Analysis Workshop

Peter Robinson

1. The aims of this workshop

This section presents, in workshop form, instances of our use of the various tools at our disposal to explore the textual tradition of The General Prologue of *The Canterbury Tales*. The workshop form and title are deliberately chosen. This is not intended as any kind of definitive statement as to the history of The General Prologue tradition (and still less, the history of the whole *Canterbury Tales* tradition). Rather, it should be seen as an introduction to our aims in attempting this analysis; as a description of the tools we use in this analysis; as a series of examples of the use of these tools; and as an opportunity for the reader to use some of these tools for himself or herself. The reader may also, from use of these tools, be able to judge how far they make possible a historical analysis of the textual tradition. Thus, the reader himself may use these tools, and exercises interspersed through the text offer instructions and suggestions on their employment. The software tools (VBase and SplitsTree) are provided on this CD-ROM and will be activated when you click on appropriate points in the text of the CD-ROM.

One aim of this workshop is to show that no definitive statement concerning The General Prologue tradition is possible. We are limited in part by the tools we must use, powerful though these may be. We are limited too, less obviously but perhaps more pervasively, by the quality of the evidence that we have gathered: in deciding not to transcribe this or that mark in certain witnesses, we may have lost a vital piece of evidence regarding the relationship of those witnesses. But most of all, we are limited by the nature of the case itself: by the ambiguities in the evidence itself and by the uncertainty of its precise application. Some of these ambiguities result from the well-known problems posed by contamination, by shift of exemplar and by accidental agreement. Other ambiguities arise from an unexpected direction: from witnesses which are such good copies of one another that the evidence of disagreement is insufficient to maintain any hypothesis of exact relationship.

The question arises: is there a point to this analysis, if no closure is possible because of these limitations? Just because there are limits to our knowledge, does not mean that there is no knowledge. In a final section, we show how the insights won from use of these tools might be used to illuminate wider issues and particular readings in the text of the *Tales*. Firstly, we briefly compare (without attempting certainty) the picture of the textual tradition emergent from this analysis of The General Prologue with that given by the earlier analysis of The Wife of Bath's Prologue. Second, in the separate Stemmatic Commentary section of this CD-ROM we present instances of how this knowledge might clarify particular readings in The General Prologue. The Stemmatic Commentary is therefore dependent on the analysis presented in this workshop, and should be read alongside it.

This final section and the Stemmatic Commentary are therefore intended as illustrative of the way in which this analysis might help us read with more intelligence. Elsewhere, I have called this approach 'new stemmatics'[1] : our aim is not to help editors edit, but to help readers read. Of course, a different analysis would lead to a different reading. We hope that by presenting our analysis as a workshop we will encourage readers to make their own analyses, and their own reading.

---

2. The Project's aims and its tools

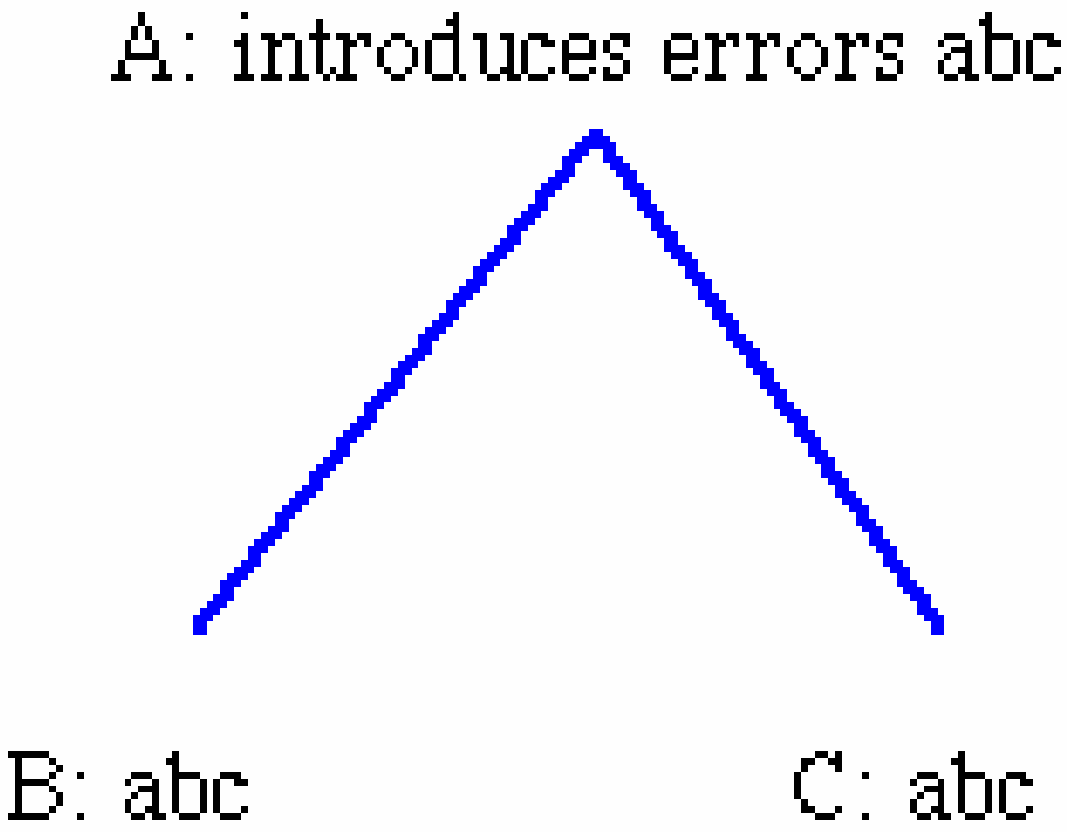
The aim of the *Canterbury Tales* Project, from the first, has been to use all the materials it is gathering — all the transcripts and collations — in order to arrive at a clearer picture of the history of the textual tradition. A crucial factor in the conception of the Project is the belief that computer-based methods of analysis may help towards this aim. The amount of information concerning the various readings in the various witnesses is so huge (around 300,000 pieces of information in The General Prologue alone) as to overwhelm manual methods of analysis. It is precisely in handling such vast arrays of information that computers excel.

Accordingly, the *Canterbury Tales* Project has assembled various computer-based techniques, and sought to apply these to the task of understanding the history of the tradition. Some of these methods were

first developed in other fields of knowledge, notably in evolutionary biology and in mathematical theory. Their use to explore this textual tradition rests on our perception of a fundamental identity between the situations facing workers in these fields and ourselves, working with all these witnesses to Chaucer's text. Evolutionary biologists sorting information on the characteristics shared and not shared by particular species, in search of a historical explanation of the relations among these species by descent from common ancestors, are similar to ourselves, as we sort information on the readings shared and not shared by particular witnesses, in search of a historical explanation of the relations among these witnesses by descent from common ancestors. We think this similarity is more than similarity, and is actual identity. Therefore, tools and perceptions developed in one arena (evolutionary biology) can be translated directly for use in another arena (stemmatic analysis). However, this perception itself must be tested, and one function of this workshop is to present instances of our use of these tools, so that the reader may judge how far their use in this new context is justified.

2.1 Is historical analysis of textual traditions possible?

The aim of historical analysis of textual traditions is to try to reconstruct the history of the copying of the text, and hence determine, so far as is possible, its earliest state. This method, also known as stemmatics, is usually identified with the methods espoused by Karl Lachmann and his followers last century. The clearest statements of the techniques of classic stemmatic analysis (for example, the accounts by Paul Maas and by W. W. Greg) presume that if we find two manuscripts (B and C) which agree in the erroneous readings abc, then those two manuscripts will share a common ancestor (A). It was this common ancestor which introduced the erroneous readings abc; the readings then descended from this ancestor into its copies, thus:


A: introduces errors abc
B: abc
C: abc

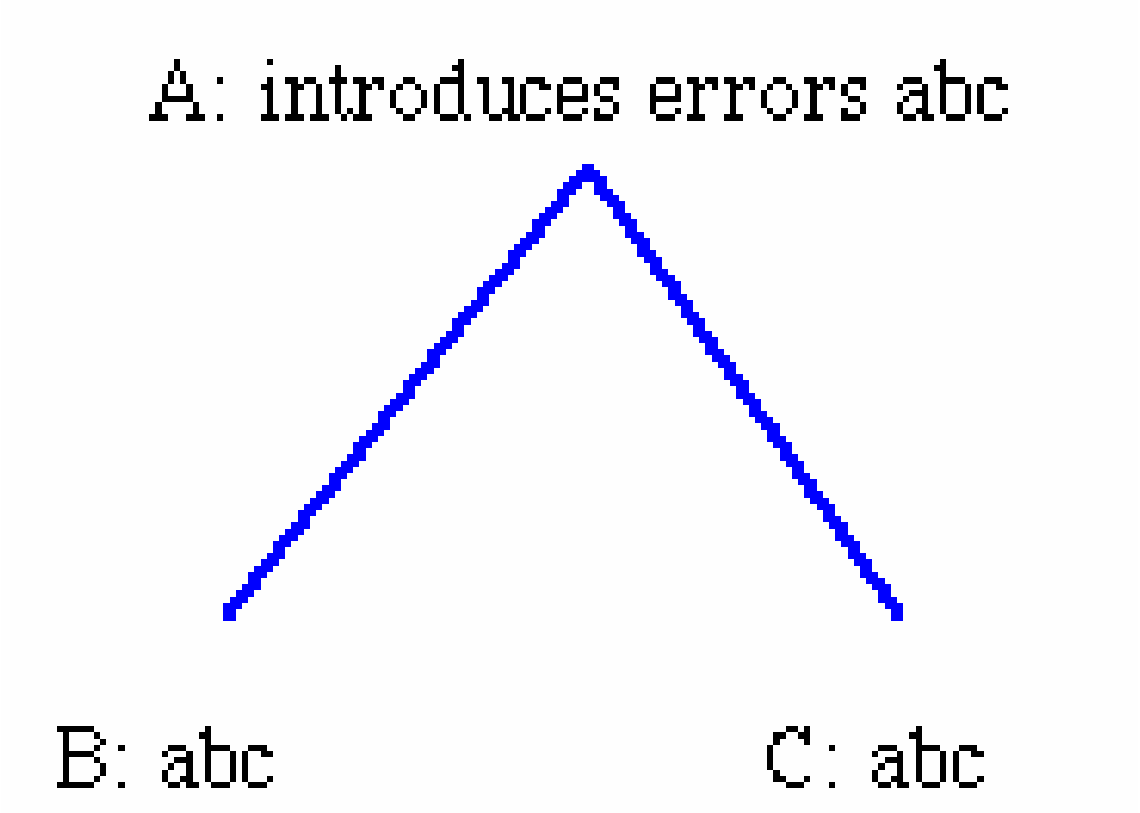


Figure 1: A simple tree of descent

According to this method, an editor should proceed as follows:

1. Identify the erroneous readings in the manuscripts
2. Identify the manuscripts which agree in erroneous readings, and so have common ancestors
3. Thereby, reduce the many manuscripts to a few family groups, each descended from a distinct and exclusive common ancestor
4. Having reduced the manuscripts to a few distinct family groups: when faced with a choice of readings where (say) two groups have reading d and one group has reading e, one may choose reading d as more likely to have been in the archetype.

