. A transcription of the two sections attributed to Chaucer in Sl3 is in Appendix 1.

Because the sections in Sl3 have been significantly altered from those in Chaucer's *Astrolabe*, it was not possible to include it in the detailed analysis of the relationship between manuscripts that follows. These versions of sections from Chaucer's text are nonetheless significant, since they provide further evidence of the astrological turn that Chaucer's reputation took in the fifteenth and sixteenth centuries, despite his statements against judicial astrology in section II4 of the *Treatise on the astrolabe*. As Edgar Laird explains:

> The existence of that [late medieval scientific] culture, with its broad strain of astrologism, explains the continued currency and frequent recopying of the *Astrolabe*. The appeal of vernacular scientific writing seems to have been felt in diverse social and economic settings, and the *Astrolabe*, as variously presented in manuscript, could meet various demands that Chaucer himself may not have contemplated.[17]

Laird suggests that the unfinished state of the *Treatise on the astrolabe* made it particularly attractive to adaptation, rewriting, and reordering and encouraged scribes to rework the text to their own ends. He speculates that for this scientific audience, Chaucer as a person was less important than he was to his literary audiences, to whom he was becoming an iconic author.[18]

[15] See Eagleton (2003).
[16] Reidy (1988).
[17] Laird (1999), p. 158.
[18] Ibid., p. 159.

This paper reconsiders the relationship between copies of Chaucer's *Treatise on the astrolabe* by using new techniques derived from evolutionary biology. We begin by describing the method by which we have compared them. We propose a stemma showing all manuscripts but four fragmentary copies of the *Treatise on the astrolabe*, along with the first printed edition. From this, we suggest that some version of the diagrams present in nine of the manuscripts was in the original version of the text. Then, we place the fragmentary copies of the *Astrolabe* relative to other manuscripts by considering the stemmas generated by consideration of subsections of the text. Study of data from subsections of the text also allows us to analyse the nature of conflation in manuscript copies. Finally, we consider the impact this new understanding has on the authenticity of the so-called 'supplementary conclusions' (sections II 41–46, 41ab, 42ab and 43a, which appear in only some copies of the *Astrolabe*), and conclude with some suggestions about how the text of the *Treatise on the astrolabe* changed and developed during the fifteenth century.

2. Methods

In order to determine the dissimilarity between surviving manuscripts, we considered changes to the text that have occurred at the level of individual words: changes of spelling, additions and deletions.[19] From each section of the *Treatise on the astrolabe* a sample of the first ca. 200 words (or the whole section, if it is less than 200 words long) was entered, with the exception of the prologue, from which two sections of 200 words each were analysed. This gave us a main data set of 11375 words, taken from all sections of the text, a large enough sample that we are confident of the validity of the results drawn from it.[20]

The data set was based on Eisner's edition of the text, and on standardised Middle English spelling and dialect.[21] Drawing the data from Eisner's edition ensures that

[19] An issue of concern to the critics of the use of phylogenetics for the study of manuscript texts is that of the quality of the variants. According to this view, it is less important that two manuscripts share a particular variant reading in a particular place, than that the variant reading is a significant change that cannot easily be attributed to chance. However, there is no objective view of 'quality' on which we are to base such judgements, returning us once again to the subjective, unreproduceable study. For Lydgate's *Kings of England*, stemmata from different kinds of variants (e.g: insertions/deletions of important and unimportant words, word substitutions affecting/not affecting meaning) were all much more similar than would be expected by chance. See Spencer et al. (2004). This suggests that although some kinds of variant are more subject to convergence than others, all are transmitted along the same lines of descent. Low quality variants are therefore more likely to represent random noise than bias.

[20] See below for neighbour-joining and NeigborNet methods, and Appendix 2 for details of the data used and the distance calculations. One of the crucial factors for the reliability of a stemma is the size of the data set used to construct it. For Chaucer's *Miller's tale*, about 1000 words or short phrases is probably enough for a reliable stemma using neighbor-joining, and 500 is not much worse. See Spencer, Bordalejo, Robinson, & Howe (2003). We do not have equivalent results for NeighborNet, but we expect it to perform at least as well, particularly because NeighborNet is designed to reveal conflicts in the data that could lead to incorrect reconstructions using neighbor-joining. We therefore think that the main data set (11375 words) is large enough for reliable results. We cannot be so confident about the section-by-section analyses, which are based on data sets of between 200 and 1000 words each.

[21] We assume throughout that transcription errors are rare enough to be ignored. A few such errors in Eisner's edition have been identified (Rand, 2005), but the mistranscriptions noted by Rand represent less than 1% of the data in the sections checked. Recent work has shown that large numbers of transcription errors have the potential to affect stemma reconstruction in surprising ways (Susko, Spencer, & Roger, 2005), but the effect of a small number of transcription errors is likely to be negligible if there are many correctly transcribed variants.

any decisions about what is a significant variant, and what a spelling or dialect difference, were made by one editor, which lends some consistency to the data set. Although this standardisation excludes some potentially important variation, it ensures that the differences between copies are not overestimated due to their different dialects, and ensures that we do not group manuscripts together simply because they share dialect.[22]

We first constructed matrices of pairwise distances among manuscripts. For each pair, we first excluded any reading that was illegible, affected by physical damage or otherwise undeterminable in either manuscript. We then counted the number of remaining readings different between the two, and divided this by the total number of remaining readings to obtain an estimate of pairwise distance. This has range 0 to 1 and is a reasonable estimate of the number of changes of reading separating the manuscripts, so long as distances are small enough that multiple changes at the same location are rare (as was the case here). We treated insertions or deletions of single words in the same way as substitutions of one word for another, and treated insertions or deletions of more than one adjacent word as being single events. We treated places where a manuscript was lacking words because of damage or illegibility, and omissions of entire sections, as missing data. One copy (N) had some sections that appeared twice, and for this we counted both occurrences of these sections equally when estimating distances, allocating the sigil N2 to the duplicate sections, in line with Reidy's discussion of this manuscript.[23] From these raw distances we constructed separate matrices for each section, and also a matrix for the main data set, calculated by summing the number of words at which a pair was different and the number of words at which neither had missing data over all sections.

We reconstructed stemmas from the distance matrices using NeighborNet, a recently developed method for representing evolutionary relationships among living organisms.[24] Methods from evolutionary biology are increasingly used to reconstruct stemmas,[25] as well as in related fields such as linguistics[26] and archaeology.[27] Reconstructing a stemma is difficult because the number of possible stemmata is very large. For example, with ten surviving manuscripts, the number of possible rooted tree-like stemmata is more than 100 billion.[28] A moderate-sized text tradition may contain thousands of variants, all of which may contain useful information. The task of choosing the best possible stemma is therefore too large for an unassisted human, and even a fast computer cannot usually evaluate every possibility. Several textual critics have proposed computer algorithms for stemma reconstruction.[29] These methods have not been rigorously analyzed, so we do not know how well they perform. In contrast, many phylogenetic methods have been studied in great detail by mathematicians, biologists and computer scientists, and have been tested on both simulated[30] and real

[22] Most stemmatic analyses in Middle English ignore spelling, dialect and punctuation differences, as these are believed to be more likely to reflect scribal habits than stemmatic relationships (for example, Robinson 1997).
[23] Reidy (1988).
[24] Bryant & Moulton (2002).
[25] See, for example, Lee (1989), Robinson & O'Hara (1996), and Robinson (1997).
[26] See, for example, Gray & Atkinson (2003).
[27] See, for example, Tehrani & Collard (2002).
[28] Flight (1990).
[29] For example, Dearing (1974), or Flight (1994).
[30] Nei (1991).

data,[31] including text traditions.[32] As far as we know, A. R. Lee was the first to apply phylogenetic methods to a real text tradition,[33] although the similarities between textual criticism and phylogenetics had already been discussed.[34]

Using an algorithm rather than editorial judgement for stemma reconstruction has been harshly criticised. For example, Ralph Hanna used a quotation from Housman as an epigraph to a paper on the stemma of Chaucer's *Wife of Bath's prologue*:

> They must have a rule, a machine to do their thinking for them. If the rule is true, so much the better; if false, that cannot be helped: but one thing is necessary, a rule.[35]

What Hanna sees as a weakness, we see as a strength. The alternative to a clearly defined algorithm is subjective judgement that cannot be evaluated, justified or reproduced.

Neighbor-joining is a popular, fast and fairly reliable method for reconstructing phylogenetic trees from distance data.[36] It progressively combines manuscripts into larger and larger groups in such a way that, at each step, the smallest possible amount of change in readings in required. The neighbour-joining algorithm can be briefly summarised as follows:

> Define a pair of neighbours as a pair of manuscripts connected directly to the same hypothetical ancestral manuscript in an unrooted, bifurcating tree. Unrooted here means that the direction in which time runs along edges is not known; bifurcating means that every extant manuscript is directly connected to one hypothetical ancestor, and that every hypothetical ancestor is directly connected to three other extant or hypothetical manuscripts. Define an unresolved node as a hypothetical ancestor to which more than one extant manuscript is connected.

1. Start with a star-like topology in which all extant manuscripts are connected directly to a single central hypothetical ancestor.
2. While the tree contains extant manuscripts connected to unresolved nodes, consider each unresolved node and:
   a. Try making two of the extant manuscripts attached to this node neighbours. Do this by connecting a new hypothetical ancestor to the unresolved node, and connecting the pair of extant manuscripts to the new hypothetical ancestor.
   b. Estimate the lengths of the edges connecting all manuscripts by a method similar to least-squares.
   c. Do this for each such pair in turn. Choose the pair for which the sum of edge lengths over the whole tree is the smallest, and replace this pair with their new hypothetical ancestor.

