

# Dante's *Monarchia* as a test case for the use of phylogenetic methods in stemmatic analysis

Heather F. Windram

Department of Biochemistry, University of Cambridge,  
Cambridge, UK

Prue Shaw

Department of Italian, University College London, London, UK

Peter Robinson

Institute for Textual Scholarship and Electronic Editing, Graduate  
Institute of Theology and Religion, University of Birmingham,  
Birmingham, UK

Christopher J. Howe

Department of Biochemistry, University of Cambridge,  
Cambridge, UK

## Abstract

Dante's *Monarchia*, a fourteenth century treatise on political theory which survives in 20 manuscripts and the *editio princeps*, has been studied extensively by scholars using traditional analytical methods to establish textual transmission. It was selected as a suitable tradition for a blind study to test the application of computer-based phylogenetic methods to the stemmatic analysis of manuscript relationships. Our results show that these methods—maximum parsimony, NeighborNet and the Supernet algorithm—are capable of producing stemmata in very close agreement with those produced by traditional stemmatic analysis, including the identification of texts that change exemplar in the course of copying. The phylogenetic methods can correctly indicate the affiliations both before and after the point of exemplar change. The maximum chi-squared method (developed to detect recombination in DNA sequences) is able to indicate the region of exemplar change, allowing the precise location to be ascertained by textual analysis.

### Correspondence:

H. F. Windram, Department  
of Biochemistry, University  
of Cambridge, Tennis  
Court Road, Cambridge  
CB2 1QW, UK.

### E-mail:

h.f.windram@btinternet.com

## 1 Introduction

### 1.1 Traditional analysis and phylogenetic analysis

The principle of reconstructing the transmission history of a manuscript tradition of a text by

analysing common errors and other textual alterations to establish relationships between extant manuscripts was formulated in the nineteenth century and has been standard practice for scholars working on medieval texts for well over 100 years (Timpanaro, 1985; Salemans, 2000). This approach

has produced critical editions of outstanding distinction in many languages, including those of Dante's *Divine Comedy* by Petrocchi (1966–67) and Chaucer's *Canterbury Tales* by Manly and Rickert (1940). These traditional stemmatological methods were tested in a study by Hiram Peri (Pflaum) (1956), where they were successfully used to analyse an 'artificial' tradition (where present-day scribes copied a text, and the copying history was known).

There are many similarities between the inheritance and accumulation of scribal alterations through successive generations of copying, and the inheritance and accumulation of genetic mutations in molecules of DNA through successive generations of living organisms. In both cases, a string of information (words in the case of a manuscript, or nucleotides—the repeating chemical units that make up the molecule—in the case of DNA) is copied reasonably accurately, resulting in two copies of the same string of information. However, in both cases, changes can occur, resulting in alterations to the text or mutations in the DNA. Many common types of textual change have parallels in DNA mutations. Substitution of one word for another is analogous to the substitution of a nucleotide. Similarly, insertion or omission of words is analogous to the insertion or deletion of nucleotides.

A scribe using more than one exemplar while copying a text creates a hybrid 'contaminated' text, which may either contain sections from each exemplar used, or incorporate individual readings from a second copy into the original base text. This has parallels with genetic recombination, where related DNA molecules produce a hybrid, or lateral gene transfer, where DNA is exchanged between distantly related organisms (Barbrook *et al.*, 1998; Howe *et al.*, 2001; Windram *et al.*, 2005).

The similarities between textual transmission and biological inheritance have been utilized in many studies, where phylogenetic techniques developed in evolutionary biology have been applied to textual traditions. (e.g. Platnick and Cameron, 1977; Lee, 1989; Robinson and O'Hara, 1996; Barbrook *et al.*, 1998; Salemans, 2000; Stolz, 2003; Lantin *et al.*, 2004). The availability of methods and computer

programs from evolutionary biology that can handle large amounts of data (see Page and Holmes, 1998 for an overview of many commonly used methods) makes their application potentially very powerful.

However, this approach to textual studies has attracted opposition from some scholars who argue that it is not possible to make any attempt to understand the true relationships between manuscripts without a deep understanding of the text and context of the work (e.g. Hanna, 2000; Cartlidge, 2001). A realistic position is probably that these phylogenetic methods offer a valid approach to manuscript analysis, but that their real value comes in their power to assist the textual scholar rather than to attempt to replace the scholar's expertise. Once the initial coding of texts into a form that can be handled by the computer programs has been completed, phylogenetic methods offer an extremely rapid insight into the transmission history of the text. They allow the researcher to study in a few minutes the effects of, for example, including only certain sections of the text in an analysis, or omitting certain manuscripts.

Our early studies with a variety of traditions, both artificial (Spencer *et al.*, 2004; Baret *et al.*, 2006) and genuine (Mooney *et al.*, 2001; Windram *et al.*, 2005; Eagleton and Spencer, 2006) showed that phylogenetic methods supported the results of traditional analysis. However, these test studies generally used texts that were rather short or had a simple copying history. In order to test the applicability of phylogenetic methods to longer texts, with a more complex copying history, we used the methods to analyse a well-studied longer text, Dante's *Monarchia*, which, excluding notes and apparatus, runs to some fifty pages in the printed version.

The phylogenetic analysis was done 'blind' by two of us (H.F.W. and C.J.H.) who did not know the identity of the text, but just worked from the variant list and coded matrix prepared by the textual specialists in the team. This ensured that the initial interpretation of the results and the identification of features of interest within the tradition could not have been influenced by prior knowledge. We then combined all members of the team and compared

the results of the computer analyses with those of more traditional textual scholarship using recent publications (Shaw, 1995, 2006).

## 1.2 The *Monarchia* tradition

Dante's *Monarchia* is a treatise on political theory written in medieval Latin in the early fourteenth century. It comprises three books which address specific questions central to Dante's vision of the role of the Monarchy, or universal empire, in the way human society should ideally be ordered. Book I considers whether monarchy in this sense is necessary to the well-being of the world. Humanity as a whole has a purpose and therefore needs an overall leader, a single sovereign authority (Monarch) set over all lesser rulers. Book II tackles the question of whether the Roman people took on this role by right due to their nobility and inherent superiority as a race. Dante uses poetic testimony from classical literature to support his case, giving an episodic view of Roman history rather than a chronological historical account. In Book III, he considers whether the monarch derives his authority directly from God, or indirectly from the Pope. He uses arguments based on the scriptures and historical evidence to argue against the primacy of papal authority, concluding that there is a direct line of authority from God to the Monarch (Shaw, 1995).

This work survives in twenty-one witnesses dating from the fourteenth century to the eighteenth century, one of these being the *editio princeps* of 1559, which is not copied from any extant manuscript, and which is accorded manuscript status. Over the last 150 years there has been a continuing history of scholarly analysis of the *Monarchia* tradition reflected in the publication of four influential editions, those of Witte (1874), Bertalot (1918), Rostagno (1st edn 1921; 2nd edn, 1960) and Ricci (1965) (see Appendix I for full details). The last of these is the Edizione Nazionale published to coincide with the 700th anniversary of Dante's birth in 1265. More recently, Shaw has published two editions of *Monarchia*, one in standard book format (1995) and the other as a DVD-ROM (2006). Shaw's critical text will replace Ricci's as the Edizione Nazionale of the Società Dantesca Italiana, in a volume which is forthcoming.

## 2 Methods

### 2.1 Encoding of text for phylogenetic analyses

Data from the transcripts of each witness were converted into a matrix, or NEXUS file (Maddison *et al.*, 1997), by encoding the texts as symbols. If several witnesses agreed on a reading at a particular site, they were given the same symbol at that location, while each alternative reading was coded with a different symbol. When text is missing at a particular location, there are two ways of encoding this. It can be treated as a number of independent omissions with each character (e.g. word) encoded separately as an omission, or it can be treated as a single insertion or omission. The occurrence of significant omissions is often important for establishing relationships between texts ('il principio degli errori', Gianfranco Contini, 1986). However, if each word in a region of omitted text is coded independently as a separate event, then undue significance may be given to the omission which, in reality, usually represents a single event e.g. an eye-skip between two occurrences of the same word. NEXUS files were prepared in accordance with both strategies.

In this study, the analysis was carried out initially using a NEXUS file encoded under the first approach, and was then repeated using a NEXUS file in which insertions or omissions were encoded as single events. In this instance, there was no significant difference between the conclusions from the two analyses; the results presented here are from the second analysis. The NEXUS file was analysed using a range of methods.

### 2.2 Phylogenetic methods involving all of the manuscripts

#### 2.2.1 *Maximum parsimony*

The maximum parsimony method considers all possible bifurcating trees and identifies the one that requires the smallest number of changes (i.e. is most parsimonious). The various species, or manuscripts, are in this way grouped according to their shared derived characters. Not all sites are informative for maximum parsimony analysis. A parsimony informative site is one where there is

more than one variant recorded, and where at least two variants are each recorded at least twice. All other sites are considered parsimony uninformative and are ignored by the analysis. Maximum parsimony analysis was carried out using PAUP\* (Swofford, 2001). Trees generated using different sections of the text were compared to identify witnesses whose position varied according to the section used, and whose exemplar might therefore have changed during the copying process.