So described, stemmatics could aspire to the rigour of 'scientific method.' Indeed, the dry tone of Maas's account, and the neo-mathematical notation

used by W. W. Greg, encourage a belief that editing can be made a merely mechanical process, where strict adherence to correct method alone can guarantee a correct text.

Such confidence invites attack, and for over a century scholars have, indeed, attacked. Joseph Bédier pointed out that scholar after scholar found that each manuscript tradition had precisely two, and just two, branches. With high sarcasm, Bédier pointed out how convenient this was for editors: it meant that they did not have to follow the iron rule of 'two branches against one' (as in step 4 above) but could happily choose whichever reading they liked. A. E. Housman, at various points in his *Prefaces to Manilius* and other writings, mocked editors who preferred mathematics to thought: 'manuscripts are to be weighed, not counted' (Housman 1903 and 1922).

In the field of Middle English texts, the most cogent and influential criticism of the stemmatic method, as outlined above, has been by George Kane. In a series of articles and books over the last decades, Kane has argued that historical reconstruction of Middle English vernacular textual traditions on the basis of shared agreement and disagreements is not possible (for example, in his edition of the A text of *Piers Plowman*, 1960, and his comments on Manly and Rickert, 1984). Kane (in company with his co-editor of *Piers Plowman*, Talbot Donaldson) levels two particular criticisms at traditional stemmatics:

1. Step one above requires that the editor determine what readings are erroneous and (by implication) what readings are original, before commencing analysis. As Talbot Donaldson first remarked (1970: 107), if one can identify original readings as this supposes, then the editor might as well just collect these original readings into an edition and not bother with stemmatics at all.
2. Further, Kane asserts that even if one wanted to carry out stemmatic analysis it would not be possible, in Middle English at least. Step two above asserts that when manuscripts do agree in erroneous readings, this must be because they have an exclusive common ancestor.

Kane, along with other critics of stemmatics, observes that stemmatic analysis must fail because of this latter requirement alone. It is a commonplace of stemmatics that very often manuscripts may agree in unoriginal readings for two other reasons than common ancestry.

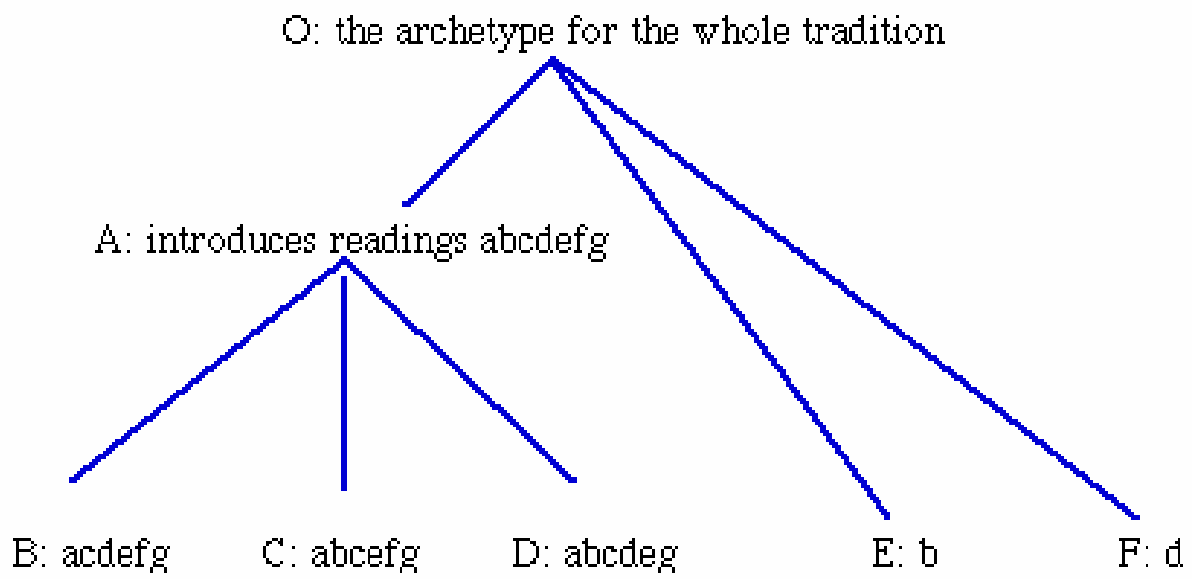
The first reason why manuscripts may agree in erroneous readings other than by common ancestry is contamination. Here, a scribe will deliberately import readings not in the exemplar into the copy. Perhaps the scribe may be using two different manuscripts, and while copying mostly from one will introduce occasional readings from the second. Or, a reader might copy readings from one manuscript into the margin of a second manuscript, and a later scribe copying the second manuscript might move some of these readings from the margin into the text itself. Alternatively, a scribe might remember readings from a manuscript copied or read elsewhere, and introduce these into the text. This circumstance, if at all widespread, would render any kind of stemmatic analysis impossible: it would be impossible to know, for any set of agreements, whether they had arisen through common ancestry or through contamination. Thus, Maas declares in the last words of his book: against contamination there is no specific. The second reason why manuscript might agree in erroneous readings other than by common ancestry is simple coincidence: two (or more) scribes, engaged on copying the same text, might just happen to make the same mistake, quite independently of one another.

One aspect of Kane's argument is specially relevant: that contamination and accidental agreement between manuscripts are so widespread as to make it impossible to extract any meaningful genetic hypothesis concerning the relations between the manuscripts. Kane in his prefaces to his editions of *Piers Plowman* (e.g. that of the A text, 1960) demonstrates how the evidence of apparent groupings of manuscripts evidenced by particular patterns of agreement in particular readings, is (in his opinion) continually undercut by different patterns of agreement in other readings, suggesting quite incompatible groupings.

Clearly, some method must be devised to deal with these problems, if any kind of stemmatic analysis is to be possible. The foundation of the work of The *Canterbury Tales* Project is our belief that computer methods offer ways past these difficulties, and so make possible a historical analysis of the tradition which would not otherwise be available. The discussion of our method which follows is based on that given in O'Hara and Robinson 1993, but amplifies and updates that article in certain respects.

2.2 The concept of fundamental witness groups

The key to analysis as we practice it, and to what I have named the 'new stemmatics', is what we call 'fundamental witness groups.' The theory begins with the perception of Lachmann as described by Maas (see above). If historical reconstruction is to be possible, we must presume that at a particular point in the history of the text, a scribe or editor made a copy we might call A. In that copy A, the scribe introduced various readings not present in the exemplar. Further copies of this copy A were then made, and these copies inherit many of the readings introduced into A by the scribe. We can therefore identify manuscripts which belong to this group because they characteristically include a significant number of the readings introduced in copy A, readings which are not found in other manuscripts. As I explain below, the introduction of the concept of a 'significant number of readings' is a modification of the classic stemmatic theory. In our terminology, the manuscripts which are characterised by possession of these readings form a fundamental group: in this case, the group 'A', sharing an exclusive common ancestor (or 'hypearchetype') in A itself, the manuscript which introduced these readings into the tradition. This may be expressed in graphic terms as follows:



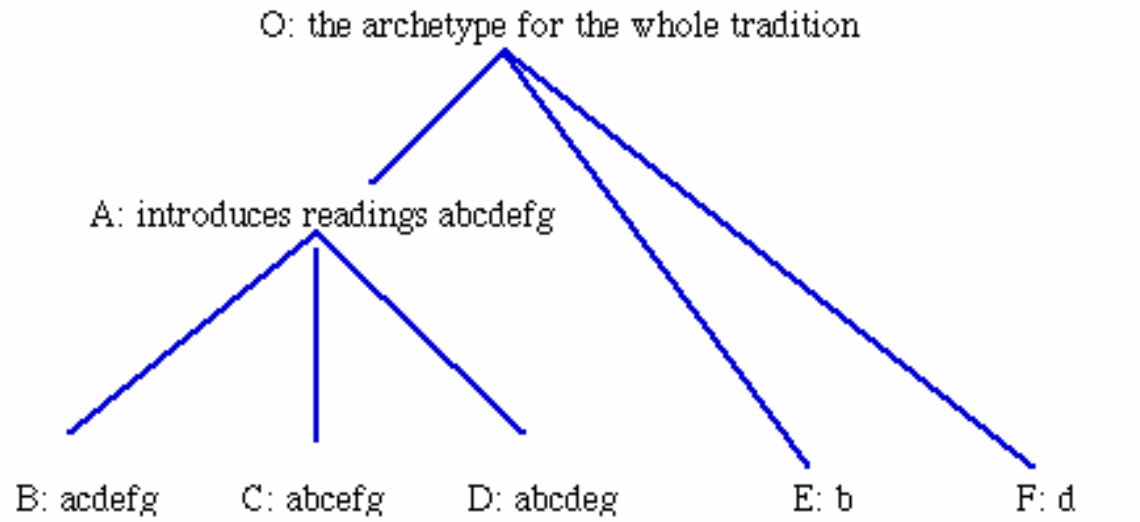


Figure 2: Fundamental witness groups

In our terms, the manuscripts BCD constitute a 'fundamental witness group ' (obviously, if A were still extant, it too would be a member of this group). Notice that in this formulation it is not necessary that each of the witnesses BCD have all of the seven readings abcdefg which we postulate were introduced by A. It is only necessary that each witness have a significant number of these readings: in this case, each of the witnesses BCD has six of the seven. Note too, that some (or even all) of these readings might also appear in manuscripts not descended from A: reading b (but no others) in witness E, reading d (but no others) in F, as shown here. However, because neither E nor F has a significant number of the readings abcdefg (only one each) then this might be simple chance, and neither E nor F might be descended from A.