Neighbor-joining is a consistent method in the sense that, given sufficient data, it will estimate the relationships among manuscripts correctly if those relationships are tree-like and the distances among manuscripts have been estimated appropriately. However, we cannot

[31] Hillis, Bull, White, Badgett, & Molineux (1992).
[32] Robinson & O'Hara (1996), and Spencer, Davidson, Barbrook, & Howe (2004).
[33] Lee (1989).
[34] Poole (1974), Platnick & Cameron (1977) and Cameron (1987).
[35] Hanna (2000): the quote is on p. 163, and cited from *D. Ivnii Ivvenalis Satvrae* (A. E. Housman, Ed.) (London: E. Grant Richards, 1905), pp. xii–xiii.
[36] Saitou & Nei (1987), Gascuel (1994).

assume that the true relationships will be tree-like in a potentially conflated data set such as the *Treatise on the astrolabe*. We therefore used a related method, NeighborNet,[37] which does not assume each extant manuscript had only one direct ancestor. NeighborNet has some similarities to the split decomposition method, which has been used to analyse manuscript traditions including Chaucer's *Wife of Bath's prologue*.[38] Like split decomposition, NeighborNet can represent contamination and conflicting information in the data as sets of parallel edges, with each pair of edges corresponding to a grouping of manuscripts for which there is some support. The lengths of edges are proportional to the amount of change occurring along them. NeighborNet is better than split decomposition at resolving the relationships among large numbers of manuscripts. Finally, when tested, NeighborNet correctly resolved most of the relationships among artificial manuscripts with a known true stemma.[39] The NeighborNet algorithm differs from the neighbor-joining algorithm in one important way. Neighbor-joining immediately replaces a selected pair of manuscripts by their hypothetical common ancestor (at step 2c, above), whereas NeighborNet retains the pair and their ancestor, and does not perform a replacement until a manuscript has been chosen as a member of a second pair. It then replaces both pairs with a pair of linked hypothetical ancestors. This change results in a method that gives a network instead of a tree, if the true relationships are not tree-like.

It has often been remarked that there is a close analogy between cladistic phylogenetic methods (such as maximum parsimony) and the usual approach to stemma reconstruction.[40] In both cases, it is assumed that minimising the number of changes (for example, in readings or morphological features) required to explain the observed data patterns is an appropriate goal. This is usually justified by appeals to Occam's Razor, although the validity of such arguments in not self-evident.[41] Neighbor-joining and NeighborNet are not cladistic methods, since they are based on pairwise distances rather than on discrete characters, and do not minimise the total amount of evolutionary change. Their intellectual forebears are the numerical taxonomy methods that later formed the basis for a wide range of mathematical clustering methods.[42] Interestingly, important foundational work on distance methods in phylogenetics was done in the context of manuscript classification.[43]

As we outlined above, the majority of previous work on the relationships between copies of the *Treatise on the astrolabe* has been based either on significant differences like the spelling of 'zenith' as some form of 'signet' by ***β*** group manuscripts, the presence or absence of sections, or the raw number of shared variants that a pair of manuscripts has. While these methods are useful, analysis using NeighborNet allows a more sophisticated treatment of the data, providing graphical representations of the relationships among manuscripts. An important feature of this method is that it is repeatable—NeighborNet uses clearly defined algorithms, so if another manuscript of the *Treatise on the astrolabe*

[37] Bryant & Moulton (2002).
[38] See, for example, Barbrook, Howe, Blake, & Robinson (1998), and Huson (1998).
[39] However, we expect to find support for conflicting relationships even in the absence of contamination. See Spencer et al. (2004).
[40] For example, Lee (1989), Platnick & Cameron (1977), Cameron (1987), and Salemans (1996).
[41] For a brief review, see pp. 138–145 in Felsenstein (2004).
[42] See ibid., pp. 123–124.
[43] Buneman (1971).

were found, it would be straightforward for us or other scholars to use the same methods, and the same data set, to place that new fragment with respect to the copies currently known. Finally, the method we have used measures all differences between two copies of a text, significant or not. Therefore, if our results agree with other scholars' interpretations of significant differences, we can be more confident about both the subjective and the algorithmic results.

3. The relationship between manuscripts of the *Treatise on the astrolabe*

We constructed the main stemma using combined data for all sections (a total of 11375 words): Appendix 2 gives the distance matrix for the main stemma. We omitted four fragmentary manuscripts—Cp (contains sections II16–36), M2 (II31–37, 41–45), R2 (II37–40) and Rcp (II36 only)—each of which had no content in common with at least one other manuscript and therefore had undefined pairwise distances.[44] We later discuss section-by-section analyses that allow us to place these fragments, as well as examining the sources for conflated manuscripts and the printed edition in more detail.

There are several points to bear in mind when interpreting the stemma shown in Fig. 1. First, the extant manuscripts are represented by labelled nodes, and hypothetical relationships by edges. Unlabelled nodes where two edges meet represent hypothetical manuscripts with more than one descendant. Even if this reconstruction were correct, it would not represent every manuscript in the tradition. Each edge may represent a chain of copying events, but intermediate manuscripts with only one descendant are not represented. Lost manuscripts with no extant descendants will not appear at all. Second, the edges are drawn to scale. The length of each edge is proportional to the estimated probability of change per reading between the nodes at its ends. A long edge may represent either a single inaccurate copy or a chain of several intermediate copies. Third, the stemma is unrooted. We do not know the direction of change along each edge. We can imagine placing the archetype of the whole tradition at any point along any edge, with time flowing outwards towards the extant manuscripts from this point. Fourth, ambiguities in relationships among manuscripts appear as parallelograms. In such cases, the pair of long edges represents the hypothesised relationship that is more strongly supported. The pair of short edges represents an alternative relationship with weaker support. Such ambiguities may arise from contamination or from large numbers of convergent changes.

Fig. 1 agrees to a remarkable extent with Eisner's summaries, based mainly on Reidy's groups, of the relationships between the manuscripts. In two cases we agree with him against Price or Pintelon: in placing Bl1 and Ad2 together rather than with Br, and Hg and R1 together rather than placing R1 with Sl2. However, there are differences, notably that our stemma places manuscript J with Ph and Rl2 rather than with the core of the β group (A2, Ad1, Dg and Eg). However, Eisner does point out that this group is not particularly close to J when the number of shared variant readings is considered. One manuscript, A3, appears in a different place on the stemma than one might expect from previous study, but this is a difficult copy of the text, which will be considered in more detail in a

[44] The text in these manuscripts begins after II3, when the copies like A3 and W have no further text. This means that it is not possible to compare them, since NeighborNet cannot deal with a matrix in which there are unknown values.

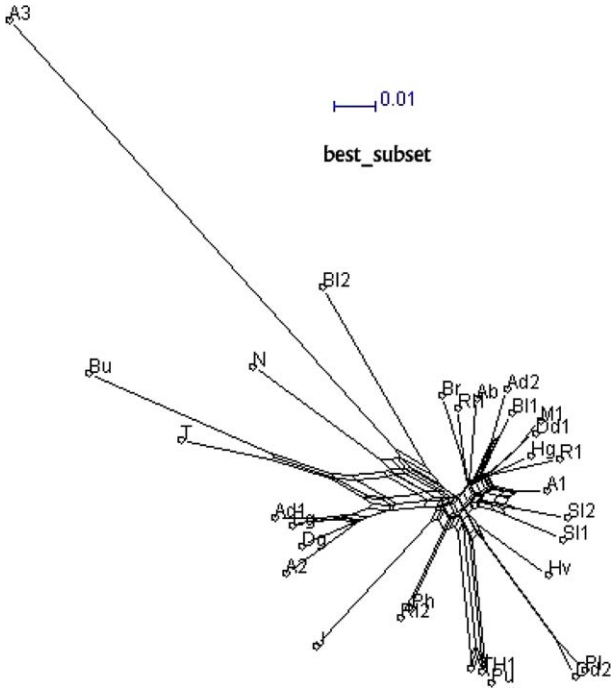

Fig. 1. Stemma from main data set of 11375 words.

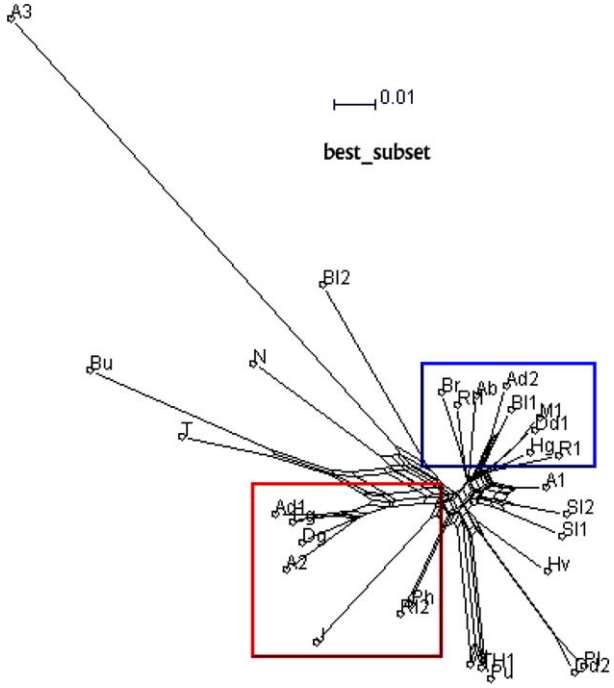

Fig. 2. Stemma from main data set, with α and β groups marked.

later section of this paper—its position on this stemma might not be reliable because of the heavy abbreviation of its version of the text. It appears with Dd2, Hv and Pl, and not particularly close to Bu and T, to which it is thought to be linked.

In spite of these differences, the many similarities mean that this stemma can be considered a reliable representation of the relationships between copies. Despite the fact that we have considered all differences, significant or not, our stemma shows the same manuscripts in the **α** and **β** groups as does Eisner's classification based primarily on significant variant spellings and on the section ordering (in Fig. 2, the **α** group manuscripts are outlined at the top right of the diagram, while the **β** manuscripts are outlined at the bottom left).

4. Diagrams in the *Treatise on the astrolabe*

Eleven surviving copies of the *Treatise on the astrolabe* contain diagrams illustrating the text, or, in the case of Br, spaces where diagrams should have been inserted. Marking them on the stemma generates the striking pattern shown in Fig. 3.