Bootstrapping was used to give a measure of the statistical support for a tree structure generated by maximum parsimony (Felsenstein, 1985). This technique uses a large number of replicate data sets, generated from the original data set by sampling a set of sites at random (and with replacement) from the original data set. A phylogenetic tree is generated individually for each data set and the results are then represented as a consensus tree which shows only manuscript groupings supported by a chosen percentage of the individual trees. In this study, 1,000 replicate data sets were analysed, and the consensus tree showed only those groupings supported by at least 50% of the individual trees. The bootstrap support values are shown on the maximum parsimony trees and indicate the percentage of the individual trees supporting that particular grouping.

It is important to note that bootstrapping indicates how certain we are about the estimate of phylogeny from a given method, but not whether that estimate is actually correct. A bootstrap support value of 100% for a particular grouping tells us that adding more data is unlikely to change our opinion about the grouping. However, if the method is in some way inappropriate, it is possible to get 100% bootstrap support for incorrect groupings.

### 2.2.2 *NeighborNet*

The NeighborNet method (Bryant and Moulton, 2002, 2004) represents the relationships between species, or manuscripts (Spencer *et al.*, 2004), as a network rather than as a simple branching tree, thus enabling multiple affiliations to be represented. The distance between pairs of manuscripts is calculated as the total number of differences between

them. The closest pairs of manuscripts, and pairs of pairs etc., are then grouped and when a member of one pair has been chosen as a member of a second pair, then both pairs are replaced with a pair of linked hypothetical ancestors to give an overall network showing the relationships between the various manuscripts. The analyses were performed using the SplitsTree4 package v 4.3 (Huson and Bryant, 2006).

### 2.2.3 *Supernetwork analysis*

The Supernetwork algorithm (Huson *et al.*, 2004) forms part of the SplitsTree4 package (Huson, 1998; Huson and Bryant, 2006). For this analysis the input data are the trees produced by other phylogenetic methods such as maximum parsimony. Information present in the topologies of individual trees (e.g. the tree for one gene or, in the case of *Monarchia*, for an individual book) is used to generate an overall tree or network that gives the best representation of the relationships between individual species or, in this case, manuscripts. The resultant network enables conflicts between the input trees to be visualized. One advantage of this method is that the input trees do not all need to contain exactly the same set of species or manuscripts, although the accuracy of the results may be weakened if many of the species or manuscripts are not common to all of the input trees.

The trees produced by all of these phylogenetic analyses show the relationships between the manuscripts according to the assumptions of each method. All of the trees are unrooted and give no indication of the origin of the tradition, which can best be established by textual analysis. It should be noted that, as with the analysis of biological data, the results are not affected by instances of convergent variation (also known as polygenetic variation), where scribes may make similar changes independently. This type of change tends to be distributed randomly throughout the manuscripts and gives no statistically significant bias to the results. Indeed, one of the main benefits of a numerical analysis is that it avoids giving undue significance to this type of unsystematic error.

## 2.3 Methods involving pairs of manuscripts

### 2.3.1 Maximum chi-squared analysis

This method is based on the chi-squared method developed by Maynard Smith for the detection of recombination in pairs of DNA sequences (Maynard Smith, 1992). For each selected pair of manuscripts, the program compares the data at each character number in the two rows of the NEXUS matrix that correspond to the chosen manuscripts (Windram *et al.*, 2005). It creates a linear sequence of differences indicating whether the two manuscripts being compared are the same or different at each location in the NEXUS file. If the two manuscripts are similarly related throughout, then on average, any differences between them should be evenly spaced along the length of the sequence. However, if, for example, the two sequences are very closely related, but only in the second half, then the differences would tend to occur predominantly in the first half. The method moves an imaginary breakpoint along the sequence of differences, and at each point calculates a chi-squared value which represents the difference between the observed number of differences on either side of the breakpoint and the expected number of differences in the absence of any recombination (i.e. if the differences are evenly distributed, see Fig. 1). The result is a graph showing the chi-squared value



**Fig. 1** The distribution of the differences (marked as asterisks) between sequences A and B are compared by moving an imaginary break point (solid arrows) along the sequence of differences. In (i), the texts are similarly related throughout and the observed number of differences at each point would equal the expected number of differences (i.e. an even distribution of differences throughout the text). In (ii), the texts are particularly closely related in the second half and the maximum chi-squared value would occur at the location (dotted arrow), where there is the greatest discrepancy between the observed distribution of differences and that expected for sequences that are similarly related throughout

(vertical axis) for each position of the window (horizontal axis). The maximum chi-squared value occurs at the location where there is the greatest discrepancy between the observed and the expected number of differences, and this indicates the most likely breakpoint or point of exemplar change (Windram *et al.*, 2005). A probability (*P*) value can be assigned to the maximum chi-squared measure, which indicates the chance of getting this chi-squared value, or a larger one, in the absence of recombination. Thus a lower probability value indicates a *more* significant shift.

### 2.3.2 Distribution of differences

This analysis provides an indication of how different two manuscripts are within a particular window. A window of chosen size is moved along the same sequence of differences as was used for the maximum chi-squared analysis. A window of 1,501 characters proved optimal for the *Monarchia* data and was used for all the analyses reported here. The number of locations where a difference has been recorded is calculated as a percentage of the total number of locations in the window, and this percentage of differences may then be plotted against the character number at the mid-point of the window. The resultant graph shows clearly how the percentage of differences varies along the length of the two sequences and whether the two manuscripts become more or less similar after any possible recombination point. The maximum chi-squared program and the percentage of differences program were both written and implemented in Matlab (The Mathworks, Inc., Natick, MA, USA).

## 3 Results

### 3.1 The complete tradition

The data for phylogenetic analysis consisted of a 21 witness  $\times$  16,548 character matrix (the NEXUS file). This divides into the three books as follows:

Book I NEXUS chrs 1–4,659 = 4,659chr.

Book II NEXUS chrs 4,660–10,209 = 5,550chr.

Book III NEXUS chrs 10,210–16,548 = 6,339chr.

Some of the witnesses are incomplete. There are many minor and several sizeable regions of missing



**Table 1** The location and explanation for sections of data of thirty or more characters missing from the *Monarchia* tradition

Witness	Region(s) missing (Chr. No. in NEXUS file)	Reason
Q	3,680–16,548	Only the first gathering of the witness survives
P	9,031–9,777	A folio has gone missing from the exemplar from which P was copied
R	13,316–13,663 14,028–16,548	The witness is missing its last few pages and an isolated page near the end
A	5,728–5,791 5,868–5,912	The missing text consists of two lengthy quotations omitted by the scribe
T	5,728–5,791 5,868–5,912	The missing text consists of two lengthy quotations omitted by the scribe
C	14,908–15,814	A folio is missing near the end of the text
Z	8,951–8,990	Omission of five lines due to eye-skip by the scribe

data, including several sections of thirty or more characters in length (Table 1).

This NEXUS file was used for maximum parsimony, NeighborNet, and Supernet network analyses using both the complete tradition and the individual books.

### 3.1.1 Maximum parsimony analysis

The maximum parsimony analysis of the complete tradition (Fig. 2) was then compared with the relationships between witnesses established by traditional stemmatic analysis (Table 2). Ricci (1965) suggested that the extant witnesses could be divided into a two-branched stemma, with the majority falling into the  $\beta$ -group, while three witnesses, including the *editio princeps* (K), formed a separate  $\alpha$ -family. Shaw (2006) has shown that the tradition may be better represented by a three-branched stemma, with the *editio princeps* deriving from the archetype independently of the  $\alpha$ -family manuscripts (A and T). It should be borne in mind that

Shaw's analysis incorporates two manuscripts, U and Ph, which were unknown to Ricci (Shaw, 1996).