As I observed above, this introduction of the notion of a 'significant number of readings' represents a modification of the classic stemmatic method, as described by Maas. According to the classic theory, these variants introduced in A should appear in manuscripts descended from A, in all of those manuscripts, and only in those manuscripts. According to this theory, only the variants aceg meet this test, of being present in all three of BCD, only in BCD, and therefore present in A. However, the introduction of the qualification, that a manuscript may declare itself as a

descendant of a particular ancestor by having only a significant number of the variants introduced in that ancestor, permits us to see the variants bdf as also likely to have been present in A.

This modification has far-reaching effects. Firstly, it is much closer to what common sense tells us is really likely to happen in the copying of a manuscript tradition. If manuscript A introduces (say) 70 readings, and then is copied in turn by the scribes of BCD, then simple human inconsistency will see to it that no one of these three will copy every one of the 70 readings, and indeed each of the three will copy a slightly different selection of the 70. One might see 60 of the readings in B, 45 in C, 55 in D, and so on. In addition it is likely that in a real tradition some of these variants might (by simple accident, or contamination) also appear in other manuscripts not descended from A. Presume, in this case, that we had only the manuscripts BCDEF with variants distributed as below:

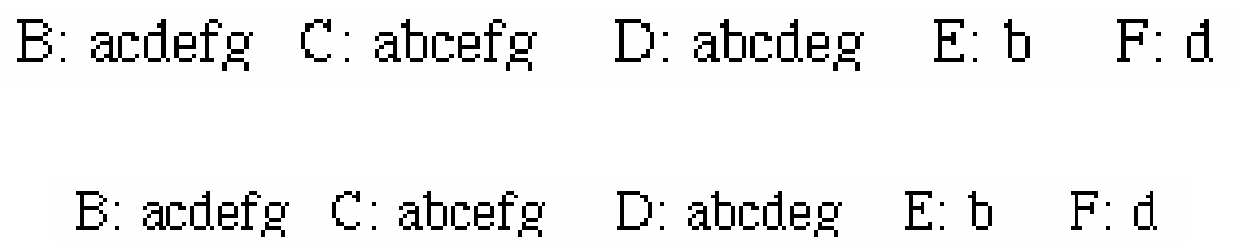


Figure 3: Hypothetical variant reading distribution

By ordinary processes of deduction, we might deduce that BCD had a common ancestor, and that this common ancestor introduced all the readings abcdefg (not just aceg). We might further presume that the presence of the unoriginal readings b and d in EF is likely to be the result of simple chance only. This would agree precisely with the modified stemmatic theory presented above, and give exactly the stemma given in Figure 2. Indeed, manuscript scholars have usually taken just such a pragmatic approach. In fact if not in theory, they have accepted that a particular manuscript will show itself as a member of a given group by the presence of a significant number of variants from that group.

The second effect of this modification is that it makes identification of the fundamental witness groups in large traditions considerably more difficult.

In this case, it seems reasonable to presume that because all three of these manuscripts BCD have the four readings aceg then they are likely to share a common ancestor which introduced these four readings. One could identify this fundamental group just on the basis of these four readings aceg. One might then note the presence of two of the three readings bdf in each of BCD and deduce that these three readings were likely also to have been present in this common ancestor. With just three manuscripts, in a small tradition, it is likely indeed that there will be just such a core of introduced readings present in all three, and only in those three.

But, the more manuscripts we have, the more likely it is that for any given reading present in the common ancestor, there will be at least one manuscript which fails to include that reading. This is especially so when there are several intervening copies between the manuscript and the hypearchetype, with each copyist removing some of the introduced readings, either by introduction of a different error or (as is commonly the case) by re-introducing the original reading. In the analysis of The Wife of Bath's Prologue (1997) I refer to this phenomenon, where successive copies retain fewer and fewer of the introduced variants,'variant drift.'[2] On the most basic level, it is very likely that in a large number of manuscripts, the lacunae in those manuscripts will overlap to such an extent that indeed, it will be impossible for all the manuscripts of that group to share introduced readings, simply because for every line or text segment, there is at least one manuscript which is 'out' and does not have that line or segment at all. Of course, in such a large tradition even if it happens that every one of the descendants has a particular reading introduced by the hypearchetype, it is very probable that some other unrelated manuscript will have this reading. This will destroy the ideal situation of the reading occurring in every manuscript of this group, and only in every manuscript of this group.

It is conceivable then that for a large group of (say) ten or more manuscripts descended from a common ancestor, that there may be very few or even no readings introduced in that common ancestor which actually survive in every one of those copies. This was exactly the case in our analysis of the fundamental group of witnesses named group cd (following Manly and Rickert) for The Wife of Bath's Prologue. We conclude that there are about twenty manuscripts in this group. The uncertainty in the number is due in part to shift of exemplar, whereby

manuscripts leave or join the group in the course of the text. It is also due in part to 'variant drift' where the number of cd variants in particular manuscripts may have been reduced to the point where it is not certain that a particular manuscript was or was not a member of the group. The cd variants are an extremely distinctive and numerous set of variants. On our analysis, the common ancestor of cd introduced some 250 readings into the text, about one every three lines. Yet, there is actually not a single reading in the Wife of Bath's Prologue which occurs in every cd witness, and only in the cd witnesses. Commonly, the cd readings are found in between 10 and 25 witnesses, with manuscripts leaving and joining the group at individual variants with no discernible pattern.

In theory (and indeed in fact, if our analysis of the Wife of Bath's Prologue is correct), for a large group of manuscripts in a large tradition one must expect that there may not be a single case where the classic stemmatic requirement, that the introduced reading should be in every witness of this group and in no other witness, is satisfied. But how then is the scholar to identify this group? Further, for any one manuscript, what might constitute a significant number of variants, sufficient to warrant that manuscript being a member of a particular group? In our hypothetical case of the seven variants abcdefg one might indeed agree that the presence of a different six of these in each of BCD suggests that there was indeed a hypearchetype A and that each of BCD is descended from A. Conversely, the presence of only one of these seven in each of EF is likely to signify nothing more than chance. But what if we had a further manuscript, G, which had (say) three of these seven readings? at some point, one would have to say that the evidence simply is not decisive, either way, and that it is just not possible to determine for a particular witness whether it belongs to a group or not. The matter will become yet more complex if one is unsure whether given readings are original to the ancestor of the whole tradition or introduced below this archetype. In the case of readings bd, present in the 'outside' manuscripts EF, for example, if one felt the readings might be original to the whole tradition then their presence in EF — and in BCD — would mean nothing at all.

In a large and complex manuscript tradition one will find both instances where there seems clear evidence of the relation of particular manuscripts (as in BCD in this instance), and instances where the evidence permits no certain conclusion (as in the putative G). One needs to be able be see the pattern, where there is one; and also able not to see a pattern, where there

is not one. Human beings are very good at seeing patterns in language, but not so good in seeing patterns in numbers. It is likely that, left to ourselves and when grappling with large numbers of variants in large numbers of witnesses, scholars will both not see patterns though they are there, and also think they see patterns when they are not there. In the case of the *Piers Plowman* tradition, Kane refused to accept that any group of manuscripts existed unless there were variants present in all the manuscripts of the group, and only in the manuscripts of the group. In the circumstances of a real tradition, as argued here, this will happen very rarely, and so this rigid approach must often fail to see groups which actually are related. On the other hand, Manly and Rickert appear so eager to construct larger and larger manuscript groupings, and thus to simplify their stemma, that the evidence for their grouping becomes extremely problematic.