All copies containing diagrams are in the **α** group of manuscripts (Br, Rl1 Dd1, M1, R1, Hg), or in conflate manuscripts with strong links to the **α** group (A1, Sl1, Sl2, Hv, Bl2). However, they are not all directly descended, one from another. Two manuscripts—Bl2 and Hv—contain diagrams for which there is no evidence of a direct link to the text. Hv has on f. 6 a diagram showing the *mater* of an astrolabe, and Bl2 has a sketch of a shadow square on f. 14v. It is possible that these diagrams derive from a more completely illustrated version of the text, but it may also be that these copies were illustrated indepen-

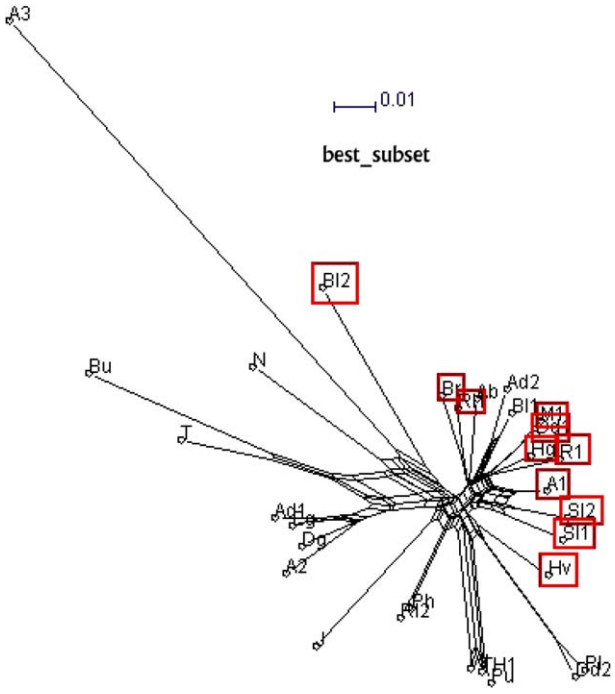

Fig. 3. Stemma from main data set, with manuscripts containing diagrams marked.

dently of the diagrammatic tradition specific to Chaucer's *Treatise on the astrolabe*, of which the other nine copies provide evidence.

The remaining nine manuscripts contain versions, of differing completeness, of what was probably the original set of illustrations for the *Treatise on the astrolabe*. All are close to one another and all are in copies that are in group α, or which were based in part on an α manuscript. Because of this, we can see that the ancestor of the α group probably had a set of diagrams, and—because most modern scholars think that the archetype was like the ancestor of the α group—so did Chaucer's original version of the *Astrolabe*. This view is supported by the many references in the text to them. Sections of the text in many manuscripts end with a statement 'lo here the figure' referring the reader to the images, and the striking similarity between most of the sets of diagrams.

5. Fragments and incomplete copies

Despite the detail given by this stemma, it was not possible to include four fragmentary manuscripts on it.[45] Therefore, we considered subsections of the data in order to locate these copies relative to the more complete manuscripts. The smallest amount of surviving text is in Rcp, which contains only section II36, the first quire(s) of the manuscript having been detached at some point in its history after it was copied. In addition, the text has been partially erased, meaning that only eighty-three words of the 245 words in the data set for II36 are legible in Rcp. This means that the stemma generated is much less reliable than

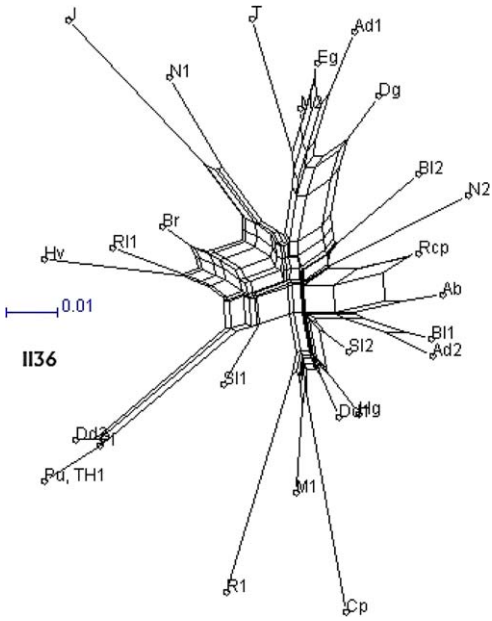

Fig. 4. Stemma for section II36.

[45] See Appendix 2.

those based on larger data sets, but in this case is important since it allows us to tentatively place Rcp in relation to other copies (Fig. 4).

Rcp appears to be most similar to Ab, in the α group, and also with some relationship to Bl1 and Ad2. Significantly, when Rcp was listed in a fifteenth-century booklist, it was described as 'Milk et breed'. Some variant of this subtitle 'Bread and milk for children' is also found in manuscripts Bl2, Ab, Bl1 and M1, all of which appear close together on this stemma.[46]

Also present in II36 is manuscript Cp, which contains only sections II16 to II36. These sections were considered; the stemma for II23–25 is shown in Fig. 5 (data set of 552 words).

This diagram shows Cp with manuscript T, related most closely to the α group, but in particular to manuscripts Hg and R1. This relationship is also seen in other subsections, summarised in Table 2, particularly in II18–19, II23–25 and II36. There is also a strong link between Cp and manuscripts Ab, Bl2 and N, particularly in II15–17, II20–22, II26–28, II29–30 and II35:

As Eisner points out, in terms of the number of shared variants Cp is closest to manuscript T, and in sections II16 to II30 this is certainly the case. However, after this point T shifts to a different exemplar, while Cp continues to show strong links to the α group (Dd1, Hg, M1 and R1) as well as to Ab, Bl2 and N.[47] In other words, Cp and T are only together where T is with Hg and R1 or Ab, Bl2 and N. Cp should therefore be classified as an α group manuscript, along with Ab, Hg and R1, rather than as a conflated manuscript because of its apparent association with T. In this case, consideration of the raw number

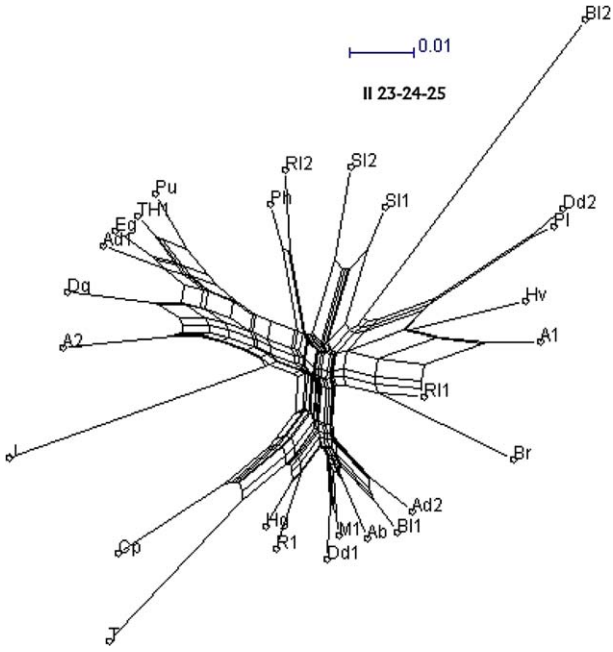

Fig. 5. Stemma for sections II23–24–25.

[46] See Eagleton (2003).

[47] T, J, and N are discussed later in this paper. N is at this point based on an α group exemplar similar to Ab.

Table 2
Closest manuscripts to Cp, from section-by-section analysis

II15–17 (487 words)	II18–19 (267 words)	II20–22 (347 words)	II23–25 (552 words)	II26–28 (584 words)	II29–30 (420 words)	II31–34 (619 words)	II35 (221 words)	II36 (245 words)
with T, related to J and Bl2	with T, related to Hg	with T, related to Bl1 and Ad2, and Ab	with T, related to Hg and R1	with T, related to Ab, Bl2 and N	with T, related to Bl2 and N	closest to A1, Sl2 (T now with A2, Ad1, Dg, Eg)	with Ab and N	with Dd1, Hg, M1 and R1

of shared variants of Cp and T misleadingly suggests that Cp is a conflated manuscript, and we have been able to show that, although fragmentary, Cp witnesses the α version of the *Treatise on the astrolabe*.

Manuscript R2, first described by Edgar Laird,[48] contains only four sections—II37 to II40. These were considered in two groups: II37 (data set of 213 words), and II38–40 (data set of 685 words), and the resulting stemmas are shown in Fig. 6.

In both cases, R2 is clearly an α group manuscript, closest to Br and Rl1, but with variants from them that mean that they probably have some common ancestor rather than being more directly linked.

The final fragment that was not included on the main stemma is found in manuscript M2. This copy of the *Treatise on the astrolabe* includes sections II31 to II45 (but not II38 to II40, in common with core β group manuscripts), and its location with respect to other copies can be seen in the diagrams for II36 and II37 (Figs. 4 & 6). Table 3 summarises its closest relatives in each section.

From this, it is clear that M2 is a β group manuscript, close to Ad1 and Eg, but with links (especially in conclusions after II41—the so-called 'supplementary conclusions') in particular to manuscripts N and T which have by this point switched to a β group exemplar.[49]

6. Copying and conflation: scribes who used more than one manuscript source

These fragmentary copies are all relatively stable and do not seem to shift to a new exemplar in the sections present; other manuscripts are more complex and draw on several exemplars. In order to look at the sources of the conflated texts, we considered subsections of the main text. In order to have a basis for comparison, a group of stable copies was identified, and the section-based stemmas used to check that they do not obviously change exemplar anywhere. They are the α group (Ab, Ad2, Bl1, Br, Dd1, Hg, R1, Rl1) and the two main parts of the β group (Ph, Rl2; and A2, Ad1, Dg, Eg), and the other manuscripts were considered in relation to these stable copies.[50] The sample sizes for individual sections are small, so we cannot be as confident about these results as about the main stemma. We therefore briefly report only the main features of these analyses.