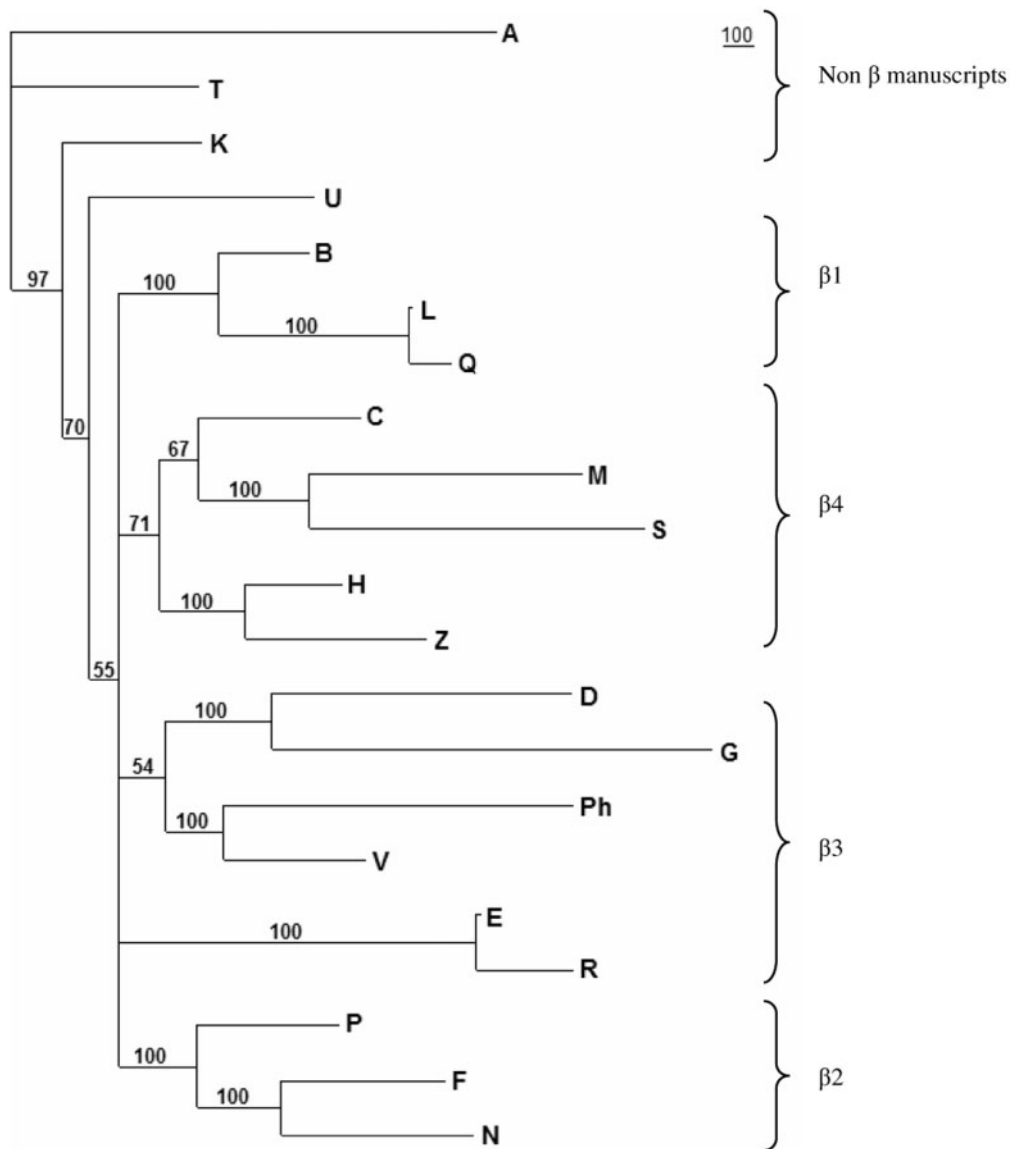
The evolutionary tree produced by maximum parsimony strongly resembles the stemma constructed by Shaw (Appendix II). Traditional analysis has shown that the  $\beta$ -manuscripts divide into four subgroups, termed  $\beta_1$ – $\beta_4$ , with A shifting from the  $\alpha$ -family to the  $\beta_3$ -group for the second half of the text. The two halves of manuscript A are referred to as A<sup>1</sup> and A<sup>2</sup>. The main groupings of the traditional analysis were retained in well-supported groups in the maximum parsimony analysis. As a general guide, groupings in biological analyses are considered to be well supported if they have bootstrap support values of more than 70% (Hillis and Bull, 1993). Manuscript A grouped clearly with T (97% bootstrap support) when the complete data set was analysed. The  $\beta_1$ - and the  $\beta_2$ -manuscripts had 100% support for all their groups. The paired manuscripts in the  $\beta_3$ - and  $\beta_4$ - groups all had 100% support, but the grouping of these pairs was less reliable, especially the grouping of G/D with Ph/V which was very uncertain, having only 54% support. The E/R pair was not linked to the other  $\beta_3$ -texts.

### 3.1.2 NeighborNet analysis

The NeighborNet results (Fig. 3) were largely in agreement with the maximum parsimony results and the traditional analysis.

The  $\beta_1$ - and  $\beta_2$ -groups were clearly defined. The  $\beta_3$ -texts Ph, V, E, and R were grouped together, but the G/D pair grouped separately from the other  $\beta_3$ -manuscripts and showed linkage with S/M. Manuscripts G, D, S, and M all had long-branch lengths, indicating that each contains many unique variants. This linkage of S/M and D/G also disrupted the  $\beta_4$ -group, with just C, H, and Z grouping together. Manuscript A was linked with T, which shows some slight affiliation (indicated by a reticulation) to K.

The E/R pair was typical of a grouping where one text (in this case R) has been copied directly from the other (E). In such cases, the exemplar typically has a very short individual branch away from its copy, and forms an almost internal node within the tree (Spencer *et al.*, 2004; Baret *et al.*, 2006). Thus E and R formed a very narrow network that



**Fig. 2** A maximum parsimony tree of the complete *Monarchia* data set. The phylogenetic trees in this article are all unrooted, and so give no indication of the origin of the tradition. The tree is a bootstrap consensus tree using 1,000 replicate data sets, and shows groupings that are supported by at least 50% of the individual trees. Each branch is labelled with its bootstrap support percentage. The horizontal branch lengths are proportional to the number of substitutions inferred. The figure also shows the traditional grouping of the witnesses into groups  $\beta 1$ – $\beta 4$  and non- $\beta$ -manuscripts

appeared more as a forked branch. The copy, R, not only contains almost all of the error variants present in the exemplar, E, but also contains its own collection of unique errors. In contrast, the manuscript

pairs F/N and H/Z were typical of manuscripts copied independently from a common exemplar. They each formed a tight pair, as they have many variants in common, but the individual manuscripts

each contain independent errors found uniquely in that witness. Some scribes introduce more errors/variants than others (e.g. manuscript Z contains more unique variants than manuscript H), but the branch length of H indicates that it was not a direct exemplar for Z.

Sometimes a branch length can be apparently misleading, for example if large sections of one

manuscript are missing. Q is known from traditional analysis to be copied directly from L, but although the two manuscripts group together, L has a longer branch length than one might expect for the exemplar. However, manuscript Q consists only of part of the first book and is missing from chr 3,681–16,548, including all of Book II and Book III. L can then be compared only with the remaining manuscripts, and so has many more unique variants, and a longer branch length. It is likely that most of these variants would also have been found in the remainder of Q. A comparison of the results of the traditional analysis and of the phylogenetic analysis for each manuscript group is summarized in Table 3.

**Table 2** The manuscript groups of the *Monarchia* tradition according to traditional analysis

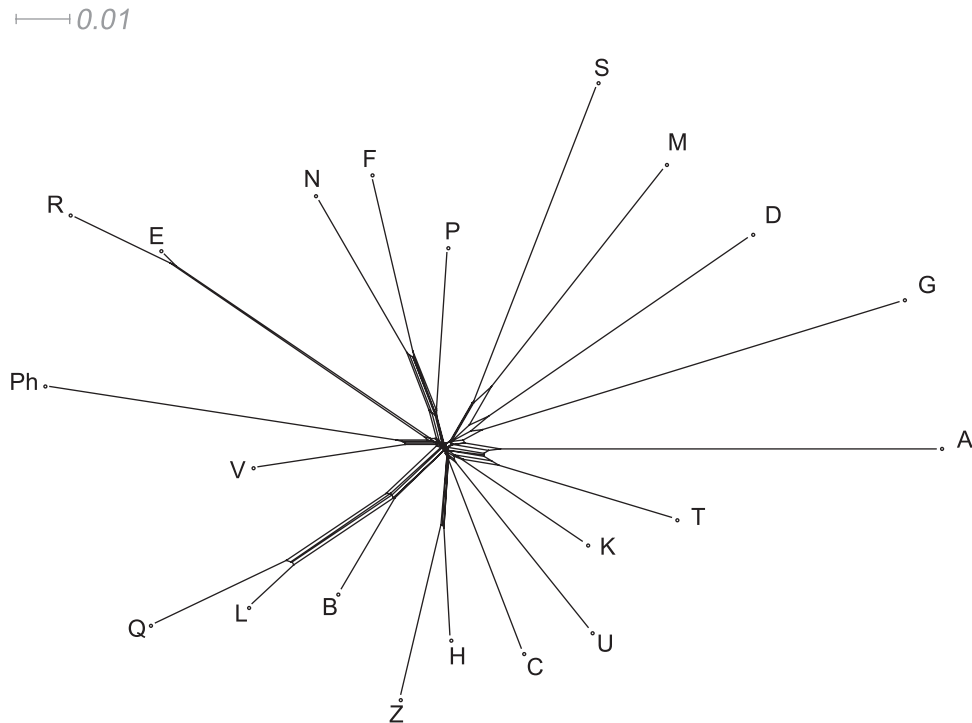
Manuscript group	Manuscripts in group
<i>Editio princeps</i>	K
(non- $\beta$ -witness) <sup>a</sup>	
Non- $\beta$ -manuscripts <sup>a</sup>	A <sup>1</sup> , T
$\beta$ 1	B, L, Q
$\beta$ 2	P, F, N
$\beta$ 3	Ph, V, E, R, G, A <sup>2</sup> + D (contaminated)
$\beta$ 4	C, H, Z, M, S + U (contaminated)

<sup>a</sup>Manuscripts A<sup>1</sup>, T form the  $\alpha$ -group (Shaw, 2006). Earlier analysis (Ricci, 1965) also included K in the  $\alpha$ -group.