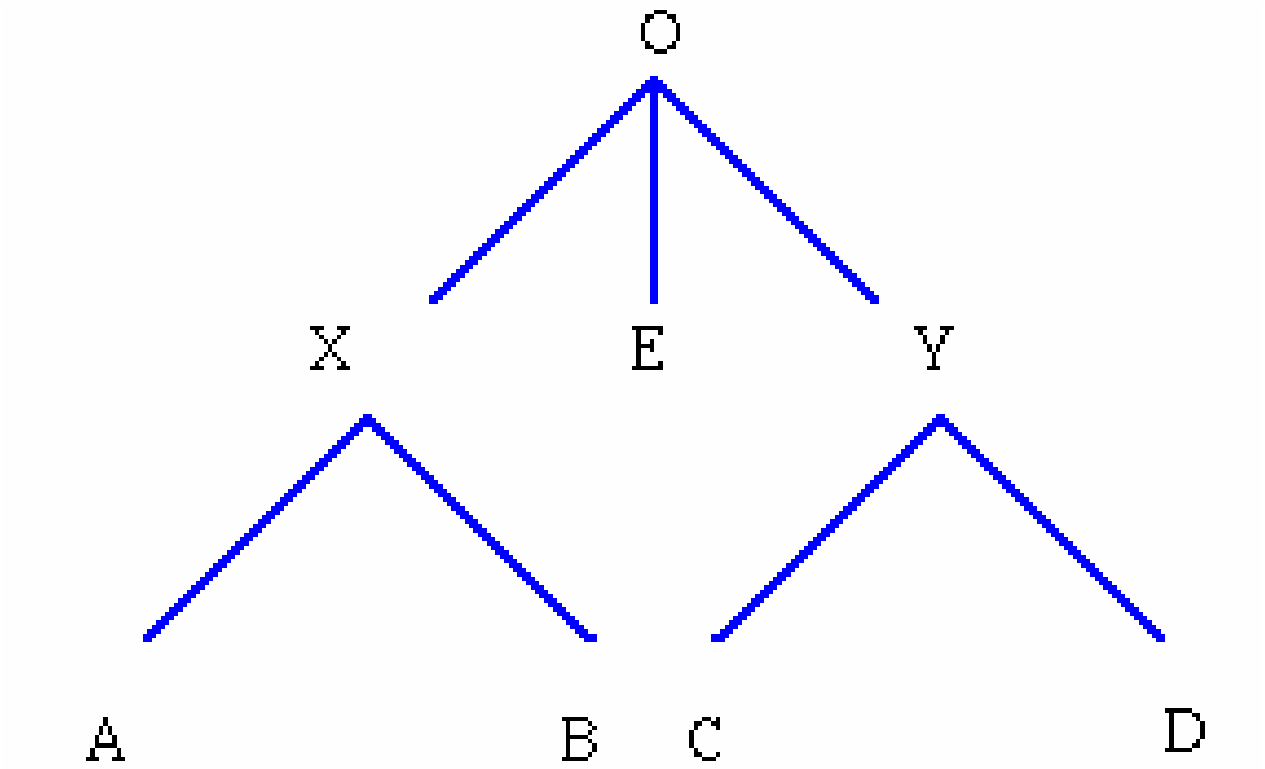
Computers are notoriously poor at seeing patterns in language, but excellent at seeing patterns in numbers. Perhaps we could use computers in the analysis of manuscript traditions to help us find what otherwise we could not.

2.3 Evolutionary biology: unrooted trees

Over the last decades, workers in evolutionary biology have been faced with exactly the same problems in their attempts to determine the family relationships of apparently related organisms from study of the characteristics they share and do not share. As in stemmatics, so in evolutionary biology there is the problem of determining what is original and what is introduced; there is the problem of the same characteristics appearing in unrelated species apparently by chance or by hybridization, analogous to accidental agreement and contamination; there is the problem of identifying groupings on shifting sands of agreement and disagreement, where some individual organisms fail to preserve the characteristics introduced by an common ancestor or introduce characteristics from outside the common ancestor.

Up to around 1960, these problems led many evolutionary biologists to abandon genetic hypotheses, just as textual scholars were abandoning stemmatics.[3] However, from 1960 on two factors have transformed

evolutionary biology, and brought about a renaissance of what is called phylogenetics: the reconstruction of the history of organisms in terms of their family groupings. The first factor is the theoretical work of Willi Hennig and later thinkers. Hennig and his followers addressed the issues of original and introduced characteristics, and of the failure of organisms to preserve only the characteristics introduced by the common ancestor. Their response to the first problem, which in stemmatic terms is the problem of how does one determine what is 'original', what 'introduced', is this: you do not have to determine it. Suppose one has organisms ABCDE descended from an original O through two distinct ancestors, X and Y, as follows:



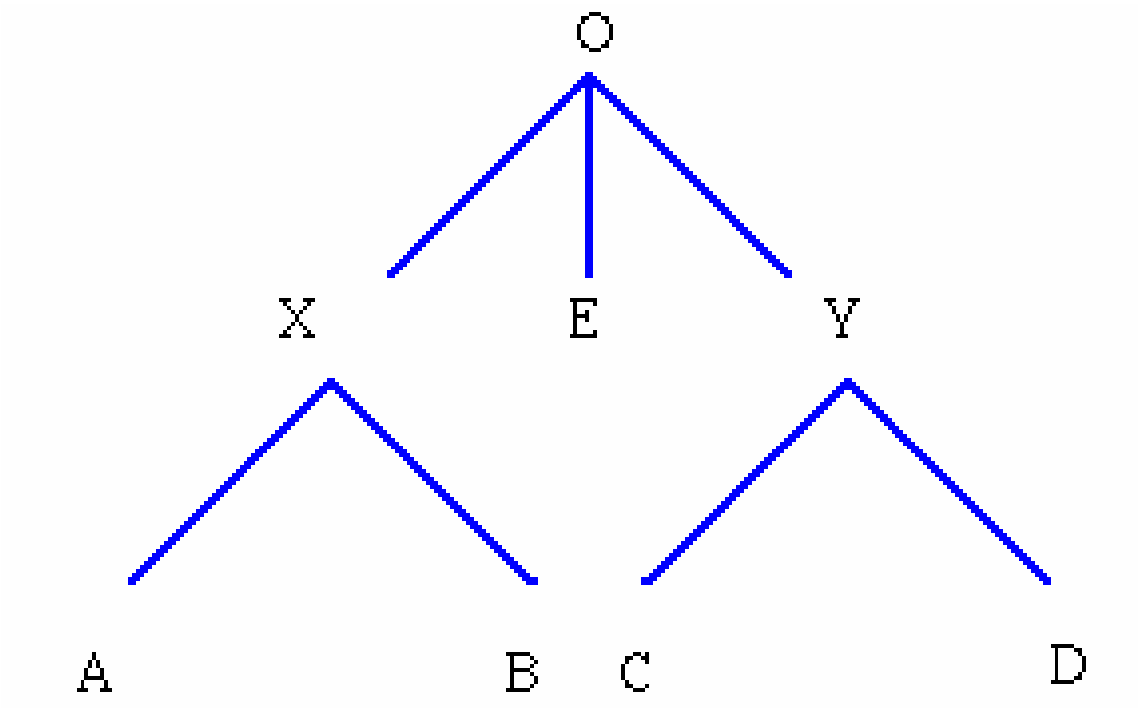
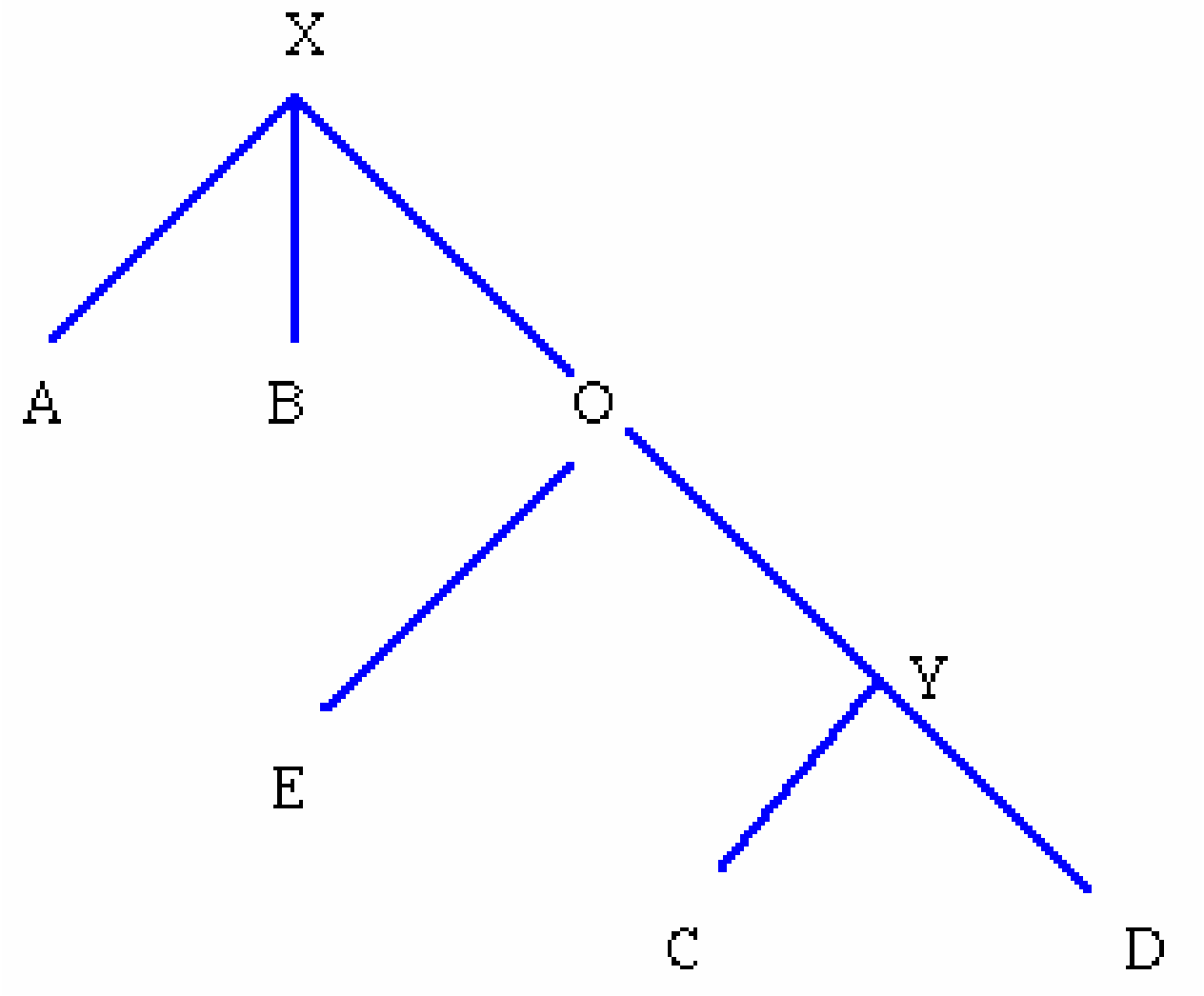


Figure 4: a tree rooted at O

Suppose that instead of rooting the tree at O, we place the root of the tree at X. The tree now appears as follows:


X
A
B
O
E
Y
C
D

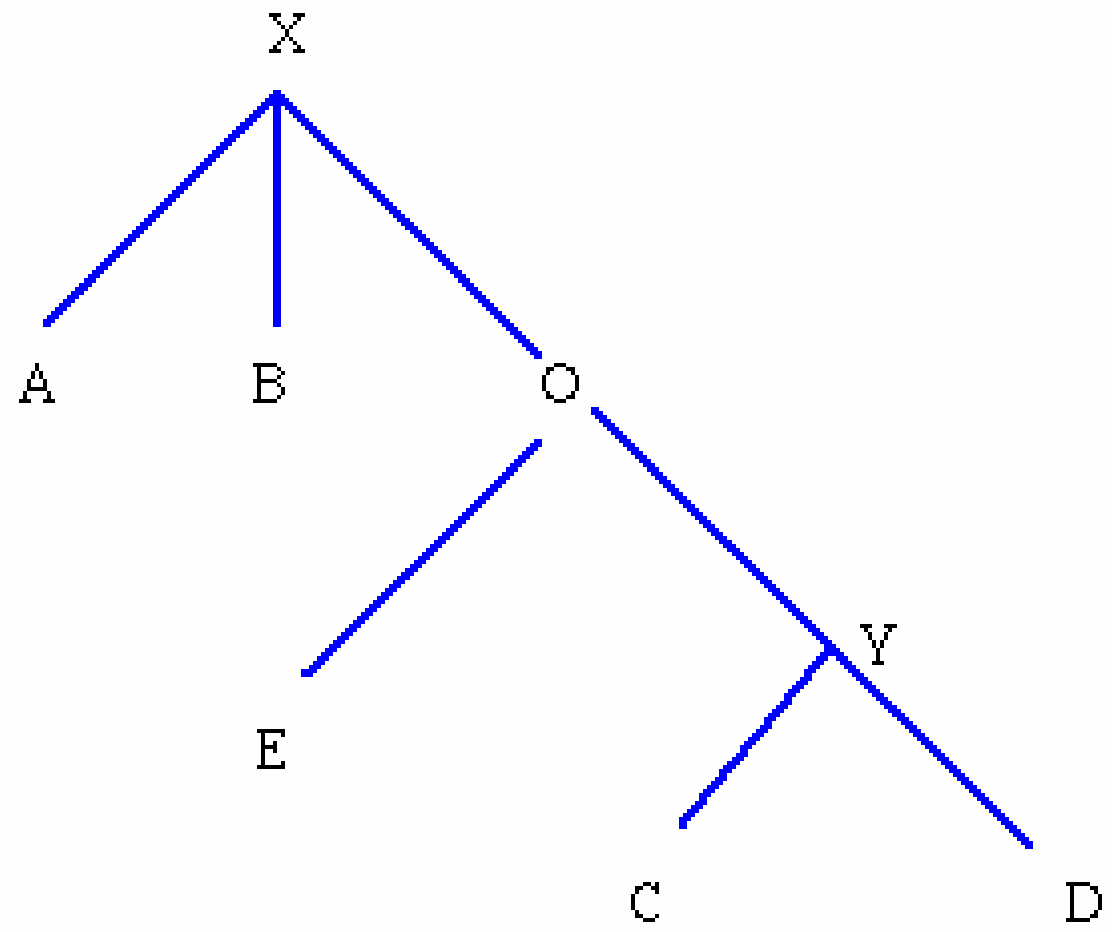


Figure 5: the same tree as Figure 4, rooted at X

This second tree, though differently rooted, has exactly the same internal relationships between its parts as the first tree. That is: E is descended directly from O; AB is separated from O by a common ancestor X; CD is separated from O by a common ancestor Y. Indeed, no matter how you root the tree, the different parts have exactly the same relationship with each other.

Hence, evolutionary biologists arrived at the notion of an 'unrooted' or ' unoriented' tree. If one has a tool which can recreate the relationships of objects one can postpone the identification of the ancestor to a later point in the analysis. One can create a tree showing the relationships, as in

figures 4 and 5, examine the groupings which emerge from the tree (here: AB/E/CD), determine which of these groupings appears most likely to be nearest the original, and so root the tree accordingly. Thus in this case: after making the tree, we could then examine the variants present in each of AB, E, and CD. We would conclude that those found only in the pairs AB and CD are unoriginal in character, while those found only in E are original. Therefore, the correct rooting of the tree is that given in figure 4.

This has far-reaching consequences. Because identification of the original can safely be postponed till the last stage of the analysis, it means that one can concentrate just on identification of the family groupings, as a necessary first stage. Thus, one only (only!) has to identify these groupings. At just the period in the 1960s when Hennig and others were rethinking this theory, vast amounts of additional data concerning agreements and disagreements among organisms were becoming available, through DNA sequencing. Evolutionary biologists found themselves faced with exactly the same problem confronting stemmaticists: how does one, from information about tens of thousands of agreements in hundreds of organisms, distinguish real patterns showing real relationship, and so determine the relationships themselves?

The response of evolutionary biologists was the same as that suggested at the end of the last section. The problem can be put in terms of finding patterns in numbers: what computers are good at, and human beings are not. Therefore, one should develop computer programs capable of finding these patterns.

2.4 SplitsTree and PAUP

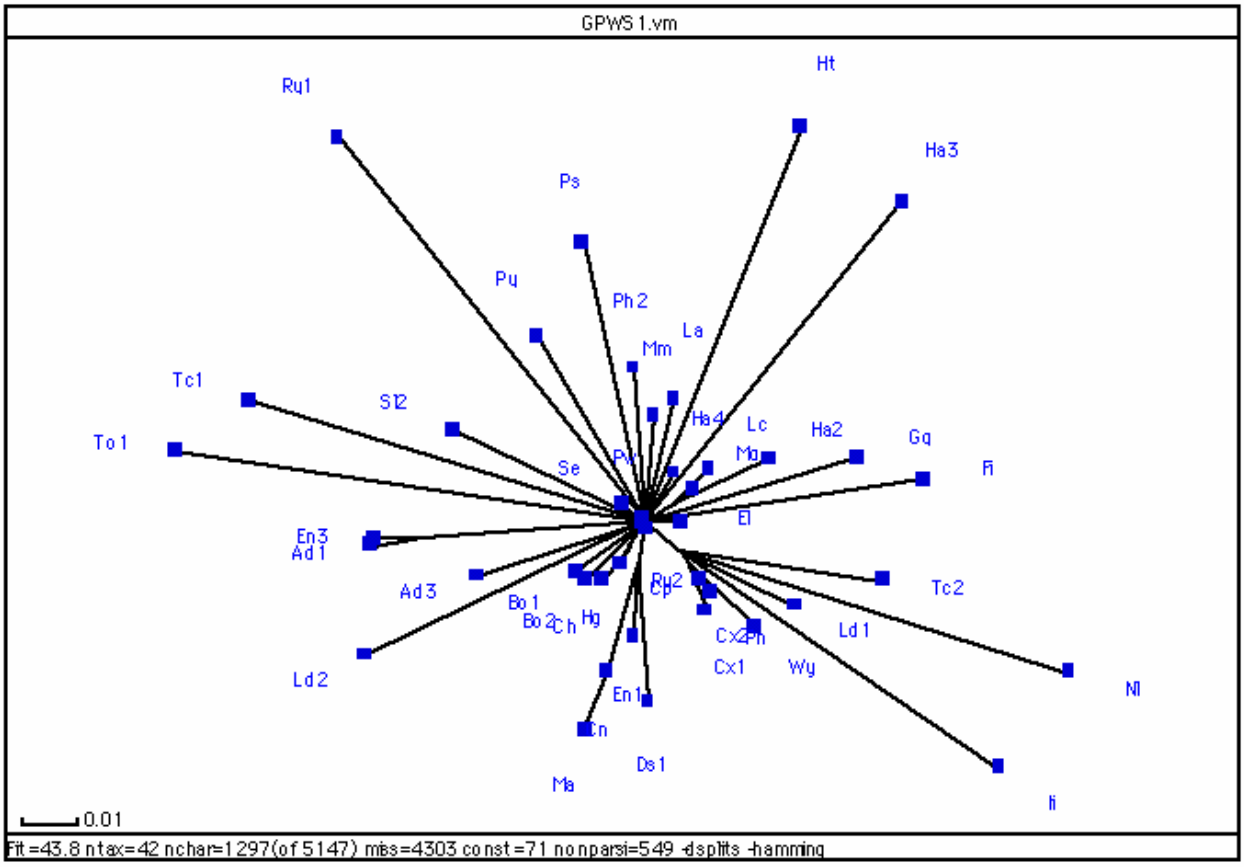
Accordingly, immense effort over the last decades has been devoted to the writing and testing of computer programs which carry out phylogenetic (or 'cladistic' — from the Greek clados, 'branch') analyis: the making of trees showing the evolutionary relationships among organisms. Some idea of the scope of this effort can be gained from two websites: the website of the Willi Hennig society, at http://www.nhm.ac.uk/hosted_sites/hennig/hennig2.html, and Joe Felsenstein's website for his PHYLIP suite of programs, at http://evolution.genetics.washington.edu/phylip.html. This widespread

activity has led to the development of many different tools, using distinct methods and focussing on distinct problems. One can use different tools, and different methods, to cross-check one's results: if contrasting methods suggest the same relationship, then one may be more confident about that relationship. Further, the circumstances of individual situations in evolutionary biology are every bit as diverse as are those in textual traditions. No two cases are the same and this has fostered the creation of many different tools, for many different situations. Thus, it is the more likely that one can find just the right tool for the particular case.