[48] Laird (2000).
[49] N and T are discussed later in this paper.
[50] The results of this analysis are summarised here. The authors would welcome queries from anyone who would like to see detailed results of the section-by-section analysis.

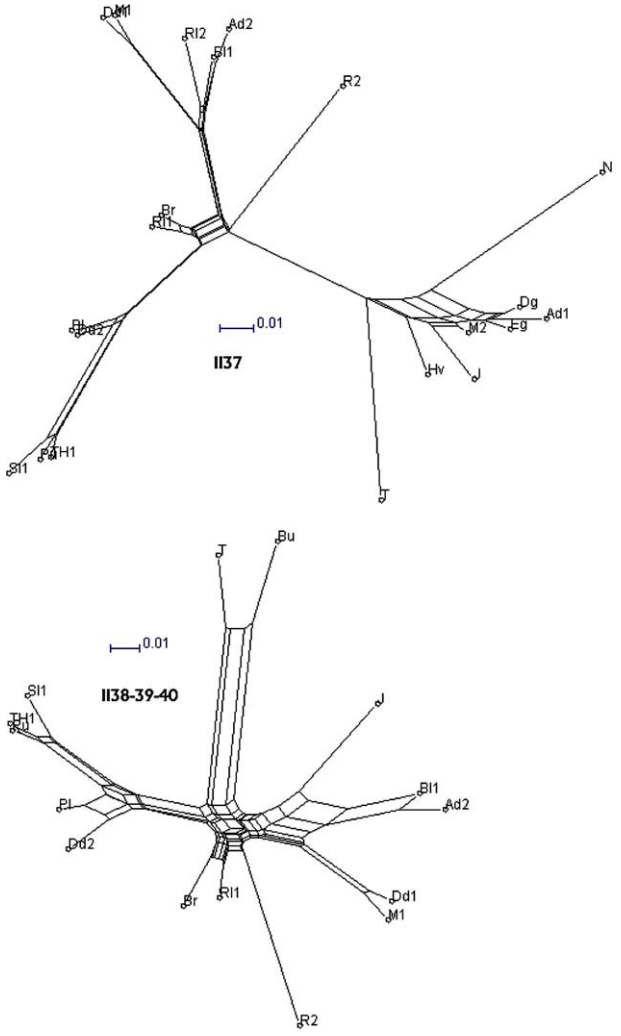

Fig. 6. Stemmas for sections II37 and II38–39–40.

Table 3
Closest manuscripts to M2, from section-by-section analysis

II31–34 (619 words)	II35 (221 words)	II36 (245 words)	II37 (213 words)	II41, 42, 43 (614 words)	II41ab, 42ab, 43a (715 words)	II44–45 (437 words)
β group, closest to Ad1 and Eg	*β* group, closest to Dg and Eg	*β* group, closest to Ad1 and Eg	*β* group, closest to Ad1, Dg and Eg	*β* group, closest to Ad1 and Eg, less close relation to N and T	*β* group, closest to Eg, less close relation to N and T	between Ad1 and Eg, and J, N and T

As seen on the main stemma, and summarised by Eisner, A1 and Sl2 share features of the α and β groups—they preserve the α group section ordering, but sometimes use a form of 'signet' for 'zenith'—but we did not see any obvious changes of source between sections, which would have been represented as changing locations of those manuscripts on the section-by-section stemmas. The β group manuscripts to which they are most closely related are Ph and Rl2 rather than the core of the β group. Drawing on Sl2 is Sl1, a sixteenth-century manuscript edition of the *Treatise on the astrolabe* whose editor, Walter Stevins, has left marginal notes in Sl2. Sl1 shows influence from early printed editions throughout, but especially in the prologue, sections II29–30, and from section II36 onwards (see the figures above for this change of location, and therefore of source).

Manuscripts Dd2, Hv and Pl often appear together on the stemmas, showing heavy influence from the α group in part I (especially copies like Br and Rl1), and the β group in part II (usually between Ph/Rl2, and Ab/Bl2/N). After section II35 Hv separates from Dd2 and Pl, drawing on a source like the core of the β group. Price suggests that Hv could represent a missing link between his group Ib (Dd2, Pl and Br) and group II (Ph, Rl2, T, J and the core manuscripts of the β group).[51] However, although Hv does show stronger links to the β group than do Dd2 and Pl, this may have more to do with the probable change to a β group exemplar at the end of section II35 of Hv than with its being an intermediary between the two groups.

Bl2 and N are generally similar to Ab, but with some influence from the β group. In particular, N includes sections from II35 onwards from a different source from those in part I and part II1–36 (there is duplication of II35 and II36). This second source, given as N2 on the diagram for II36 (Fig. 4), is a β group manuscript similar to that on which T draws at the end of part II.[52]

Related to Ab, Bl2 and N is manuscript J. It is usually put into the β group since it has the characteristic displaced order of conclusions. However, unlike other manuscripts in the β group, J does not include II45, but does include part of II38 (II38–40 are omitted by copies in the core of the β group). Eisner suggests that J might represent an early version of the β text, and that later scribes missed out the fragment of II38. This is also the situation depicted in the stemma proposed by Price, who placed J with Ph and Rl2 between Dd2 and Pl (conflated manuscripts drawing on the α group), and the main β group (A2, Ad1, Dg, Eg). However, Eisner argues that evidence against this hypothesis is provided by the fact that J has a higher rate of variants than the core of the β group, and that many of these variants are not shared with the β group.[53] We suggest instead that the high rate of variants in J indicates that it is a copy deriving from some common ancestor of it and the core of the β group, lacking most of II38–40 but not yet having acquired II45. Our analysis agrees with Price in that J is much closer to Ph and Rl2 than to the core of the β group, but we can also see influence from or features of manuscripts like Ab, Bl2 and N, and Pl and Dd2. The link to Ab, Bl2 and N is particularly strong in II3–4 and II7–9, while the relationship with Dd2 and Pl is stronger in II15–20, but it is unclear whether this indicates that J is switching sources (and therefore draws on more than one exemplar) or whether it is simply intermediate between the α and β groups. Our analysis, summarised in Table 4, agrees with Price in that J is much closer to Ph and Rl2 than to the core of the β group.

[51] Price (N.d.).
[52] T is discussed later in this paper.
[53] Eisner (2002), p. 66.

Table 4
Closest manuscripts to J, from section-by-section analysis

Prol(a) (225 words)	Prol(b) (203 words)	I1–4 (266 words)	I5–8 (465 words)	I9–11 (274 words)	I12–14 (218 words)	I15–17 (445 words)	I18–20 (277 words)	I21–23 (332 words)
Ph/Rl2	Bl2, also Hv/Pl	N/Ab	Bl2, also N	Hv and Pl/Dd2 but distant	[Data unclear]	Ph/Rl2	N/Bl2, also A3/Bu/T	Ph/Rl2

II1–2 (602 words)	II3–4 (433 words)	II5–6 (409 words)	II7–9 (284 words)	II10–11 (311 words)	II12–14 (428 words)	II15–17 (487 words)	II18–19 (267 words)	II20–22 (347 words)	II23–25 (552 words)
near Ph/Rl2 & Bu/T/N/A3/Bl2	N/A3/Bl2	β group, also Ph/Rl2	Ab/N, also Bl2	Hv, also β group	N/T/Bl2	Bl2, also T/Cp & Pl/Dd2	Hv & Pl/Dd2	Ph/Rl2 & β group	β group, also Ph/Rl2

II26–28 (584 words)	II29–30 (420 words)	II31–34 (619 words)	II35 (221 words)	II36 (245 words)	II37 (213 words)	II38– 40 (685 words)	II41, 42, 43 (614 words)	II41ab, 42ab, 43a (715 words)	II44–45 (437 words)
β group, also Ph/Rl2	Ph/Rl, also β group	near Ph/Rl2 & Sl1/Sl2	β group & Hv/N2/T & Sl2	N	β group & Hv & N2	Bl1/Ad, also Dd1/M1	Bu	Bu	N/T

Table 5
Section ordering in manuscripts Bu and T

Bu	II2		38*	I17*	18 to 20	21*	II38*	39 to 40	2a[54]	3*	4 to 11	41 to 43		41ab, 42a	43a	42b		
T	II2	3*					II38	39 to 40	2a	3*	4 to 35	41 to 43	44 to 45	41ab, 42a		42b	36 to 37	

[54] Section 2a is a spurious passage—see Eisner (2002), p. 173. It is present only in manuscripts A^2, Ad^1, Dg, Eg, Pu, W.

Manuscripts Bu and T are distinctive in that their section order is highly disturbed, the idiosyncratic order providing strong evidence of a link between them. Apart from the insertion in Bu of sections from part I in the middle of II38, perhaps because of displacement of a leaf, Bu and T share the distinctive section ordering until II12, when Bu is missing sections that are present in T. Section-by-section analysis confirms this link, showing Bu and T close together, although with some signs of influence from a text like J after II41, seen in Table 5 and Fig. 7. (* indicates that only part of a section is present).