### 3.2 The individual books

#### 3.2.1 Phylogenetic analyses

The data file was split and each book examined separately to identify manuscripts whose affiliation



**Fig. 3** The NeighborNet graph of the complete *Monarchia* data set. The edge lengths are proportional to the number of differences between manuscripts (expressed in the scale bar as the mean character difference)



**Table 3** A comparison of the results of traditional and phylogenetic analysis of the various manuscript groups of the *Monarchia* tradition

Traditional analysis	Phylogenetic analysis
<p><b>The non-<math>\beta</math>-manuscripts—K, T, and A<sup>1</sup></b>  T and A<sup>1</sup> are closely related.  K is separate from the T, A<sup>1</sup> group (Shaw, 1996)</p> <p><b>The <math>\beta</math>1 manuscripts—B, L, and Q</b>  Manuscripts B, L, and Q form a subgroup, with L and Q being closer to each other than to B. Manuscript Q is copied from L. Many of the individual readings in Q correspond to characteristics of L (e.g. abbreviations or corrections) not understood by the scribe of Q.</p> <p><b>The <math>\beta</math>2 manuscripts—P, F, and N</b>  Manuscripts P, F, and N form a related group.  F and N share a common ancestor.  N is older than F, but F cannot be a direct copy of N, because there are some omissions in N that do not occur in F.</p> <p><b>The <math>\beta</math>3 manuscripts—Ph, V, E, R, G, and A<sup>2</sup></b>  Manuscripts Ph and V are closely related, sharing a common ancestor. They share errors and omissions with the other <math>\beta</math>3 manuscripts.  R is believed to be copied from E.  Manuscript A<sup>2</sup> is a member of the <math>\beta</math>3 group.</p> <p><b>The <math>\beta</math>4 manuscripts—C, H, Z, M, and S</b>  Manuscripts S and M are closely related.  C generally shows similarities with S and M.  H and Z share a common ancestry, but their pattern of variants shows that they are not copied one from the other.</p> <p><b>The Contaminated manuscripts—D and U</b>  Manuscripts D and U do not fit neatly into any of the groupings.  D contains features of <math>\beta</math>3- and <math>\beta</math>4-manuscripts. It has links to both M and G.  U is a <math>\beta</math>-manuscript (mainly <math>\beta</math>4), but it shares some significant good readings with K and T, where the <math>\beta</math>-manuscripts are corrupt. The scribe either copied from a contaminated exemplar, or contaminated as he copied.</p>	<p>T and A form a networked group with 97% support. K groups with T, A with 70% support.</p> <p>Manuscripts L, Q and B, L, Q group with 100% support. L and Q are closer to each other than to B. Q has a longer branch length than L reflecting its greater number of unique readings.</p> <p>Manuscripts F, N and P, F, N group with 100% support. F and N are closer to each other than to P. The branch lengths for F and N are similar, supporting the view that neither is directly copied from the other.</p> <p>Manuscripts Ph and V group together (with 100% support). In the NeighborNet analysis, G shows linkage with D. Manuscripts E and R group with 100% support. R branches off from the E branch and has a longer branch length, supporting the idea of it being copied directly from E. It contains almost all of E's errors and others of its own. For A<sup>2</sup> affiliation see later results.</p> <p>Manuscripts S and M group together (100% support). NeighborNet analysis grouped C with H, Z. Maximum parsimony analysis grouped C first with S, M, and then with the H, Z pair.  H and Z group together (100% support). Their branch lengths indicate that neither is copied directly from the other.  Z contains more unique variants.</p> <p>Manuscripts G, D and manuscripts S, M occur as two pairs (each with 100% support). The NNet analysis shows some linkage between the two pairs, especially between D and M. Manuscript U does not appear to link closely with any of the manuscript groups.</p>

shifted between books. The maximum parsimony results for each of the books are shown in Fig. 4. Manuscript Q does not include Book II or Book III, and contains only part of Book I. The manuscript groupings are still largely retained in the individual books, but there are some minor differences. For example in the complete tradition, manuscript C grouped with the M, S, H, and Z manuscripts

with a fairly low level of support, and this grouping was retained only in Books II and III. However, a clear shift was identified with manuscript A which grouped strongly with T (100% support) in Book I, weakly with T (58% support) in Book II, and with E, R (78% support) in Book III. Although there was some support for an A/T/K group (70%) in the complete tradition, the A/T/K grouping was

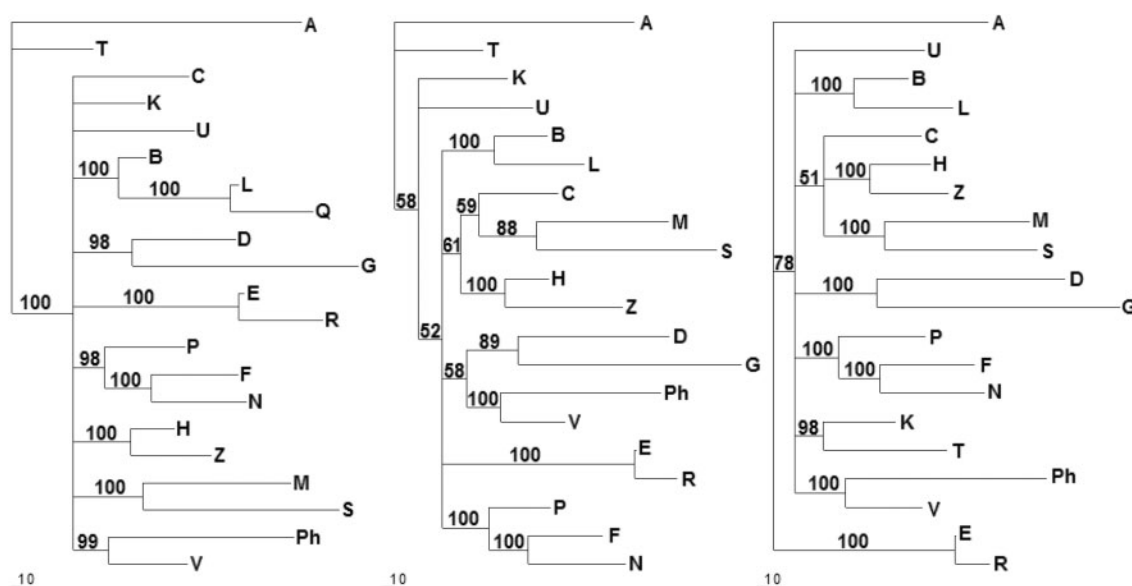


Fig. 4 The maximum parsimony trees for Book I (chr 1–4,659), Book II (chr 4,660–10,209), and Book III (chr 10,210–16,548) of *Monarchia*

not supported in any of the individual books, but there was strong support (98%) for a T/K pair once the A manuscript had moved away towards E/R. The exact relationship between K and T is indicated by traditional textual analysis, which indicates that the grouping of K and T in Book III is caused by both texts being far more accurate than all other extant texts in Book III, whilst not necessarily indicating that they are both derived from a common ancestor other than the archetype of the complete tradition. The strong linkage of T with A in Book I and then with K in Book III may result in the A/T/K group seen in the complete data.

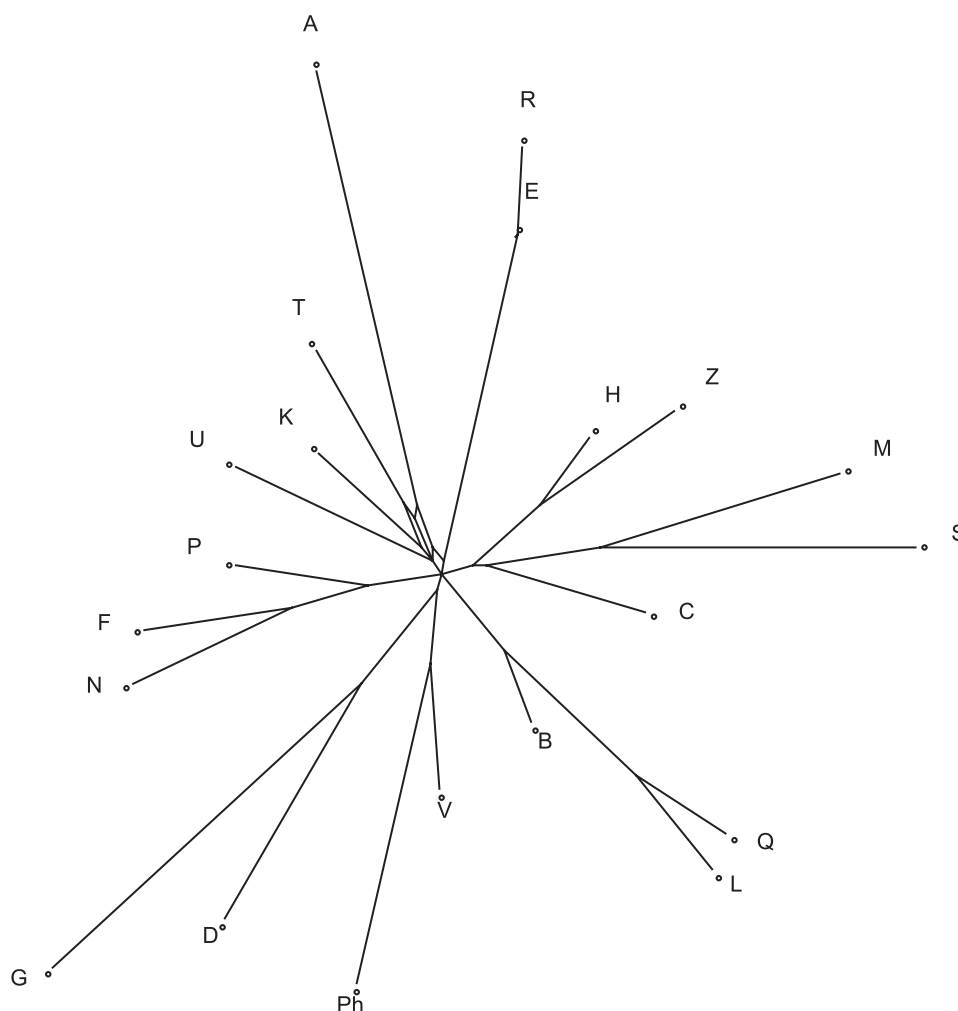
The dual affiliation of manuscript A could also be seen in the overall supernetwork analysis of the three individual maximum parsimony trees (Fig. 5). Manuscript A occurred between, and linked to, both manuscript T and the E/R manuscript pair.