The *Canterbury Tales* Project has identified two tools developed in evolutionary biology as particularly suitable for our purposes. The first of these tools is the program SplitsTree, developed by Daniel Huson of the University of Bielefeld, Germany, on the basis of suggestions by A-J. Bandelt and A. W. M Dress. I am grateful to Chris Howe of the Department of Biochemistry, University of Cambridge, and to Linne Mooney of the University of Maine for introducing me to this program. SplitsTree differs significantly from what might be termed the classic evolutionary biology technique of cladistic analysis as exemplified by programs like PAUP (described further below). SplitsTree is based on the premise that in any mass of data relating to a real evolutionary history, there will be some data which supports a tree-like relationship for at least some of the possible related objects. However, there will also be data which does not support a tree-like relationship: in a manuscript tradition, for example, such a situation would be data arising from contamination or shift of exemplar. In Huson's words, one can use SplitsTree to indicate ' how tree-like given data is' (see the document 'splitstree.doc', included on this CD-ROM, in the directory 'docs'). The means used by SplitsTree are complex, and a full account lies beyond the scope of this introduction. In summary: SplitsTree uses sophisticated graphing methods ('canonical decomposition') to transform the account of the relations between objects at every point where there is variation into a single 'splits graph'. In lay terms, one could think of this as creating a tree at every point where there is variation, and then laying all these on top of one another to create a single tree for the whole tradition. A detailed account of how SplitsTree works is contained in the file 'splitstree.doc', incuded on this CD-ROM, in the directory 'docs'. This file presents, in a Microsoft Word document, Daniel Huson's own account of SplitsTree.

This concept, that within a range of data there will be some areas which

are ‘treelike’, others which are not, is very close to the view of manuscript relations presented in section 2.1 above. One might use SplitsTree over different ranges of data, in different manuscripts, to determine whether that range of agreements and disagreements in those particular manuscripts is ‘ treelike’, and so can be explained in terms of the family relationships among those manuscripts. Many datafiles are included on this CD-ROM which you can use with the SplitsTree program, included on this CD-ROM. You may click here to see the program at work. SplitsTree will start, and it will load a file named ‘GPWS1.vm’ (you can ignore the warning message). This file contains data on the agreements and disagreements for 42 witnesses to lines 1-250 of The General Prologue: all the manuscripts which contain reasonably complete texts of these first 250 lines. SplitsTree will process this, and make a tree which should appear as follows:



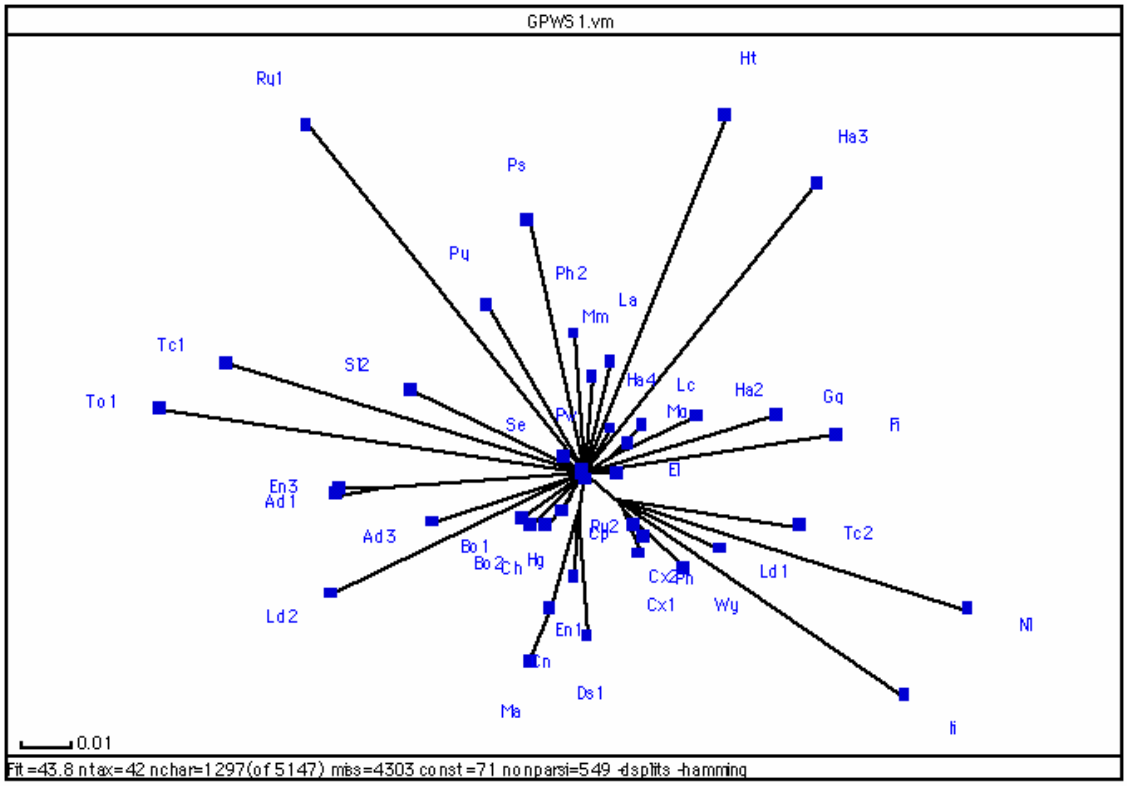


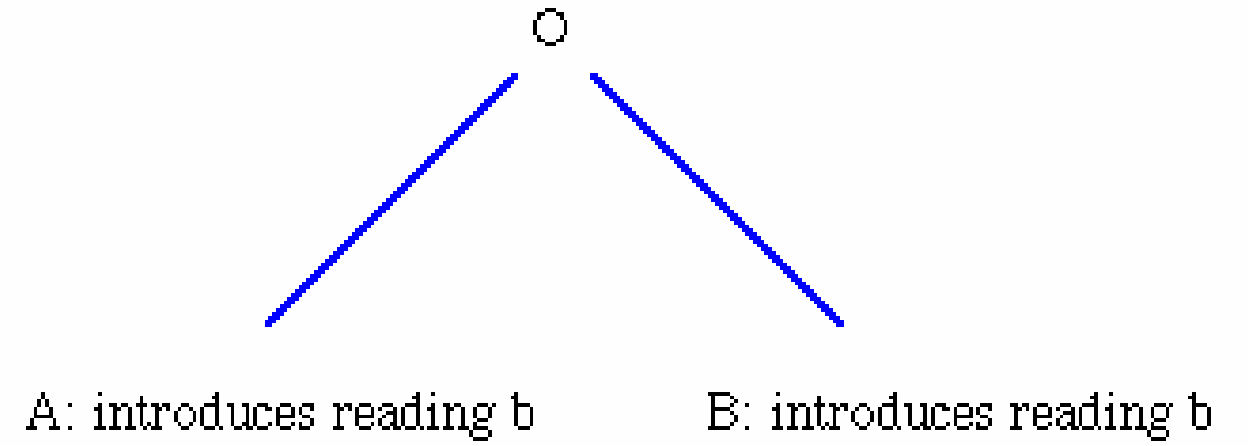
Figure 6: A SplitsTree graph for the file GPWS1.vm. Manuscripts excluded: BASE Ad^4 Bw Dd Dl Do Gl Ln Ne Ra^2 Ra^3 Sl^1

Note that in this graph, the majority of the manuscripts of The General Prologue simply radiate straight out of from a single central point. The length of each line from the centre represents 'distance' from that point: in gross terms, one could think of this as the comparative number of variants. A short line implies the manuscript has few variants from the notional centre; a longer line, that it has more.

For these manuscripts, for this data, SplitsTree suggests that they do not appear to be related in a 'tree-like' manner. However, SplitsTree does suggest that for some seventeen manuscripts, for this data, there are ' tree-like' relationships. These are the four manuscripts Cn/Ds^1/En^1/Ma, in the bottom centre of the graph; the manuscripts Lc/Mg/Ha^2, in the centre right; the group Cx^1/Cx^2/Pn/Wy/Ii/Ld^1/Nl/Tc^2 in the bottom right; and the pair En^3/Ad^1 on the centre left. All these small groups appear to radiate

from a single point separate from the centre; in some cases, radiation from separate points may suggest further archetypes within the group (as in the grouping $Cx^1/Cx^2/Pn/Wy/Ld^1/Nl/Tc^2$).

If SplitsTree finds that the data is indeed treelike, it will itself suggest a tree of relationships, as it does in the instance here given. However, SplitsTree is not designed to find the very best possible tree: in phylogenetic terms, the tree which offers the most economical ('parsimonious') tree of descent explaining all the agreements and disagreements between the objects under study. There are many evolutionary biology programs which do aim to provide just such a tree. The Project has used David Swofford's Phylogenetic Analysis Using Parsimony (PAUP) for this purpose. The thinking behind PAUP is quite different from that behind SplitsTree. PAUP (like other cladistic programs) presumes that the agreements and disagreements among a population of objects are to be explained by the process of 'descent with modification': by the location of the objects in a ' tree of descent'. It then seeks to find this tree of descent by hypothesizing many different possible trees, comparing those trees with the data, and calculating for each tree how well it 'explains' the data on agreements and disagreements. The best tree will be the tree which is the most ' parsimonious': in cladistic terms, this is the tree where agreements and disagreements are explained most economically by the fewest possible number of 'changes in state'. One might think of this as the tree where manuscript variants are explained most often by occurring in common ancestors, and then being inherited in their descendants, rather than occurring independently, over and over again, at separate points in the tradition. For instance: presume we have two manuscripts A and B containing the unoriginal reading b. We could explain this as follows:



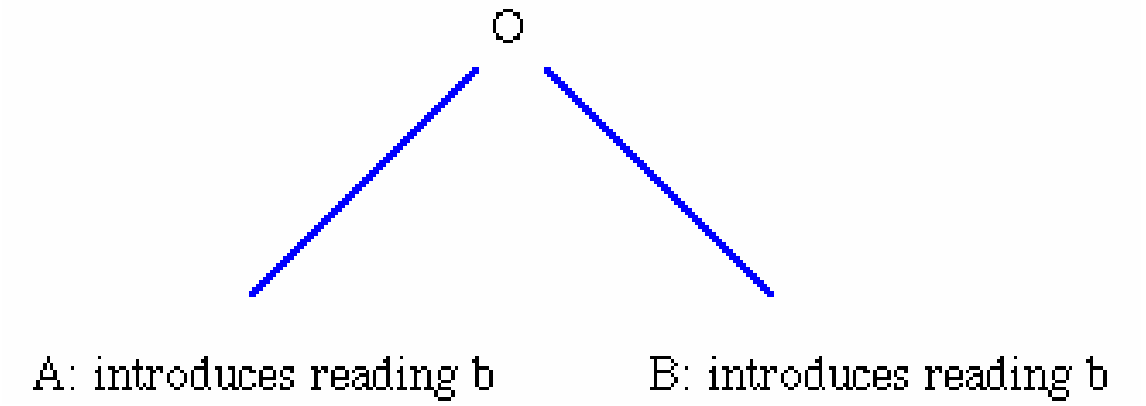
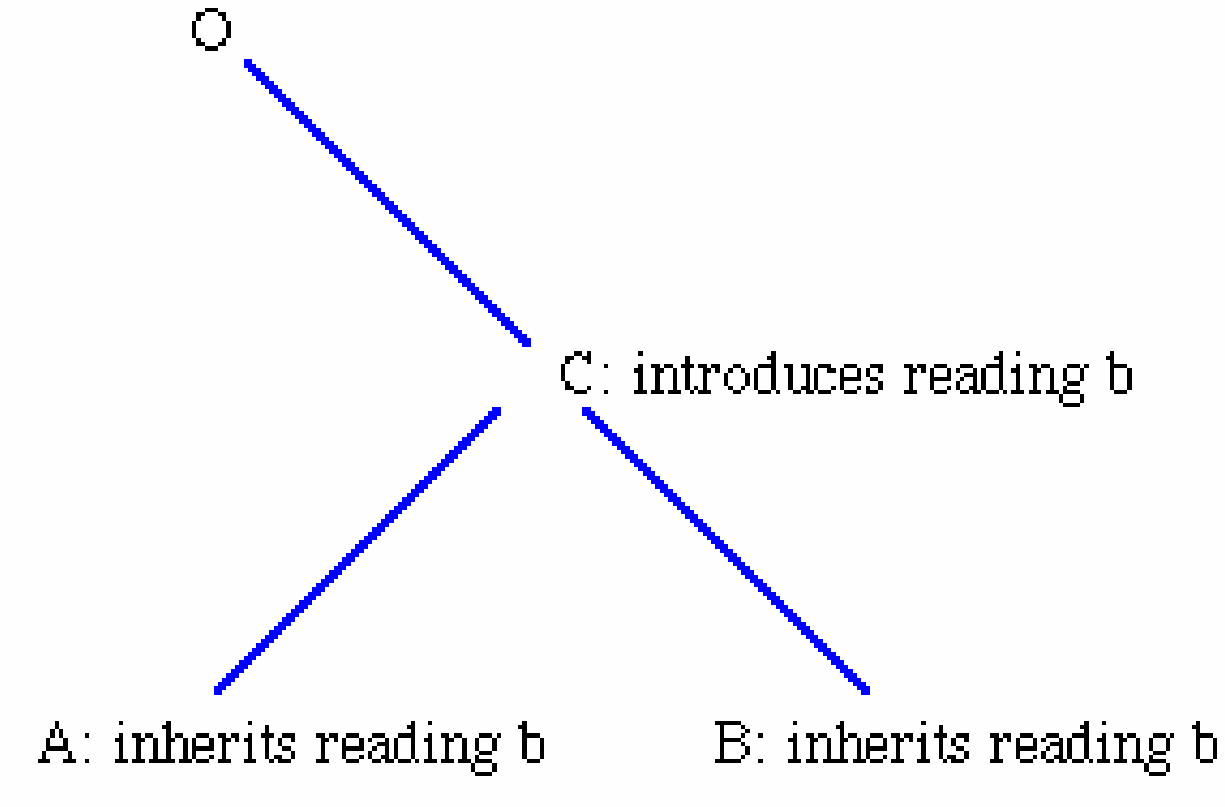


Figure 7: an ‘unparsimonious’ tree

In cladistic terms, this is seen as two ‘changes of state’, from ‘not-b’ to ‘b’: the introduction of reading b into A is one ‘change of state’, and the introduction of reading b into B is the second change of state.

Given this data, a cladistic program would actually presume the existence of a subancestor, C, which introduced reading b:


O
C: introduces reading b
A: inherits reading b
B: inherits reading b

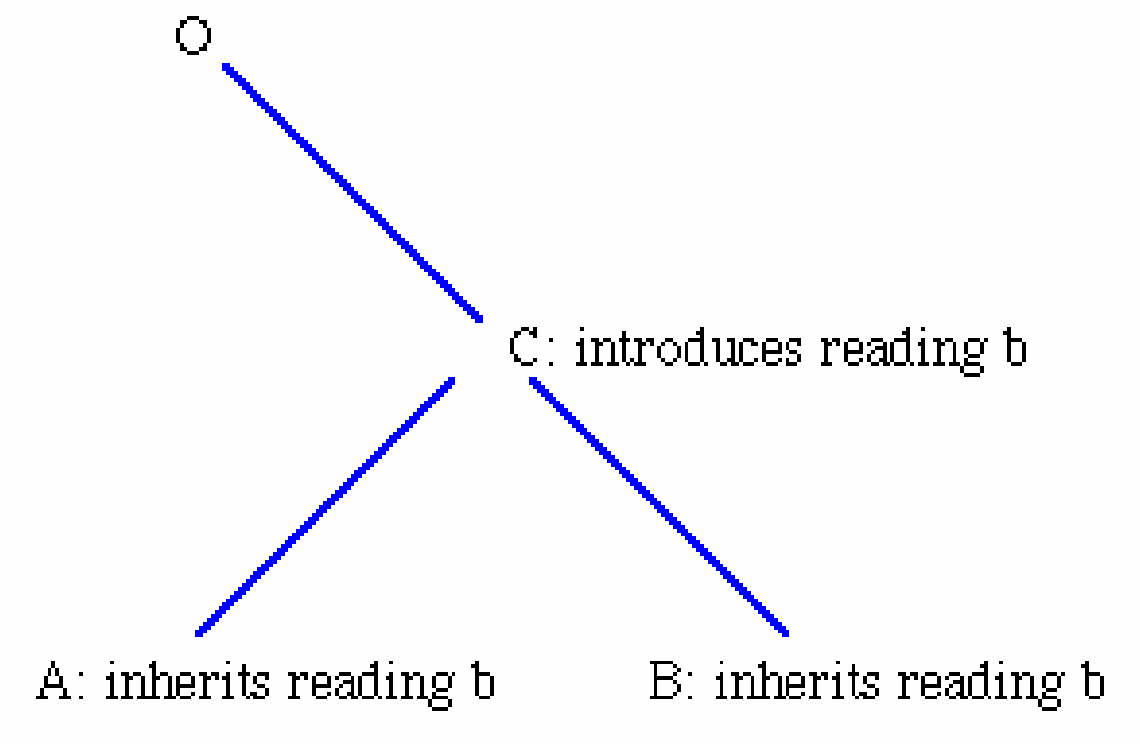


Figure 8: a more parsimonious tree

This reading is inherited into the descendants of C in A and B. In this second tree, there is only one 'change of state', occurring at C, not two changes of state, as in the tree in Figure 7. A cladistic program will see this second tree as more 'parsimonious' than the first tree.

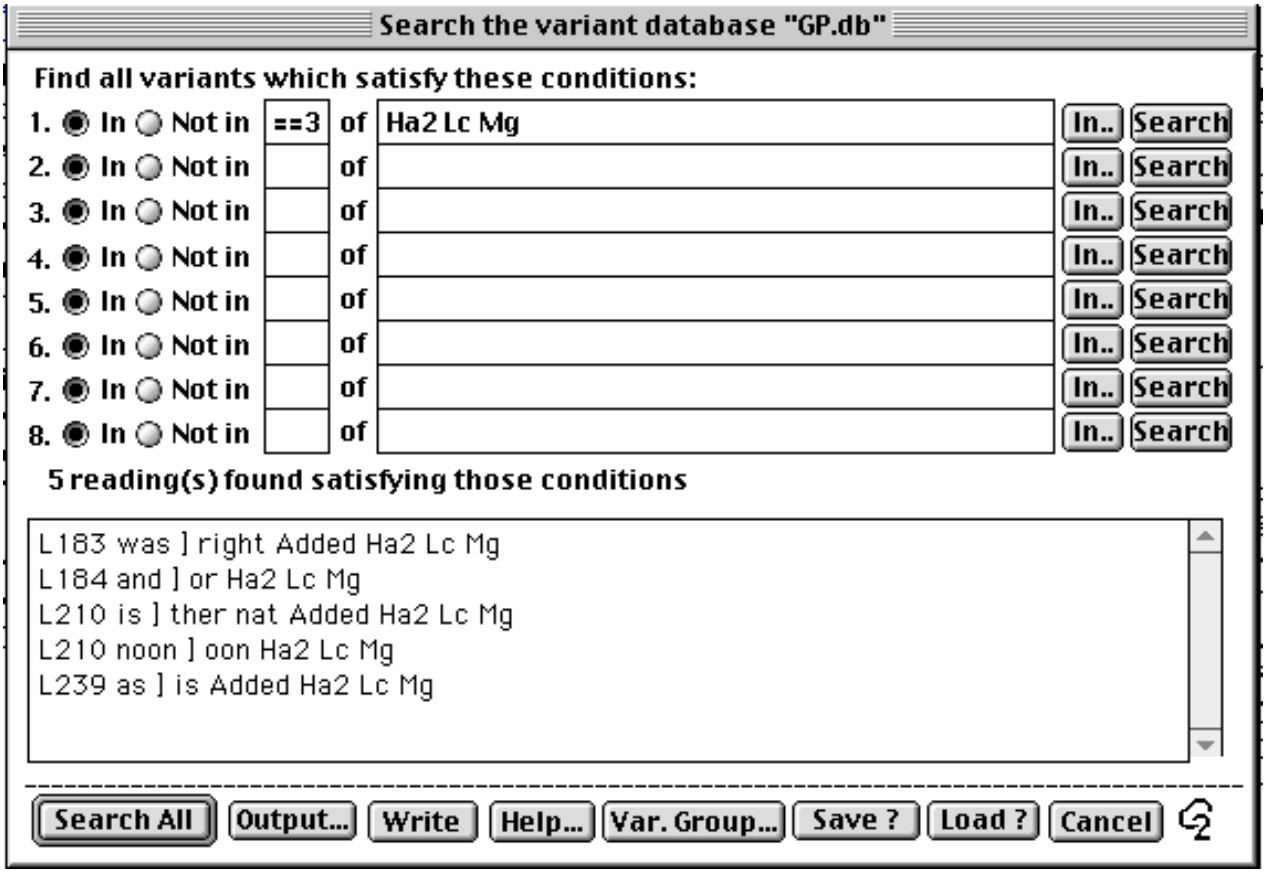
Where the data is indeed 'tree-like', this method works very well indeed. However, such cladistic programs can produce rather misleading results when the data is not tree-like, as they will find trees whether there are trees to be found or not. Accordingly, the Project uses PAUP only on sections of data in manuscripts where SplitsTree suggests that the cladistic method might be useful. PAUP is not provided on this CD-ROM. However, data files in PAUP format are provided in the directory 'docs.' There is a PAUP file, with the suffix .nex, corresponding to each SplitsTree file, with the suffix .vm, contained in this directory 'docs.' If you have PAUP, you can run PAUP over these files and study the results for yourself.