T's ancestry is described by Eisner as follows:

> up through I.14 T is clearly with A2, Ad1, Dg, Eg (c.f., e.g., line 95). In the remaining sections of Part I, T is independent; most of its variants are unique, and the others are not shared in any particular pattern.[55]

We agree with Eisner that T is clearly drawing on a ***β*** group source at the start of part I, but section-by-section analysis somewhat clarifies the confusing situation after I14. In I15–17 T is between Dd2 and Pl, and Ph and Rl2, and then from I18–23 T is closer to the group including Bl2 and N, although often somewhat distant from these manuscripts. It is possible that this change of source at I15 resulted from the ancestor of T not containing

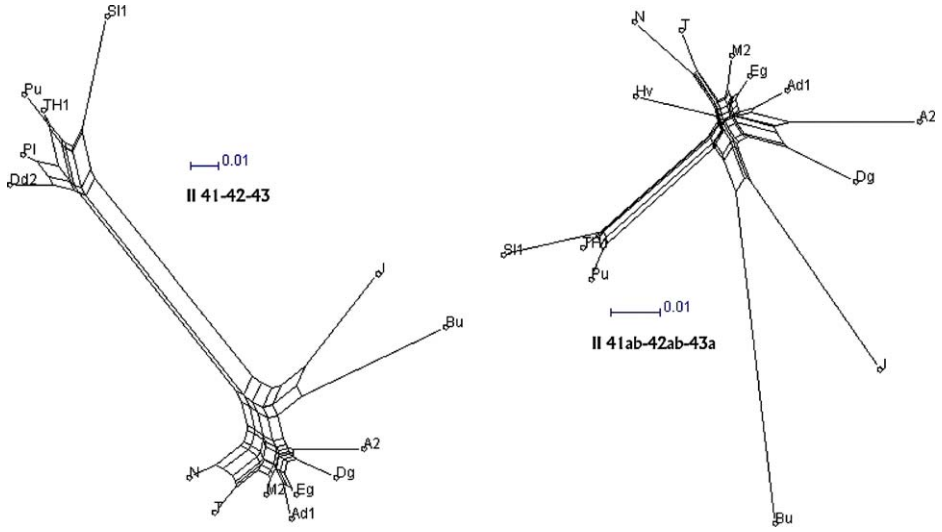

Fig. 7. Stemmas for sections II41–42–43 and II41ab–42ab–43a.

[55] Eisner (2002), p. 80.

Table 6
Closest manuscripts to T, from section-by-section analysis

II2	3*	II38–40	2a	3*	4	5–6	7–9	10–11	12–17	18–25	26–30	31–35	41 to 45	41ab, 42ab	36–37
Ab, Bl2 and N	**β**	Dd2 and Pl	**β**	**β**	**β**	Dd2, Pl and Hv	[unclear]	Hg and R1	Bl2, N	**α**, esp. Hg, R1	Ab, Bl2, N	**β**	**β**, esp. M2 and N2	**β**, esp. M2 and N2	**β**

sections I15–23 (in common with the other manuscripts of the core of the **β** group), so the scribe provided them from another exemplar.

In part II things then start to get very odd. We expected that the changes of source identified could be linked to the disturbance in the section order, and so constructed Table 6, showing the closest relations of T in the various sections, in the order in which they appear in T.

It is clear that T is drawing on at least four sources—an **α** group manuscript, perhaps like Hg and R1; a manuscript like Ab, Bl2 and N; another like Dd2 and Pl; and a **β** group source. However, it does not appear that the section reordering resulted from this combination of sources—except perhaps at the end, when all the sections from 31 onwards draw on a **β** group source, and show the typical **β** group features of omitting II38–40, and putting II36–37 at the end, after the supplementary conclusions (II38–40 are included earlier in manuscript T, but from a source similar to Dd2 and Pl). From the distance between Bu and T it is more likely that both were copied from some common ancestor, whether direct or indirect, rather than that Bu was copied directly from T. That ancestor already had the disordered conclusions, and drew on a multitude of sources. However, crucially, Bu and T include II2a, a conclusion also present in manuscripts in the core of the **β** group, providing further evidence that Bu and T are (at that point at least) closer to them than to, say, Ph and Rl2.

A3 was very difficult to place relative to other manuscripts because it is so heavily abbreviated, especially at the start—in fact, the abbreviation in sections I1 to I11 is so severe that we were not able to include data from those sections in the stemmatic analysis, despite the fact that A3 has text there. Table 7 summarises the results of our analysis.

Because of this heavy abbreviation, A3 is difficult to place on the stemmas, and it is important to combine the NeighborNet analysis with evidence from the sections included in A3 in order to get a clearer picture of its exemplar(s). The **β** group influence in sections I12–14 and II3–4, along with strong links to Bl2 and N between I15 and II2, is good evidence to support a link to manuscripts like Bu and T. Also significant is the

Table 7
Closest manuscripts to A3, from section-by-section analysis

I12-14 (218 words)	I15-17 (445 words)	I18-20 (277 words)	I21-23 (332 words)	II1-2 (602 words)	II3-4 (433 words)
Sl1 and β group but very distant	Bu/T	Bu/T, also J/N/Bl2	Bl2, also Bu/T but distant	Bl2/N, also Bu/T	N/Bl2, also J (Bu and T with A2, Ad1, Dg, Eg)

inclusion in A3 of section 2a, a feature shared by Bu and T, and the core manuscripts of the β group. Although A3 does not show the characteristic section reordering that Bu and T do, since its text stops at II3, it is noteworthy that the text of A3 stops at exactly the same place as does that of the first partial section 3 in T—just before the end of that section. So perhaps A3 or one of its ancestors was among the sources from which the conflated and reordered Bu and T manuscripts derive. If the scribe's main source stopped here, as A3 does, he might then have added in no particular order all the sections in other copies of the *Treatise on the astrolabe* which were not contained in his truncated main exemplar.

7. Conclusion: copying, completeness and conflation

This detailed section-by-section consideration of the conflated manuscripts of the *Treatise on the astrolabe* has given much new detail about the sources for, and copying of, many individual extant manuscripts. However, stepping back and looking once more at the text as a whole we see that the detailed analyses combine to show a striking concern among scribes for the completeness of the text they were copying. If an exemplar were deficient, there is evidence that in many cases the scribe consulted other copies in order to supply the missing sections. In only one case has this led to the duplication of sections (manuscript N includes two copies of sections II35–36), suggesting that the conflation process represented more than just the collection of all possible sections of the text. Instead, the scribes copying Chaucer's text seem to have had an idea of what the *Astrolabe* should contain, and they worked towards achieving that 'complete' version of the text, by consulting more than one copy. There is other evidence of this in the manuscripts: for example, in M1 there are notes next to each gap left in the text for a diagram, reading 'prima figura', '2a figura', and so on in sequence. At one point there is a gap in the text, but no diagram supplied, and the note reads 'hic deficit figura', showing that the copyist was aware that this diagram was missing from his exemplar.[56]

Scholars have usually assumed that the archetype text of the *Treatise on the astrolabe* was like that contained in manuscripts of the core of the α group, including sections 1 to 40 in part II, along with a full set of diagrams.[57] Manuscripts like Ad2 and Bl1, and Dd1 and M1 characterise this kind of text, and are the base manuscripts most commonly chosen by editors. Then, it is likely that the later sections (37 to 40) were soon dropped: copies like Hg and R1 represent an α manuscript that stops at II36. This can perhaps be linked either to the deterioration of textual quality that modern scholars have observed in II39–40, or to the fact that these sections either repeated material found earlier in the *Astrolabe*, or describe operations for which you do not

[56] In M1 sections II.35 and II.36 are displaced to the end of the text, rather than at their usual place in the sequence of sections. Kari Anne Rand Schmidt explains that a nineteenth-century rebinding error is responsible for the rearrangement of sections (Rand Schmidt, 1993, p. 59). However, the marginal notes planning the layout and order of the diagram make clear that the sequence of diagrams was wrong at the time of copying. The diagrams in the misplaced sections II35 and II36 are labelled as number 37 and 38, following in sequence from that in II40, which is labelled as number 36. What isn't clear is whether the scribe's source was corrupt, or whether he mishandled it.

[57] See Laird (1999, 2000) and Eisner (2002), pp. 40–43.

need an astrolabe.[58] The Ab, Bl2 and N group also finishes at section II36, but shows more links to **β** manuscripts, an influence seen more strongly in copies like A1 and Sl2, which have some examples of 'zenith' being spelt as 'signet'. Dd2, Hv and Pl, although close to the **α** group, include a number of the so-called supplementary conclusions, which are otherwise only present in **β** manuscripts, and Hv has the characteristic **β** section ordering.

Manuscripts J, and Ph and Rl2, seem to fall between these developments of the **α** tradition, and the core of the **β** group. All three copies have the **β** section ordering and spell 'zenith' with some form of 'signet', but J retains part of II38 and does not yet have II45 among its supplementary conclusions, while Ph and Rl2 stop at II35, perhaps linking them to the truncated versions of Chaucer's text in manuscripts like Hg and R1, and Ab and Bl2. Indeed, we have shown that there are links between J and Ab, as well as Ph and Rl2. Then, perhaps developing from manuscripts like the ancestors of J, Ph and Rl2, the core of the **β** group (manuscripts like A2, Ad1, Dg and Eg) has the **β** section ordering, all the supplementary conclusions,[59] as well as including section II2a, but omitting I15–23. These **β** group manuscripts, or one of their ancestors, then developed into the somewhat strange A3, Bu and T version of the text, likely to have been a truncated **β** manuscript that was 'repaired' by adding in the missing sections from a variety of sources.

In short, the **α** group came first, and the copying tradition then split into a number of groups, which developed some altered features, including the loss of sections after II37, the addition of supplementary conclusions, the reordering of part II, the loss of the end of part I, and the alteration of the spelling of 'zenith' to some form of 'signet'. These, perhaps thanks to the concern of some scribes with consulting more than one source, then crystallised once again into the characteristic **β** group version of the text. The speed of these developments is striking—along with some members of the core of the **α** group, some intermediate and conflated manuscripts (for example, J, Pl and Bl2) were probably copied in the first quarter of the fifteenth century. Bu and T were probably copied in the third quarter of the fifteenth century, providing a date by which the majority of the textual changes witnessed by the surviving copies of the *Treatise on the astrolabe* had been completed. Only a few decades after his death, Chaucer's text had morphed and changed significantly, due to scribes' errors in copying, to the truncation and reordering of exemplars and copies, and also to the scribes' concern with the completeness of the text, and the common practice of conflation. It is likely that sections II37 and II38–40 dropped from the textual tradition of the *Treatise on the astrolabe* relatively early, and then the so-called 'supplementary conclusions' II41–43 were added, followed by the 'spurious conclusions' of II41ab, 42ab and 43a. Finally, sections II44 and 45 were added, and the common ancestor of the core **β** group and the A3, Bu and T group gained section 2a.[60]

[58] II37 is similar to II36; II38 describes finding north from the sun using a circular piece of metal and compasses; II39 defines a number of geographical terms, and II40 can be performed with either an astrolabe or a globe. On these sections, see Laird (2000).