Overall, the phylogenetic analyses gave support to the witness affiliations that have been determined by traditional analysis. In addition to determining the main manuscript groups, it suggested that manuscript A shifts its affiliation from the non- $\beta$ -manuscript T, in the first half, to the

$\beta$ -manuscripts, especially E/R, in the second half of the text.

### 3.2.2 Further analyses of selected manuscripts

The maximum chi-squared and percentage differences analyses were used to investigate chosen pairs of manuscripts. By detecting possible 'recombination' break points, the maximum chi-squared analysis can reveal likely instances of exemplar change, while the percentage of differences plot indicates whether the two manuscripts become more similar or more different after a possible break point, and gives a profile of the level of similarity between the two manuscripts throughout the data file. For technical reasons, the probability values required for statistical significance are higher when considering specific manuscripts (selected because they appear to shift their affiliation throughout the text) than for purely exploratory analyses which may be made between any pair of manuscripts. The significance thresholds that apply to the probability ( $P$ ) values were calculated using the Dunn–Sidak correction for multiple testing (Sokal and Rohlf, 1995, p. 239). Note that only probability values that are *smaller* than the



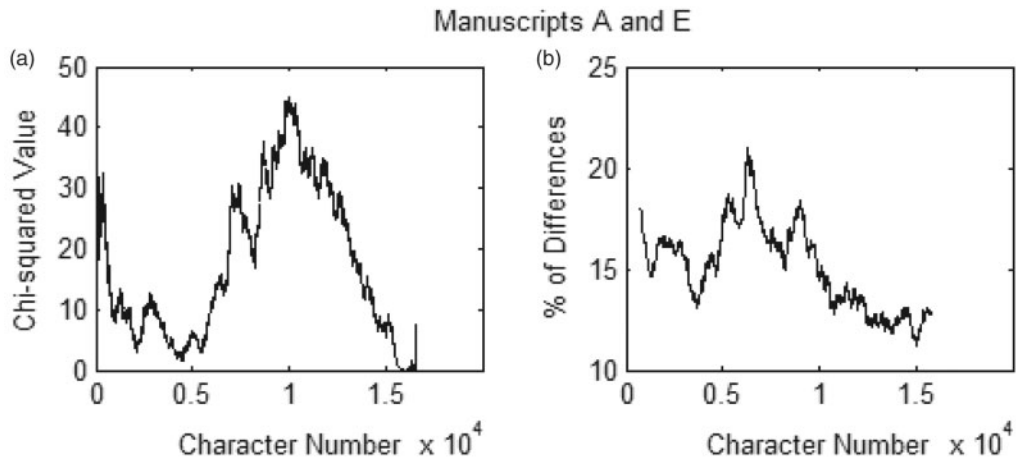
**Fig. 5** The Supernetwork graph of the individual maximum parsimony trees for Books I, II, and III. Branch lengths are proportional to the mean number of substitutions inferred

significance threshold are statistically significant. The analyses of manuscripts using the maximum chi-squared and percentage differences methods were again compared with the results of traditional studies.

**3.2.2.1 Manuscripts A and E.** Manuscript A was the only manuscript to show a clear shift of affiliation in the phylogenetic analyses of the individual books. The results of the maximum chi-squared and percentage difference analyses are shown

in Fig. 6. For this analysis, the significance threshold value =  $1.2 \times 10^{-3}$ . The maximum chi-squared value for the comparison of A and E was 45.1, occurring at character number 10,011 in the NEXUS file. The *P* value of  $8.25 \times 10^{-7}$  was much smaller than the significance threshold of  $1.2 \times 10^{-3}$ . This result could certainly be consistent with a change of exemplar occurring within Book II.

Manuscript A becomes increasingly similar to manuscript E, with Book III being more similar than Book I. The highest percentage of differences



**Fig. 6** Plot of chi-squared values (a) and the percentage of differences (b) against the location in the text for the manuscripts A and E. In the chi-squared chart, the highest peak occurs at the location most likely to be the site of manuscript recombination. The maximum chi-squared value = 45.1 at character number 10,011;  $P = 8.25 \times 10^{-7}$ . The distribution of differences chart shows the percentage of differences in the region of coded text contained in a moving window (of 1,501 characters) against the character number at the mid-point of the window

**Table 4** A comparison of the results of traditional and computer-based analysis of manuscripts A and E

Traditional Analysis	Maximum chi-squared and percentage of differences
Initially manuscript A groups with the non- $\beta$ -manuscripts. A shares a common exemplar with manuscript T up to Book (Section) II chapter vii para 8. e.g. In para 4 (NEXUS chr 7941) 'sacris' in A <sup>1</sup> , K and T 'sanctis' in $\beta$ -mss Then in para 8 (NEXUS chr 8090) 'qui' in T and K 'quod' in A <sup>2</sup> and $\beta$ -mss And in para 8 (NEXUS chr 8102) 'ut' in T and K 'quod' in A <sup>2</sup> and $\beta$ -mss These, and other examples, confirm that A shifts affiliation to the $\beta$ -manuscripts. It is particularly close to the $\beta$ 3-texts.	The shift of manuscript A from the non- $\beta$ -manuscripts to the $\beta$ 3-manuscripts (especially E) prompted the analysis. The chi-squared analysis gave a statistically significant (although broad) peak indicating a possible break point at chr 10,011. The manuscripts became increasingly similar from part way through Book II.

between manuscripts A and E occurs in Book II. A comparison between the results of traditional analysis and the results from these programs is shown in Table 4. There is good agreement, although the traditional analysis identified a change of exemplar somewhat earlier in Book II (chr 8,090—Book II vii 8) than did the computer analyses (chr 10,011—Book II xi 6). There are two likely reasons for this.

First, the maximum chi-squared peak in this analysis was broad, and flanked by lesser peaks of almost the same height, making it hard to locate the breakpoint exactly. Second, the computer analysis looks at the data purely numerically, but makes no textual judgement. The analysis can speedily point to an area of interest, but the exact location of a break may need to be refined by textual analysis of