[59] Except II46, which appears only in Bl1 and Ad2 and was probably added independently of the process of development that we sketch here.

[60] Section II46, believed by most modern scholars to be spurious, appears only in manuscripts Bl1 and Ad2. Because only two manuscripts have this section, we did not analyse the text of this section.

However, we do not agree with Laird that in this process of conflation 'the person of the compiler was relatively unimportant'.[61] We suggest instead that Chaucer's name being attached to this text was of vital importance: that scribes were concerned with preserving Chaucer's text; and with compiling their own version of it that was as complete as possible. In his preface to Sl1, compiled in 1555, the editor Walter Stevins explains that this is exactly what he has done, stating:

> I thowght it a thinge worth my laboure if I cowlde sette it in better order and frame… Wherin I confesse that besydes the amendinge of verie many wordes I have displaced some conclusions, and in some places wheare the sentences were imperfite I haue supplied and filled them, as necessitie required… this haue I done, not challengine for my selfe, but renuncynge and leauinge to worthie Chawcer his due praise for this worke.

Because it is likely that the five planned parts of the *Treatise on the astrolabe* were never finished, Stevins's comment and the concern for completeness shown in our section-by-section analysis combine to provide an explanation for the rewriting and reordering of sections, and the addition of conclusions II41–46, 41ab, 42ab and 43a. After all, how does a scribe produce a complete version of a text that might never have been completed? It is a situation open to the inclusion of extra material and, given the astrological interests of Chaucer's readers, extracting and reworking as seen in the Latin sections in Sl3.[62]

By using phylogenetic methods we have helped untangle some of the conflation that resulted in the copying, recopying and rewriting of Chaucer's text during the fifteenth century. For the first time, we have been able to place all extant manuscripts onto figures, allowing a clear visual understanding of the relationships between them. In addition, the section-by-section analyses have clarified our understanding of the relationships between the fragmentary and complete copies of the text, and between scribes and their source manuscript(s). In our own modern way, we have reworked Chaucer's text in order to better understand the text as Chaucer wrote it, and as his readers received and reworked it.

Acknowledgements

Our sincerest thanks to all those who have discussed this paper with us, and offered their suggestions and ideas so generously, but especially to Dan Ransom, Barbara Bordalejo, Peter Robinson and Laurence Totelin.

Financial support enabling Catherine to travel to study the manuscripts was provided by grants from the Medieval Academy of America, the Royal Historical Society, and the Arts and Humanities Research Board.

Matthew Spencer was supported by the Leverhulme Trust (STEMMA project) and the Arts and Humanities Research Board.

[61] Laird (1999), p. 159.
[62] See Appendix 1.

Appendix 1. Sections attributed to Chaucer in London, British Library, MS Sloane 446

This manuscript includes information about astronomy and the astrolabe, drawn from a range of sources. It was probably compiled by William Parfey in the third quarter of the sixteenth century (ownership note on f. 1v, date of 1547 on a horoscope on f. 51v). Capitalisation and punctuation are preserved as in the manuscript, with | representing a line break, and editorial insertions placed in ⟨angled brackets⟩.

⟨f. 50v⟩ 30 Chaucer Anglus

Initia duodecim domorum celi facile| constituere: dicte horam equalem et per| eandem gradum ascendentem qui horoscopus| dicitur et cuspis est prime domus et Nadir| seu gradus oppositus cadens super horizon-|tem occidentalem est cuspis septime domus| et dicitur gradus descendens: Gradus tan-| gens Lineam medie noctis Quarte domui| initium donat Cuius obicitus lineam meridi-|anam possidens exordium decime domui appe-|rit (quam regiam dicimus) Deinde posito| almuri super gradum ascendentem computa|nde in limbo vsque ad lineam meridianam| ⟨51r⟩ gradus interceptes Quod partire in tres| equas partes Et sic habebis domos 12 et 11| et nadir earundem indicat .5. et .6. domos| Similiter divide gradus ab horoscopo vsque| ad lineam medie Noctis in tres partes et| habebis initium Secunde et Tercie domus| Et nadir earum indicat 8 et 9 domos et c.| n⟨ota⟩[63] Dominus Signi id est qui in eodem Signo domicilium habet| vel si ascenderet Signum .A.U erit dominus ascendentis| n⟨ota⟩ Dominator: Significator: suos gubernator arabico| dictus almuten est ille qui in Signo ac signi gradu| plures habuerit auctoritates essentiales.| n⟨ota⟩ Sunt autem auctoritates essentiales: domus: ex eltario triplicitas: terminus et facies| n⟨ota⟩ Dux thematis est qui in toto themate potiores| et plures obtinet auctoritates cum essentiales cum acci|dentales| n⟨ota⟩ Auctoritates accidentales sunt Medium Celi: 12 in| ascendent 11. ⟨Jupiter symbol⟩ ⟨sun symbol⟩ ⟨moon symbol⟩[64] in diues aut dominatores| particularum rerum non eliguntur cum domini totius sint

⟨f. 56r⟩

Chaucer

Ad Inuenstigandam poli altitu-|dinem in quauis habitatione| vespere observa in Stellifera noctis| claritate altitudinem maximam alicuius| stelle rotantis circa polum puta stelle que| est extrema Caude vrse maioris videlicet| quando eam inveneris supra polum in linea| meridiana Quam nota ad amussim/ deinde| circa crepusculum matutinum cape eius alti|tudinem minimam videlicet quando est in linea| medie Noctis sub polo et illam altitudinem| minimam adde medietati maioris altitudinis et| numerus proveniens paudet tibi altitudinem| poli illius loci/ Exempli causa sit maxima ele-|vacio gradu .62. minima autem 21 sume medie-|tatem .62. que est 31 quam adde minori numero| videlicet .21. Et provenient 52/ Et ut certius rem| habeas suspendatur perpendiculum inter| oculum et polum illudque clare ostendet quam ipsa| stella notata sit in linea meridiana supra| aut infra: Nam polus alteraque stella ra-|dient directe sub perpendiculo

[63] All nota marks in the margin.

[64] Symbols are used for Jupiter, the sun and the moon, as well as the zodiac signs.

Appendix 2. An example of the data used for comparison, from the beginning of section II37

Ad1	Than	ley	thy	label	over	the	degre	that	ascendith
Ad2	Than	ley	thy	label	over	the	degre	that	ascendith
Bl1	Than	ley	thy	label	over	the	degre	that	ascendith
Br	Than	ley	thy	label	al	the	degre	that	ascendith
Dd1	Than	ley	thy	label	over	the	degre	that	assendit
Dd2	Than	ley	thy	label	on	the	degre	that	ascendith
Dg	Than	ley	thy	label	on	the	degre	that	ascendith
Eg	Than	ley	thy	label	over	the	degre	that	ascendith
Hv	Than	ley	thy	label	over	the	degre	that	ascendith
J	Than	ley	thy	label	over	the	degre	that	ascendith
M1	Than	ley	thy	label	over	the	degre	that	assendent
M2	Than	ley	thy	label	over	the	degre	that	ascendith
N	Than	ley	thy	label	over	the	degre	that	ascendith
Pl	Than	ley	thy	label	on	the	degre	that	ascendith
PU	Than	ley	thy	label	upon	the	degre	that	ascendith
R2	Than	ley	thy	label	on	the	degre	that	assendent
Rl1	Than	ley	thy	label	*illegible*	the	degre	that	ascendith
Rl2	∼	∼	∼	∼	∼	∼	∼	∼	∼
Sl1	Than	ley	thy	label	upon	the	degre	that	ascendith
T	Than	ley	thy	label	over	the	degre	that	ascendith
TH1	Than	ley	thy	label	over	the	degre	that	ascendith

Words that are equivalent to each other are aligned, and a '∼' indicates that word is not present in a particular copy. Where a word is illegible in a particular copy, it is entered as 'illegible' and pairwise comparisons ignore that word.

The data above is from the beginning of section II 37, and for which the distance matrix is in Table 8 below. For example, reading across the top row, Ad1 is not different to itself (the value 0 is entered), 0.15698 different to Ad2, 0.1547 different from Bl1, and so on. Using these values, NeighborNet constructs a stemma which represents the relative distances between the various manuscript copies of the text.

Table 9 shows all manuscripts (and TH1)—numbers above the main diagonal indicate the number of distances, and numbers below the main diagonal are the number of words shared by that pair of manuscript copies, in our data sample. The numbers along the main diagonal are the number of words of the main data set of 11375 words that are present in a particular manuscript copy.

The first 200 words from each section, or the whole section if it was shorter than 200 words, were entered, giving a total data set of 11375 words. Over the whole group of manuscripts, we checked the size of the data set by calculating the number of data points shared by any pair of manuscripts (the numbers below the main diagonal in Table 9). The mean number is 5398 (to 0 d.p.), and the median is 7074. Obviously, it is not possible to compare manuscripts that do not share data (represented by 0 below the diagonal in Table 9), and this was a particular problem with the fragmentary copies—especially Cp, M2, R2, Rcp. Once these four manuscripts had been taken out of the main data set, there were only

Table 8
Estimates of distance between pairs of manuscripts, for section II37

	Ad1	Ad2	Bl1	Br	Dd1	Dd2	Dg	Eg	Hv	J
Ad1	0									
Ad2	0.15698	0								
Bl1	0.1547	0.011111	0							
Br	0.13333	0.08	0.07027	0						
Dd1	0.18132	0.072626	0.063492	0.086486	0					
Dd2	0.15	0.12069	0.11475	0.059783	0.13661	0				
Dg	0.026042	0.14943	0.13661	0.1105	0.16304	0.13333	0			
Eg	0.03125	0.14451	0.13187	0.1105	0.15847	0.15	0.015544	0		
Hv	0.057292	0.13873	0.12637	0.099448	0.14754	0.12778	0.041451	0.046632	0	
J	0.057592	0.14535	0.13187	0.10497	0.15301	0.13966	0.041667	0.046875	0.0625	0
M1	0.18132	0.072626	0.063492	0.086486	0.020725	0.12568	0.16304	0.15847	0.14754	0.14754
M2	0.036458	0.13873	0.12568	0.098901	0.14674	0.13889	0.020619	0.025907	0.036269	0.031088
N	0.10526	0.17341	0.1694	0.14917	0.18579	0.17222	0.088542	0.094241	0.11518	0.12042
Pl	0.14525	0.12717	0.12088	0.065574	0.14286	0.0054348	0.12849	0.14525	0.12291	0.14045
Pu	0.1676	0.14451	0.14208	0.092391	0.16393	0.059783	0.17877	0.18994	0.1676	0.17877
R2	0.17416	0.11494	0.11475	0.077348	0.13115	0.088398	0.14444	0.15084	0.14525	0.1573
Rl1	0.13068	0.081871	0.071823	0.016484	0.093407	0.061111	0.10734	0.10734	0.096045	0.10169
Rl2	0.17266	0.050725	0.033784	0.062937	0.07483	0.10638	0.14894	0.15714	0.12857	0.14286
Sl1	0.18935	0.16279	0.15698	0.098266	0.17442	0.063218	0.18935	0.20118	0.17751	0.19048
T	0.073298	0.16279	0.14286	0.1326	0.1694	0.1676	0.057292	0.0625	0.072917	0.083333
TH1	0.17318	0.15029	0.13661	0.092391	0.15847	0.059783	0.17877	0.18436	0.16201	0.17318

	M1	M2	N	Pl	Pu	R2	Rl1	Rl2	Sl1	T	TH1
M1	0										
M2	0.14674	0									
N	0.19126	0.088542	0								
Pl	0.13187	0.13408	0.1676	0							
Pu	0.16393	0.17222	0.20556	0.065574	0						
R2	0.12568	0.15	0.16667	0.094444	0.1326	0					
Rl1	0.093407	0.095506	0.15819	0.067039	0.094444	0.079096	0				
Rl2	0.068027	0.14184	0.19858	0.11429	0.14184	0.099291	0.070423	0			
Sl1	0.17442	0.18343	0.20118	0.069364	0.028736	0.12791	0.10059	0.15385	0		
T	0.1694	0.051813	0.073298	0.16292	0.19553	0.1573	0.14124	0.15714	0.19643	0	
TH1	0.15847	0.16667	0.2	0.065574	0.016216	0.1326	0.094444	0.13475	0.022989	0.18994	0

two pairs for which there were fewer than 500 words in common: A2 and A3, and A3 and Eg, with 470 and 454 words in common respectively. This means that the main data set (all manuscripts, and TH1, omitting Cp, M2, R2 and Rcp) is robust, and likely to lead to reliable results.

For consideration of the fragmentary copies, and in order to understand where the manuscripts changed sources, stemmas were also constructed for each individual section. However, in some cases these stemmas were unreliable since they were based on as few as twenty-two words. Therefore, based on an initial analysis of the single-section stemmas, we grouped the data as shown in Table 10, in order to maximise the possible data sets, while making sure that the maximum number of manuscripts was considered in each subsection. We also tried to ensure that the conflated texts changed source at the same point as the splits in the data set, as far as possible, within the limits due to the small samples of the single-section stemmas.

Sections II35, 36 and 37 were considered individually, since these sections are more difficult to group in the same way as the others.

However, even using these combinations of sections gives data of a lower quality than that of the main data set, and these results were used only in cases where the main data set

Table 9
Raw data for all manuscripts in section II37: number of words shared by two copies under the main diagonal, and number of differences between them above the diagonal

	A1	A2	A3	Ab	Ad1	Ad2	Bl1	Bl2	Br	Bu	Cp	Dd1	Dd2	Dg	Eg	Hg	Hv
A1	**3961**	226	232	169	233	157	136	342	157	246	71	113	253	252	201	116	213
A2	2861	**8468**	101	460	396	498	475	672	498	362	289	483	722	291	330	445	526
A3	1168	470	**1266**	239	122	222	227	241	236	242	0	229	233	244	96	233	224
Ab	3800	6693	1236	**8043**	477	379	325	620	393	366	215	317	609	477	446	309	488
Ad1	2921	8329	567	6962	**9210**	534	514	696	529	373	306	514	762	360	160	448	542
Ad2	3897	6772	1238	7940	7215	**9093**	192	691	452	465	248	374	709	558	506	315	483
Bl1	3935	6800	1242	7970	7251	9062	**9139**	650	397	445	225	297	676	515	476	254	453
Bl2	3570	6421	1213	7588	6704	7724	7744	**7831**	677	462	326	662	802	751	654	648	709
Br	3943	6840	1242	7998	7294	8958	9002	7772	**9091**	474	248	369	626	556	484	320	434
Bu	1979	3308	1140	3047	3663	3760	3748	3101	3763	**5158**	0	475	598	493	352	364	442
Cp	1032	3143	0	3318	3288	3339	3347	3219	3353	0	**3415**	220	334	283	282	213	294
Dd1	3932	6817	1242	7988	7270	8988	9028	7760	9016	3767	3348	**9083**	682	535	479	213	456
Dd2	3728	7062	1181	7708	7517	8627	8662	7465	8713	4008	3286	8679	**9241**	793	719	563	540
Dg	3806	8335	1238	7854	9051	8125	8144	7581	8181	4439	3317	8160	8331	**10068**	268	459	591
Eg	2791	8345	454	6809	9035	7063	7100	6552	7147	3573	3295	7118	7369	8917	**9077**	414	499
Hg	3733	6622	1241	7793	6902	7923	7946	7780	7971	3140	3370	7970	7658	7781	6748	**8018**	423
Hv	3910	7174	1248	7928	7928	8245	8279	7719	8332	3797	3351	8299	8037	8787	7791	7909	**9121**
J	3845	7713	1219	7694	8443	8084	8128	7477	8167	4332	3193	8145	8263	9300	8324	7669	8715
M1	3928	6800	1241	7969	7250	8972	9011	7748	8997	3764	3340	9052	8660	8140	7101	7957	8279
M2	101	2174	0	1005	2860	1171	1186	1005	1197	1197	956	1189	1615	2803	2881	1005	1891
N	3170	6720	1194	6683	7623	6872	6889	6358	6944	4033	2698	6912	7157	8485	7516	6543	7602
Ph	3564	6778	949	7369	6814	7505	7537	7147	7565	2801	3101	7551	7253	7372	6712	7345	7497
Pl	3874	7254	1228	7900	7713	8856	8900	7689	8954	4150	3327	8914	9218	8578	7565	7881	8261
Pu	3686	7562	1221	7683	8301	8631	8667	7673	8720	4959	3306	8682	8868	9137	8180	7869	8705
R1	3834	6572	1235	7698	6815	7823	7854	7466	7877	3097	3126	7876	7560	7689	6660	7693	7818
R2	0	0	0	0	178	809	822	0	816	613	0	820	805	180	179	0	179
Rcp	0	0	0	81	82	81	81	82	83	0	79	81	82	83	82	81	83
Rl1	3924	6809	1234	7950	7260	8901	8947	7719	9010	3746	3329	8965	8677	8123	7111	7920	8292
Rl2	3866	6738	1229	7673	6962	7946	7991	7454	8015	3119	3061	8003	7696	7819	6813	7651	7943
Sl1	3888	7410	1235	7701	8129	8626	8659	7473	8700	4760	3134	8670	8799	8987	7998	7674	8694
Sl2	3951	6825	1238	7938	7074	8072	8105	7707	8149	3128	3304	8119	7828	7968	6924	7917	8080
T	3428	7841	1129	7232	8545	8098	8136	6959	8163	5094	3366	8155	8303	9285	8467	7160	8189
TH1	3888	7778	1219	7899	8507	8845	8890	7672	8934	4958	3300	8905	9073	9353	8386	7863	8908

(*continued on next page*)

Table 9 (*continued*)