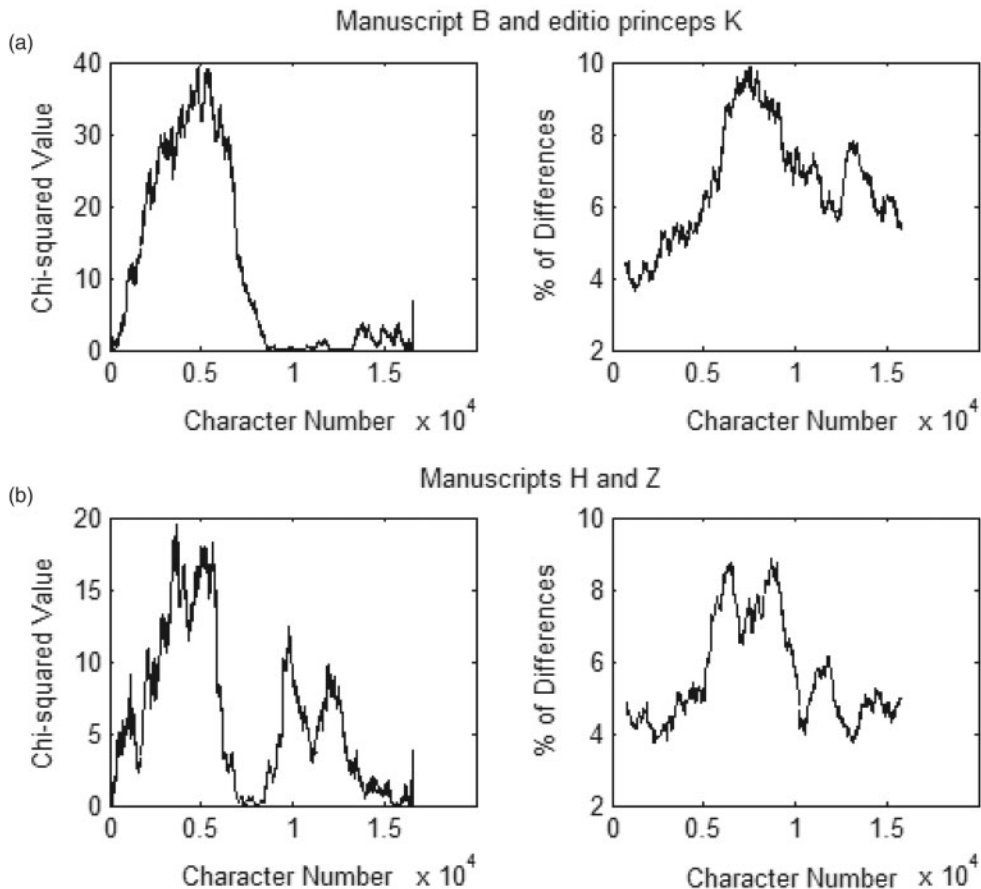


Fig. 7 Plot of chi-squared values and the percentage of differences (window size 1,501 characters) for pair-wise comparison of selected witnesses. (a) B/K—maximum chi-squared value = 39.44 at character number 5,347;  $P = 7.88 \times 10^{-4}$  and (b) H/Z—maximum chi-squared value = 19.54 at character number 3,670;  $P = 2.80 \times 10^{-2}$

individual variants. However, the combined results of the computer programs have enabled the recognition of a possible exemplar change, the identification of the affiliations both before and after the breakpoint, and its approximate location.

### 3.2.3 Exploratory analyses of other manuscripts

Manuscript A was the only manuscript that shifted its affiliation in the phylogenetic analyses. Exploratory analyses were carried out with other manuscripts, but these are only considered significant if the  $P$ -value is smaller than a lower threshold value determined on the basis of a possible full pair-wise comparison of each manuscript with all other

manuscripts. In this case, with twenty-one manuscripts, a full pair-wise comparison would have a significance threshold of  $2.4 \times 10^{-4}$ .

Various pairs of manuscripts from throughout the tradition were compared.

#### 3.2.3.1 Manuscript B and the editio princeps, K

The results are shown in Fig. 7a of a comparison involving the important witnesses B and K. Manuscript B is a fourteenth century copy, the Codex Bini, and K is the *editio princeps* of 1559. There is a large peak in the percentage of differences occurring in Book II. This peak of differences in Book II was found in almost all pair-wise

**Table 5** Number of variant forms occurring in the complete tradition in quotations from *Pharsalia* found in Book II

Total number of variants in all manuscripts in a quotation from Book II iv 6 showing locations where H and Z differ from each other (+)									
Illa profecto sacrificio cecidere Nume, que lecta iuventus patritia cervice movet;									
2	3	2	1	4	11	4	2	5	
	+	+		+					+
spoliaverat hauster, aut boreas populos ancilia nostra ferentes.									
	3	2	1	4	2	4	2	1	
						+			
(Pharsalia ix, 477-80)									
<div><b>Translation</b> - No doubt the shields, Which chosen youths bore on patrician necks, Fell before Numa as he sacrificed; The South wind or the North had robbed their bearers Of shields which now are ours.</div>									

comparisons throughout the tradition, giving rise to a maximum chi-squared peak in the region of characters 4,000–6,000. There was also a general tendency for Book I to be the region of greatest similarity.

**3.2.3.2 Manuscripts H and Z.** The results for the manuscripts H and Z are shown in Fig. 7b. These two manuscripts grouped together with 100% support in the phylogenetic analyses of both the complete tradition and the individual books. The percentage of differences is approximately doubled in Book II, although it did not cause the manuscripts to re-group.

Book II of *Monarchia* differs from both Books I and III, in that its account of Roman history is backed by poetic testimony involving a large number of quotations. These quotations are written in classical Latin, rather than the more familiar medieval Latin of the rest of the text, and often proved difficult for the scribes to copy accurately (Shaw, in press). The quotations, and the prose text which surrounds them, also contain many names. Although simple spelling variants were

removed during the collation and regularization of the files, there were still many aberrant versions of some names, too idiosyncratic to be regarded as simple spelling variants, and this gave rise to a far higher observed error rate in this region of the text.

Manuscripts H and Z form a close and constant pair, and so there was no reason to expect that they would differ greatly in all of the embedded quotations, but several of them in regions of generally high variability were examined. Tables 5 and 6 shows the results for two quotations. The Latin texts and translations are taken from the DVD-ROM (Shaw, 2006) and the numbers represent the number of versions of that word as it occurs in the NEXUS file. The ‘+’ sign indicates locations where manuscripts H and Z differ from each other. The first quotation (Table 5) is taken from Lucan’s *Pharsalia*, and occurs in Book II chapter iv paragraph 6. The second (Table 6) is taken from Cicero’s *De Officiis* and occurs in Book II chapter vii paragraph 12. A comparison of the proportion of sites that are different between manuscripts H and Z in the selected quotations, and in each book as a whole, is shown in Table 7. In this table, the number of differences is expressed as a fraction of



**Table 6** Number of variant forms occurring in the complete tradition in quotations from *De Officiis* found in Book II

Total number of variants in all manuscripts in a quotation from Book II vii 12 showing locations where H and Z differ from each other (+)									
Scite Crisippus, ut multa: 'qui stadium' inquit 'currit, eniti et contendere debet									
4	3	1	1	1	2	1	13	3	1
	+						+		
quam maxime possit ut vincat; supplantare eum quicum certet nullo modo debet'									
	5		2	3	2	3	13	3	1
	+		+				+	+	
(De Officiis, III, 42)									
<div> <b>Translation-</b> With his customary aptness Chrysippus says: 'When a man races in the arena he must exert himself and strive his hardest to win; he must not in any way obstruct his fellow-competitor.' </div>									

**Table 7** A comparison of the proportion of differences between manuscripts H and Z in all three books, and also in two quotations found in Book II

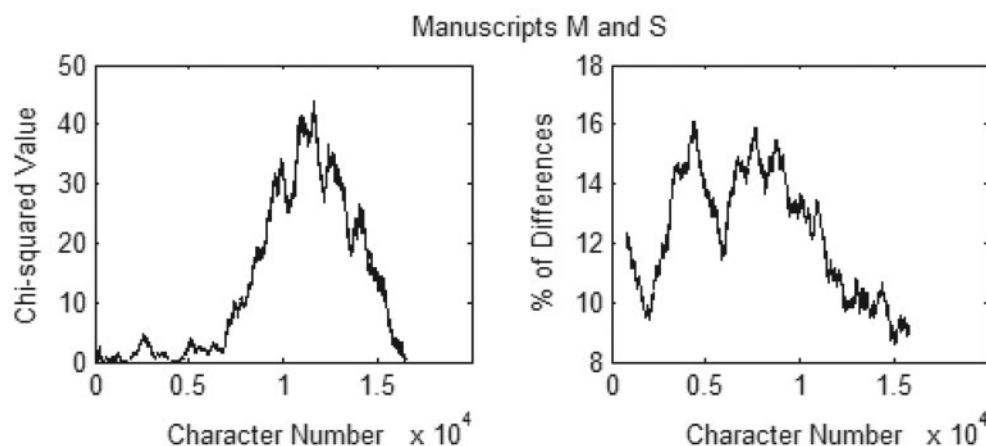
	Number of differences	Proportion
Quotation 1	5/17	0.29
Quotation 2	6/18	0.33
Book I	201/4,558	0.04
Book II	390/5,437	0.07
Book III	305/6,240	0.05

the number of variant locations present in both manuscripts. Any missing words or phrases in either manuscript reduce this total from the maximum number of characters for that book. The results show that the two quotations have a higher proportion of differences than the overall proportion for each section, and that Book II, as a whole, contains almost double the proportion of differences of either Book I or Book III. Overall, these results, while only directly reflecting the situation for manuscripts H and Z, are consistent with the hypothesis that the quoted material in Book II is the cause of the peak in the percentage of differences plot noticed in almost all of the pair-wise comparisons of manuscripts.

In both the B/K and H/Z comparisons, the *P*-values for the chi-squared peaks are above the threshold value for statistical significance, and so there is no indication of a shift in exemplar throughout the text. This is in agreement with the results of the phylogenetic analyses.

**3.2.3.3 Manuscript S.** There was some similarity between the maximum chi-squared and the percentage of differences plots for manuscripts S/M and for manuscripts A/E. In both cases, the region of greatest similarity occurred in Book III, and the maximum chi-squared value occurred near the start of Book III (Fig. 8). The *P*-value indicated that the result for manuscripts S and M is statistically significant.

When a different witness, K, was compared with S, a significant chi-squared peak also occurred towards the end of the text and the two witnesses were again closest in Book III. In contrast, a comparison of K with M gave very low chi-squared values (results of K/M and K/S analyses not shown). Manuscript S is a corrupt and degraded manuscript and traditional analysis has been divided in its view of the manuscript's textual



**Fig. 8** Plot of chi-squared values and the percentage of differences (window size 1,501 characters) for pair-wise comparison of manuscripts M and S. The maximum chi-squared value = 43.88 at character number 11,592;  $P = 5.07 \times 10^{-6}$

history, and the results of the computer analyses are also difficult to interpret. While the results of the chi-squared analysis would be consistent with manuscript S changing its exemplar to become more similar to other witnesses in Book III, the two manuscripts S and M formed a constant pair in all of the phylogenetic analyses, with no apparent shift of affiliation. Shaw (2006) has shown that S was copied from only one exemplar.

**3.2.3.4 Manuscript D.** In the phylogenetic analysis, manuscript D always grouped with G, and they were loosely networked to the S/M pair. All four manuscripts had long branch lengths (indicating a large number of unique variants), and the S/M pair grouped with the other  $\beta_4$ -manuscripts C/H/Z in the maximum parsimony analyses. This is consistent with the results of the traditional analyses (Ricci, 1965; Shaw, 2006) which indicated that D was a contaminated manuscript containing features of the  $\beta_3$ -group (mainly G) and the  $\beta_4$ -group (mainly M).

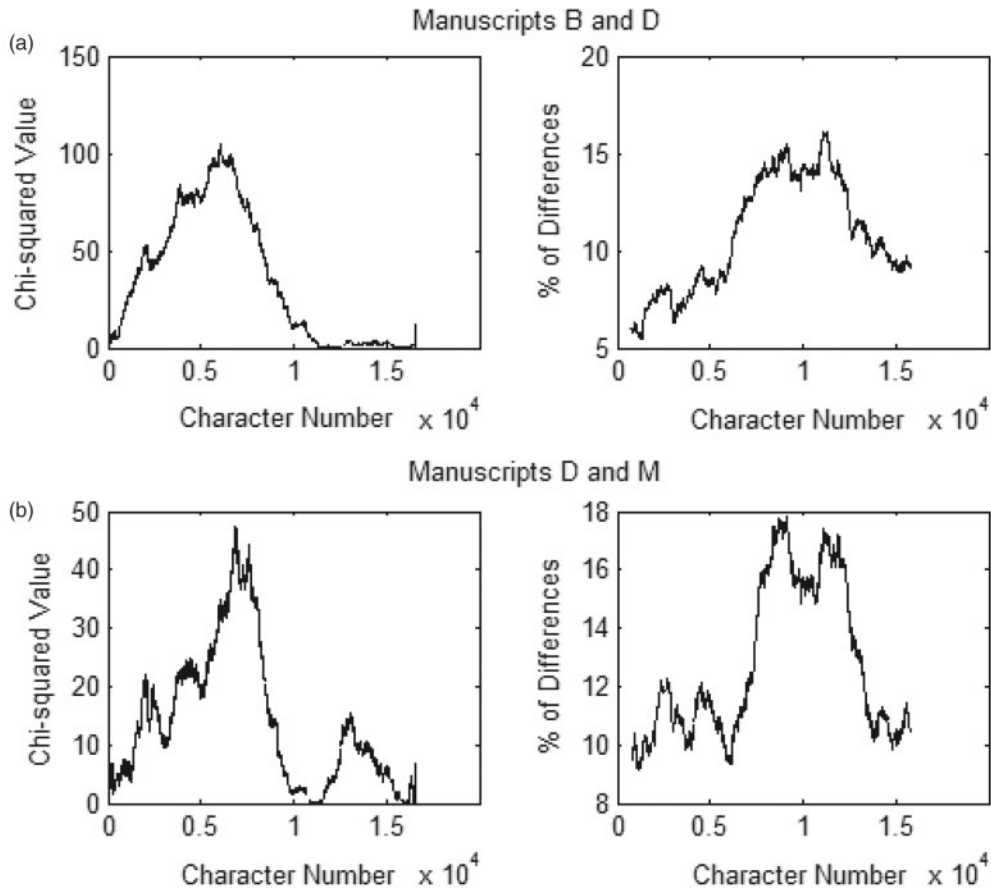
Several exploratory analyses were performed using manuscript D, and two of the results are shown in Fig. 9a and b. These analyses shared two main features. First, the  $P$ -values were generally very low, giving results that were highly statistically significant. Second, the increase in differences usually associated with Book II was present but peaked at

a higher character number than with many other analyses. The maximum chi-squared value also occurred at a consistently high character number, usually between characters 6,000 and 7,000, and some 1,500 characters after the start of Book II.

Traditional analysis (Shaw, 2006) revealed a possible explanation for the unusual results obtained with manuscript D. A section of text spanning the latter part of Book II and the first part of Book III—f.41vb (Book II, chapter vi, line 2—NEXUS chr 7,455) to f.51r (Book III, chapter viii, line 5—NEXUS chr 12,879) appears to be written in a tighter hand than the remaining text. This section of text contains more errors and unusual readings than are typical for the rest of the manuscript, and may be the work of another scribe. The combination of the normal increase in differences in Book II and the scribal hand change in Book II/III, caused manuscript D to move much more drastically away from the other manuscripts, with the peak of differences and maximum chi-squared peak occurring somewhat later than for many manuscript pairs.

## 4 Discussion

When the computer analyses were performed and interpreted, the researchers did not have access to any of the results of the traditional textual analysis



**Fig. 9** Plot of chi-squared values and the percentage of differences (window size 1,501 characters) for pair-wise comparisons involving manuscript D. (a) B/D—maximum chi-squared value = 104.83 at character number 6,100;  $P = 2.21 \times 10^{-12}$  and (b) D/M—maximum chi-squared value = 47.27 at character number 6,885;  $P = 3.00 \times 10^{-6}$

and had no knowledge of the identity of the text. The data were supplied as a numerically coded table of variants with no transcript material available. While the list of variants supplied could have been used to identify key words, no such attempts were made. The numbering of the variant list indicated that the text was divided into three sections and these divisions were used to produce the sections for analysis. It was later found that these sections corresponded with the three books of *Monarchia*. After the work was completed, the DVD-ROM and *Monarchia* literature were consulted so that the results of the computer analyses could be compared with those of traditional scholarship.

#### 4.1 Phylogenetic analysis of all texts

There was very good agreement between the maximum parsimony bootstrap tree and the traditional stemma (Appendix II). The phylogenetic analyses were able to identify not only all of the main manuscript groups as determined by traditional analysis, but also the subdivisions of these groups into smaller subgroups and manuscript pairs. The phylogenetic methods predicted a shift in affiliation (or possible change of exemplar) matching exactly the conclusions of the traditional studies. The maximum parsimony results for the complete tradition and for the individual books showed that manuscript A appeared to shift affiliation from T to E/R

in the second book. This shift had also been apparent from the traditional stemmatic analysis. It was well documented, and had been noted as far back as 1965 (Ricci, 1965).

## 4.2 Pair-wise comparisons of witnesses

- (a) The chi-squared and percentage differences results indicated that the shift of manuscript A from T to E/R occurred in Book II. The computer programs were able to identify the manuscripts involved in the shift of exemplar and to show the direction of the exemplar change, and the approximate site of the shift, although the predicted breakpoint occurred somewhat later in the book than the breakpoint indicated by traditional analysis. The phylogenetic results must be interpreted in the light of knowledge of the text in this region. The exact location of the exemplar change can be pin-pointed by detailed textual analysis and understanding of the text.
- (b) Many of the manuscript pairs were less similar in Book II than in either of the other books. The content of Book II with its many quotations written in the less familiar and much more difficult classical Latin provides an explanation for the increased level of differences found in this region of the text.
- (c) The chi-squared and percentage differences results for S/M shared some features with the A/E results in that, in both cases, the manuscript pairs were at their most similar in Book III, and a statistically significant maximum chi-squared value occurred either early in Book III (S/M), or very near the end of Book II (A/E). However, these results are difficult to interpret in the case of S/M as, unlike the A/E pair, S/M formed a constant, well-supported pair in all of the phylogenetic analyses.
- (d) Manuscript D appeared unusual, with its peak of differences occurring generally later than with other manuscript pairs. Manuscript D showed the normal increased percentage of differences at the start of Book II, but

superimposed on this, there is a section of text spanning the latter part of Book II and the first part of Book III, in which the handwriting suggests there may have been a change of copyist. This section is noticeably more error-prone than the rest of the text. These two features explain the strong increase in differences and high maximum chi-squared peaks found in this region of the manuscript. Manuscript D is a contaminated text sharing material with the  $\beta$ 3- and  $\beta$ 4-manuscript groups.

Overall, the results of the computer-based and traditional methods are in good agreement with each other. The computer analysis can be a very useful tool for the rapid identification of possible regions of interest, which can then be further examined by more traditional methods. The computer methods are rapid and flexible, allowing a variety of analyses to be performed from one data set, and for specific variants or manuscripts to be included or excluded from the analyses. They can quickly assimilate a large amount of data and, even when there are conflicting signals, can generate a tree showing the relationships between the texts. However, these trees are unrooted, and the origin of the tradition can only be ascertained by textual analysis. Although the computer methods cannot make any literary or textual judgement on the material, they can work very efficiently in conjunction with other methods to facilitate an understanding of a textual tradition.