	J	M1	M2	N	Ph	Pl	Pu	R1	R2	Rcp	Rl1	Rl2	Sl1	Sl2	T	TH1	W
W	2786	1958	1159	2804	2073	2910	2946	2666	2941	2116	0	2943	2708	2918	1933	2708	2914
A1	276	130	13	264	164	269	222	134	0	0	143	196	139	109	342	211	158
A2	699	487	130	623	457	750	646	494	0	0	473	486	581	490	647	644	207
A3	247	237	0	223	196	238	237	233	0	0	232	242	223	242	230	237	240
Ab	602	337	83	422	435	632	577	333	0	2	351	476	388	441	652	561	185
Ad1	734	525	111	676	438	789	656	498	31	7	507	483	628	490	649	655	210
Ad2	652	404	100	610	452	735	725	364	110	3	417	498	501	443	811	710	188
Bl1	638	330	88	589	407	696	700	296	104	4	342	451	454	383	776	680	179
Bl2	810	673	83	660	682	837	822	665	0	3	660	724	668	719	866	815	291
Br	642	395	85	613	438	649	673	348	75	4	239	488	454	419	798	657	200
Bu	577	491	119	546	351	618	711	374	105	0	458	395	667	387	371	703	289
Cp	326	233	93	271	247	334	359	203	0	8	243	258	239	268	207	348	0
Dd1	631	126	96	587	407	707	705	264	103	3	320	460	437	359	794	684	175
Dd2	890	699	211	822	581	212	713	623	99	8	599	650	664	640	1019	717	231
Dg	779	550	120	733	455	837	744	508	26	4	517	541	657	530	742	729	259
Eg	679	484	70	638	399	750	621	464	27	7	463	445	582	447	569	606	192
Hg	558	241	67	526	379	585	558	232	0	4	280	415	314	341	611	528	156
Hv	684	482	91	677	435	543	644	476	26	3	401	502	495	492	761	641	190
J	**9652**	653	203	826	512	913	833	599	39	8	625	581	782	610	974	845	207
M1	8129	**9067**	97	608	424	719	716	289	108	6	346	475	471	385	806	705	184
M2	2564	1189	**2897**	227	45	215	295	68	27	6	80	69	213	70	128	289	0
N	8144	6890	3221	**8688**	545	851	872	558	30	8	606	616	712	609	782	860	253
Ph	7304	7533	796	5951	**7603**	596	490	424	0	0	391	90	388	393	654	495	157
Pl	8522	8896	1658	7332	7469	**9526**	718	645	103	7	604	665	662	657	1060	730	261
Pu	9003	8671	2320	7791	7273	9149	**10086**	569	120	7	631	546	562	570	1077	111	40
R1	7589	7866	876	6426	7296	7782	7575	**7924**	0	8	324	464	357	394	642	558	195
R2	269	820	180	180	0	815	807	0	**832**	0	67	14	115	0	134	121	0
Rcp	83	81	82	160	0	82	82	67	0	**83**	3	0	5	3	9	7	0
Rl1	8121	8947	1184	6905	7514	8908	8671	7834	816	83	**9041**	437	415	369	797	616	181
Rl2	7750	7984	937	6398	7546	7915	7719	7604	141	0	7964	**8056**	443	439	732	546	187
Sl1	8961	8651	2075	7590	7281	9080	9560	7612	781	82	8653	7714	**9903**	261	944	540	162
Sl2	7863	8099	995	6721	7535	8051	7835	7822	0	79	8112	7840	7861	**8184**	718	557	184
T	8794	8143	2881	7991	6780	8549	9160	7092	786	83	8109	7180	8972	7297	**10141**	1080	302
TH1	9221	8886	2321	8003	7494	9354	10075	7792	809	82	8884	7940	9775	8050	9375	**10304**	27
W	2876	2937	0	2357	2573	2878	2855	2913	0	0	2920	2917	2878	2932	2566	3078	**3079**

Table 10
Data sections (and numbers of words in each data set) used for analysis of part I and part II

Part I	Part II
225 **Prologue (a)**	602 **1 & 2**
203 **Prologue (b)**	433 **3 & 4**
266 **1, 2, 3 & 4**	409 **5 & 6**
465 **5, 6, 7 & 8**	284 **7, 8 & 9**
274 **9, 10 & 11**	311 **10 & 11**
218 **12, 13 & 14**	428 **12, 13 & 14**
445 **15, 16 & 17**	487 **15, 16 & 17**
277 **18, 19 & 20**	267 **18 & 19**
332 **21, 22 & 23**	347 **20, 21 & 22**
	552 **23, 24 & 25**
	584 **26, 27 & 28**
	420 **29 & 30**
	619 **31, 32, 33 & 34**
	221 **35**
	245 **36**
	213 **37**
	685 **38, 39 & 40**
	614 **41, 42 & 43**
	715 **41ab, 42ab & 43a**
	437 **44 & 45**

did not give sufficient detail (for example, on the changing sources of conflated copies) or did not include a particular copy (Cp, M2, R2, Rcp).

References

Barbrook, A. C., Howe, C. J., Blake, N., & Robinson, P. (1998). The phylogeny of *The Canterbury tales*. *Nature*, *394*, 839.

Bryant, D., & Moulton, V. (2002). NeighborNet: An agglomerative method for the construction of planar phylogenetic networks. *Lecture Notes in Computer Science*, *2452*, 375–391.

Buneman, P. (1971). Filiation of manuscripts. In F. R. Hodson, D. W. Kendall, & P. Tautu (Eds.), *Mathematics in the archaeological and historical sciences* (pp. 387–395). Edinburgh: Edinburgh University Press.

Cameron, H. D. (1987). The upside-down cladogram: Problems in manuscript affiliation. In H. M. Hoenigswald, & L. F. Wiener (Eds.), *Biological metaphor and cladistic classification: An interdisciplinary perspective* (pp. 227–242). Philadelphia: University of Philadelphia Press.

Dearing, V. A. (1974). *Principles and practice of textual analysis*. Berkeley: University of California Press.

Eagleton, C. (2003). A previously unnoticed fragment of Chaucer's *Treatise on the astrolabe*. *Journal of the Early Book Society*, *6*, 161–173.

Eisner, S. (1985). Chaucer as a technical writer. *Chaucer Review*, *19*, 179–201.

Eisner, S. (2002). *A treatise on the astrolabe. Variorum Editions of the Works of Geoffrey Chaucer, Vol. VI. The Prose Treatises*. Norman: University of Oklahoma Press.

Felsenstein, J. (2004). *Inferring phylogenies*. Sunderland, MA: Sinauer.

Flight, C. (1990). How many stemmata? *Manuscripta*, *34*, 122–128.

Flight, C. (1994). A complete theoretical framework for stemmatic analysis. *Manuscripta*, *38*, 95–115.

Gascuel, O. (1994). A note on Sattath and Tversky's, Saitou and Nei's, and Studier and Keppler's algorithms for inferring phylogenies from evolutionary distances. *Molecular Biology and Evolution*, *11*, 961–963.

Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, *426*, 435–439.

Hanna, R. (2000). The application of thought to textual criticism in all modes—with apologies to A.E. Housman. *Studies in Bibliography*, *53*, 163–172.

Hillis, D. M., Bull, J. J., White, M. E., Badgett, M. R., & Molineux, I. J. (1992). Experimental phylogenetics: Generation of a known phylogeny. *Science*, *255*, 589–592.

Huson, D. H. (1998). SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics*, *14*, 68–73.

Laird, E. (1999). Geoffrey Chaucer and other contributors to the *Treatise on the astrolabe*. In T. A. Prendergast, & B. Kline (Eds.), *Rewriting Chaucer: Culture, authority, and the idea of the authentic text* (pp. 145–165). Columbus: Ohio State University Press.

Laird, E. (2000). A previously unnoticed manuscript of Chaucer's *Treatise on the astrolabe*. *Chaucer Review*, *34*, 410–415.

Lee, A. R. (1989). Numerical taxonomy revisited: John Griffith, cladistic analysis and St. Augustine's *Quaestiones in Heptateuchem*. *Studia Patristica*, *20*, 24–32.

Nei, M. (1991). Relative efficiencies of different tree-making methods for molecular data. In M. M. Miyamoto, & J. Cracraft (Eds.), *Phylogenetic analysis of DNA sequences* (pp. 90–128). New York: Oxford University Press.

Pintelon, P. (Ed.). (1940). *Chaucer's Treatise on the astrolabe*. Antwerp: De Sikkel.

Platnick, N. I., & Cameron, H. D. (1977). Cladistic methods in textual, linguistic, and phylogenetic analysis. *Systematic Zoology*, *26*, 380–385.

Poole, E. (1974). The computer in determining stemmatic relationships. *Computers and the Humanities*, *8*, 207–216.

Price, D.J. de Solla. (N.d.). On astrolabe MSS. Unpublished typescript, Harvard University, Houghton Library, MS English 920.1.

Rand Schmidt, K. A. (1993). *The authorship of the Equatorie of the planetis*. Woodbridge: Brewer.

Rand, K. A. (2005). Review of Sigmund Eisner (ed.), *A variorum edition of the works of Geoffrey Chaucer. Vol. 6. The prose treatises. Part 1. A treatise on the astrolabe. English Studies*, *86*:5, 458–461.

Reidy, J. (1988). Textual notes to A *treatise on the astrolabe*. In L. D. Benson (Ed.), *Riverside Chaucer* (pp. 1193–1197). Oxford: Oxford Paperbacks.

Robinson, P. M. W., & O'Hara, R. J. (1996). Cladistic analysis of an Old Norse manuscript tradition. In S. Hockey, & N. Ide (Eds.), *Research in humanities computing 4* (pp. 115–137). Oxford: Oxford University Press.

Robinson, P. (1997). A stemmatic analysis of the fifteenth-century witnesses to *The Wife of Bath's prologue*. In N. Blake, & P. Robinson (Eds.). *The Canterbury tales project: Occasional papers, Vol. II* (pp. 69–132). London: Office for Humanities Communication.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*, 406–425.

Salemans, B. J. P. (1996). Cladistics or the resurrection of the method of Lachmann: On building the stemma of Yvain. In P. van Reenen, P. van Mulken, & M. van Mulken (Eds.), *Studies in stemmatology* (pp. 3–70). Amsterdam: John Benjamins.

Seymour, M. C. (1995). *A catalogue of Chaucer manuscripts. Vol. I. Works before the Canterbury tales*. Aldershot: Scholar Press.

Skeat, W. W. (Ed.). (1872). *A treatise on the astrolabe addressed to his son Lowys by Geoffrey Chaucer, A.D. 1391*. London: Early English Text Society.

Spencer, M., Bordalejo, B., Robinson, P., & Howe, C. J. (2003). How reliable is a stemma? An analysis of Chaucer's *Miller's tale*. *Literary and Linguistic Computing*, *18*, 407–422.

Spencer, M., Davidson, E. A., Barbrook, A. C., & Howe, C. J. (2004). Phylogenetics of artificial manuscripts. *Journal of Theoretical Biology*, *227*, 503–511.

Spencer, M. et al. (2004). The effects of weighting kinds of variants. In P. van Reenen, & A. den Hollander (Eds.), *Studies in stemmatology II* (pp. 225–237). Amsterdam: Benjamins.

Susko, E., Spencer, M., & Roger, A. J. (2005). Biases in phylogenetic estimation can be caused by random sequence segments. *Journal of Molecular Evolution*, *61*, 351–359.

Tehrani, J., & Collard, M. (2002). Investigating cultural evolution through biological phylogenetic analyses of Turkmen textiles. *Journal of Anthropological Archaeology*, *21*, 443–463.