## Acknowledgements

This work is part of the TEXTNET project, funded by the Leverhulme Trust. The authors are grateful to Matt Spencer, Adrian Barbrook, Barbara Bordalejo, and Peter Lockhart for many helpful discussions and suggestions.

## Funding

The Leverhulme Trust

## References

- Barbrook, A. C., Howe, C. J., Blake, N., and Robinson, P. (1998). The phylogeny of *The Canterbury Tales*. *Nature*, **394**: 839.
- Baret, Ph.V., Macé, C., and Robinson, P. (2006). Testing methods on an artificially created textual tradition. In Macé, C., Baret, Ph.V., Bozzi, A., and Cignoni, L. (eds), *The Evolution of Texts: Confronting Stemmatalogical and Genetical Methods*. *Linguistica Computazionale*, Vol. XXIV–XXV. Pisa: Accademia editoriale, pp. 255–83.
- Bryant, D. and Moulton, V. (2002). NeighborNet: an agglomerative method for the construction of planar phylogenetic networks. In Guigo, R. and Gusfield, D. (eds), *WABI 2002, Lecture Notes in Computer Science*, Vol. 2452. Berlin: Springer, pp. 375–391.
- Bryant, D. and Moulton, V. (2004). Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, **21**(2): 255–65.
- Cartlidge, N. (2001). *The Canterbury Tales* and cladistics. *Neuphilologische Mitteilungen*, **102**: 135–50.
- Contini, G. (1986). *Breviario di ecdotica*. Milano: Napoli.
- Eagleton, C. and Spencer, M. (2006). Copying and conflation in Geoffrey Chaucer's treatise on the astrolabe: a stemmatic analysis using phylogenetic software. *Studies in History and Philosophy of Science*, **37**(2): 237–68.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**: 783–91.
- Hanna, R. (2000). The application of thought to textual criticism in all modes - with apologies to A. E. Housman. *Studies in Bibliography*, **LIII**: 163–72.
- Hillis, D. M. and Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*, **42**: 182–92.
- Howe, C. J., Barbrook, A. C., Spencer, M., Robinson, P., Bordalejo, B., and Mooney, L. R. (2001). Manuscript evolution. *Trends in Genetics*, **17**: 147–52.
- Huson, D. (1998). SplitsTree: a program for analyzing and visualizing evolutionary data. *Bioinformatics*, **14**: 68–73.
- Huson, D. H., Dezulian, T., Klopper, T., and Steel, M. A. (2004). Phylogenetic super-networks from partial trees. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **1**: 151–58.
- Huson, D. H. and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**(2): 254–67.
- Lantin, A.-C., Baret, P. V., and Macé, C. (2004). Phylogenetic analysis of Gregory of Nazianzus' Homily 27. In Purnelle, G., Fairon, C. and Dister, A. (eds), *Le poids des mots: Proceedings of the 7th International Conference on the Statistical Analysis of Textual Data*. Louvain-la-Neuve: Presses Universitaires de Louvain, pp. 700–07.
- Lee, A. R. (1989). Numerical taxonomy revisited: John Griffith, cladistic analysis and St. Augustine's Quaestiones in Heptateuchem. *Studia Patristica*, **20**: 24–32.
- Maddison, D. R., Swofford, D. L., and Maddison, W. P. (1997). NEXUS: An extensible file format for systematic information. *Systematic Biology*, **46**: 590–621.
- Manly, J. M. and Rickert, E. (eds) (1940) *The Text of the Canterbury Tales: Studied on the Basis of All Known Manuscripts*. Chicago: University of Chicago Press.
- Maynard Smith, J. (1992). Analysing the mosaic structure of genes. *Journal of Molecular Evolution*, **34**: 126–29.
- Mooney, L. R., Barbrook, A. C., Howe, C. J., and Spencer, M. (2001). Stemmatic analysis of Lydgate's 'Kings of England': a test case for the application of software developed for evolutionary biology to manuscript stemmatics. *Revue d'Histoire des Textes*, **31**: 275–97.
- Page, R. D. M. and Holmes, E. C. (1998). *Molecular Evolution. A Phylogenetic Approach*. Oxford: Blackwell Science Ltd.
- Peri (Pflaum), H. (1956). Une méthode expérimentale de critique des textes. In *Atti del VIII° Congresso Internazionale di Studi Romanzi*, Vol. II, Florence: Sansoni.
- Petrocchi, G. (1966–67). *Dante Alighieri, La Commedia secondo l'antica vulgata*, edited by Giorgio Petrocchi, 4 vols. (Le opere di Dante Alighieri. Edizione Nazionale a cura della Società Dantesca Italiana), Verona: Mondadori.
- Platnick, N. I. and Cameron, H. D. (1977). Cladistic methods in textual, linguistic, and phylogenetic analysis. *Systematic Zoology*, **26**: 380–85.
- Robinson, P. M. W. and O'Hara, R. J. (1996). Cladistic analysis of an Old Norse manuscript tradition. In Hockey, S. and Ide, N. (eds), *Research in Humanities Computing*, Vol. 4. Oxford: Oxford University Press, pp. 115–37.

- Salemans, B. J. P.** (2000). *Building Stemmas with the Computer in a Cladistic, Neo-Lachmannian, Way*. Ph.D. thesis, Katholieke Universiteit Nijmegen.
- Shaw, P.** (1995). *Dante, Alighieri: Monarchia*. (Edited and translated by P. Shaw) Cambridge Medieval Classics 4. Cambridge: Cambridge University Press.
- Shaw, P.** (1996). The *stemma codicum* of Dante's *Monarchia*. *Italian Studies*, **LI**: 5–26.
- Shaw, P.** (ed.) (2006). *Dante Alighieri: Monarchia*. An electronic edition on DVD-Rom jointly published by SDE-SDI (Scholarly Digital Editions and the Società Dantesca Italiana), Leicester.
- Shaw, P.** (ed.) (in press). *Dante Alighieri. Monarchia* (Edizione Nazionale delle opere di Dante Alighieri a cura della Società Dantesca Italiana, vol. V/1), Florence: Le Lettere.
- Sokal, R. R. and Rohlf, F. J.** (1995). *Biometry*. New York: W.H. Freeman and Co.
- Spencer, M., Davidson, E. A., Barbrook, A. C., and Howe, C. J.** (2004). Phylogenetics of artificial manuscripts. *Journal of Theoretical Biology*, **227**: 503–11.
- Stolz, M.** (2003). New philology and new phylogeny: aspects of a critical electronic edition of Wolfram's *Parzival*. *Literary and Linguistic Computing*, **18**: 139–50.
- Swofford, D. L.** (2001). *PAUP\*. Phylogenetic Analysis Using Parsimony (\*and other methods)*. Sunderland, MA: Sinauer Associates.
- Timpanaro, S.** (1985 [1963]). *La genesi del metodo del Lachmann*. Florence: Le Monnier, 1963. Revised edition Padua: Liviana.
- Windram, H. F., Howe, C. J., and Spencer, M.** (2005). The identification of exemplar change in the *Wife of Bath's Prologue* using the maximum chi-squared method. *Literary and Linguistic Computing*, **20**: 189–204.

## Appendix I

### IMPORTANT critical editions of *Monarchia*

**Dantis Aligherii de Monarchia** libri III, codicum manuscriptorum ope emendati per Carolum Witte. Vindobonae 1874.

**Dantis Alagherii De Monarchia** libri III, recensuit Ludovicus Bertalot. Friedrichsdorf in monte Tauno apud Francofurtum 1918; Gebennae 1920.

**Alighieri, Dante.** *Monarchia*, a cura di E. Rostagno (in *Le Opere di Dante. Testo critico della Società Dantesca Italiana*. Firenze 1921; second edition 1960).

**Alighieri, Dante.** *Monarchia*, a cura di Pier Giorgio Ricci (Edizione Nazionale delle opere di Dante Alighieri a cura della Società Dantesca Italiana, vol. V). Milano 1965.

**Alighieri, Dante.** *Monarchia*. Edited and translated by Prue Shaw. Cambridge 1995 (Cambridge Medieval Classics 4).

**Shaw, P.** (ed.) (2006). *Dante Alighieri: Monarchia*. An electronic edition on DVD-Rom jointly published by SDE-SDI (Scholarly Digital Editions and the Società Dantesca Italiana), Leicester.

**Alighieri, Dante.** *Monarchia*, edited by Prue Shaw (Edizione Nazionale delle opere di Dante Alighieri a cura della Società Dantesca Italiana, vol. V/1), Florence (in press).



## Appendix II

Traditional analysis of the *Monarchia* tradition (Shaw, 2006) resulted in the following stemma (Shaw, in press). Hypothetical ancestors are

represented as  $x_1$ ,  $x_2$ , etc. There is some very slight textual indication for a possible link between  $\beta_2$  and  $\beta_3$ .

\*U—Manuscript U is contaminated with  $\beta_1$  and non- $\beta$ .