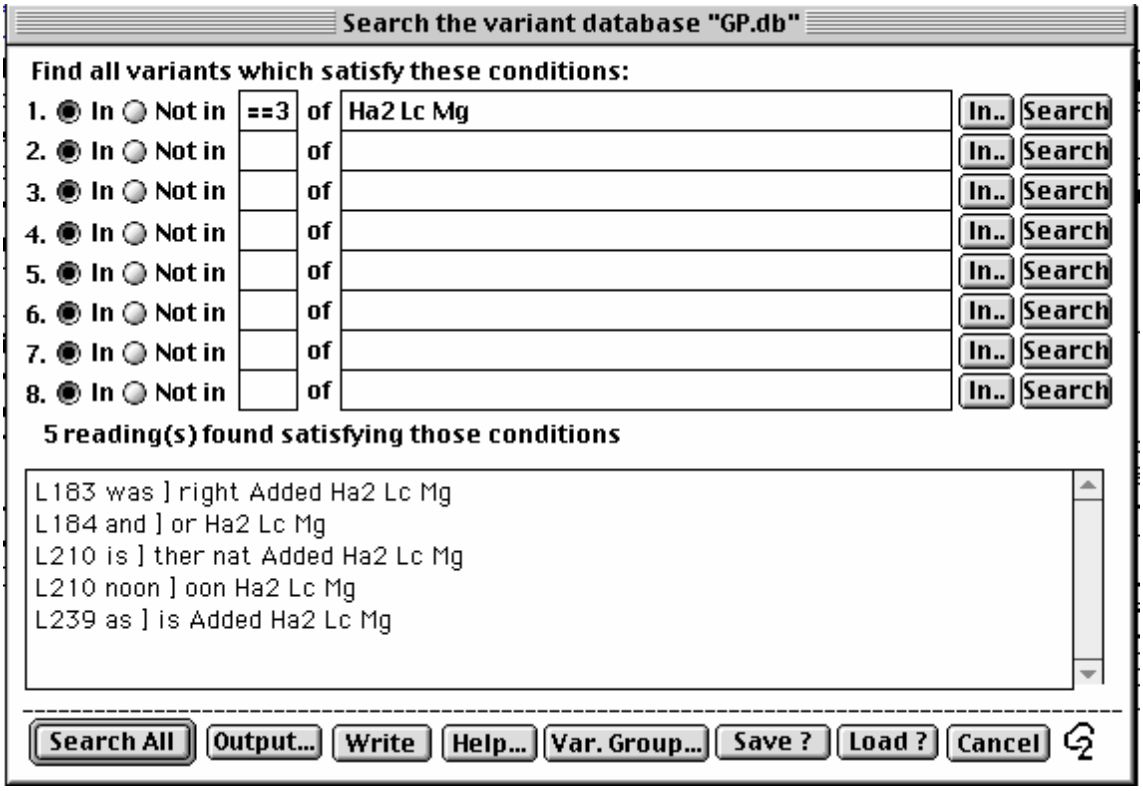
2.5 The Collate variant database: fundamental groups

From this account, it is clear that the evolutionary biology methods do not, of themselves, provide a full account of the textual history of a real manuscript tradition. In Figure 6, SplitsTree is able to group only some seventeen of forty-two manuscripts. While this figure can be increased somewhat (see 3.1.1 below), there will remain a large number of manuscripts which, for one reason or another, evolutionary biology methods are not able to place in a genetic relationship.

Further, scholars are naturally suspicious of 'black box' methods, where data goes in one end and a manuscript tree comes out the other end. We need to know just what it is in the data which causes SplitsTree to suggest (for example) that the three manuscripts Lc/Mg/Ha^2 form a single group. What are the variants in those three manuscripts which appear to have been introduced by a common ancestor? are we persuaded that these variants really are unoriginal and significant, and that their presence in these three manuscripts is truly the result of their having been introduced by a common ancestor?

To answer these questions, the Project uses an additional tool, devised specifically to explore the manuscript tradition in ways which the evolutionary biology programs do not. This tool is the variant database program VBase, developed as part of the *Collate* set of software tools. VBase is designed to identify the variants which might be evidence of a particular manuscript grouping. You can see this program at work by clicking here.


Search the variant database "GP.db"
Find all variants which satisfy these conditions:
1. In Not in ==3 of Ha2 Lc Mg In.. Search
2. In Not in of In.. Search
3. In Not in of In.. Search
4. In Not in of In.. Search
5. In Not in of In.. Search
6. In Not in of In.. Search
7. In Not in of In.. Search
8. In Not in of In.. Search
5 reading(s) found satisfying those conditions
L183 was] right Added Ha2 Lc Mg
L184 and] or Ha2 Lc Mg
L210 is] ther nat Added Ha2 Lc Mg
L210 noon] oon Ha2 Lc Mg
L239 as] is Added Ha2 Lc Mg
Search All
Output...
Write
Help...
Var. Group...
Save ?
Load ?
Cancel

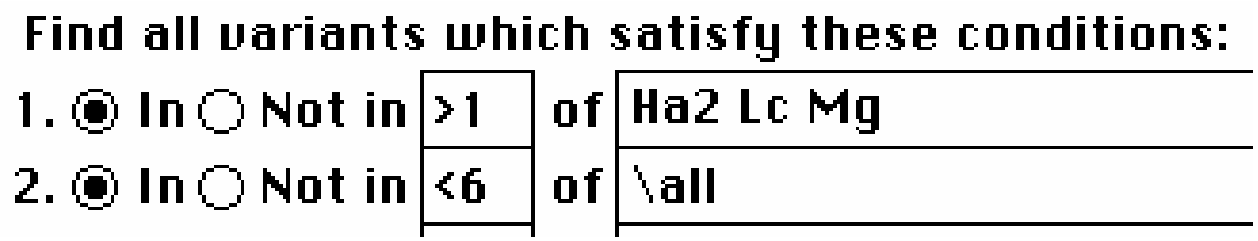


In this case, we have used the database to identify the variants found in the three manuscripts Ha^2 Lc Mg, and only in those three ('==3'). By using the **In..** button we have further specified that the search should look only in the first 250 lines of the poem. The search has found five readings in the first 250 lines which satisfy this condition, and appear to be evidence that these three have a common ancestor.

The database is also designed to meet the case described in 2.2 above, where we remark that many variants introduced by the common ancestor of a group such as Ha^2 Lc Mg will not be found in only that three, and in all that three. One of the three might, by simple chance, not have a particular variant; again by simple chance (or perhaps contamination) some of the variants will occur in other manuscripts. Thus, we might expect to find some of the variants introduced by the common ancestor in two of the three, and in perhaps three other manuscripts. You can search for the variants satisfying this more complex search as follows:

Replace the '==3' in the first box in the first row by '>1'. This will force the database to find all cases where the variants are in any two of Ha2 Lc Mg as well as all three of these

- Type '<6' in the first box of the second row, and '\all' in the second box of this row. This will force the database to find all cases where the variants are in fewer than six manuscripts in total. These two rows should appear as follows:



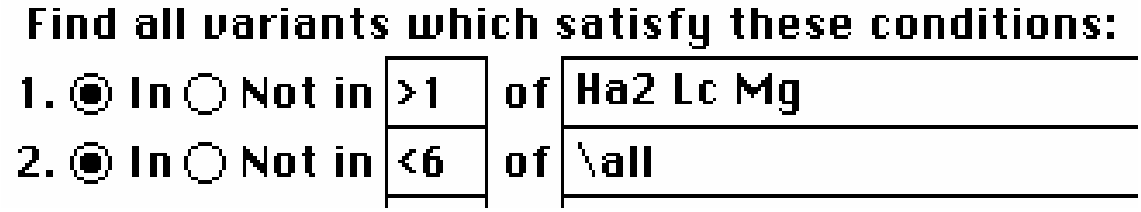


Figure 9: a database search

- Press the **Search All** button at the base of the screen. (Alternatively, you can do this search just by clicking here)

  The database should now send you the message that it has now found 31 readings satisfying these conditions: readings in at least two of Ha2 Lc Mg and in five or fewer manuscripts. You can scroll through the box at the base to see what these variants are. You will see many of these variants are actually variants on punctuation (for example, in lines 79, 108, 130), as the manuscripts Lc Mg share a distinctive system of punctuation.[4] Clearly, punctuation variants are likely to be of little value in determining genetic relationships, but are likely to result just from different scribal habits. We can have the variant database filter these out, by adding the condition 'with !punct' to the first row of the query, as follows:

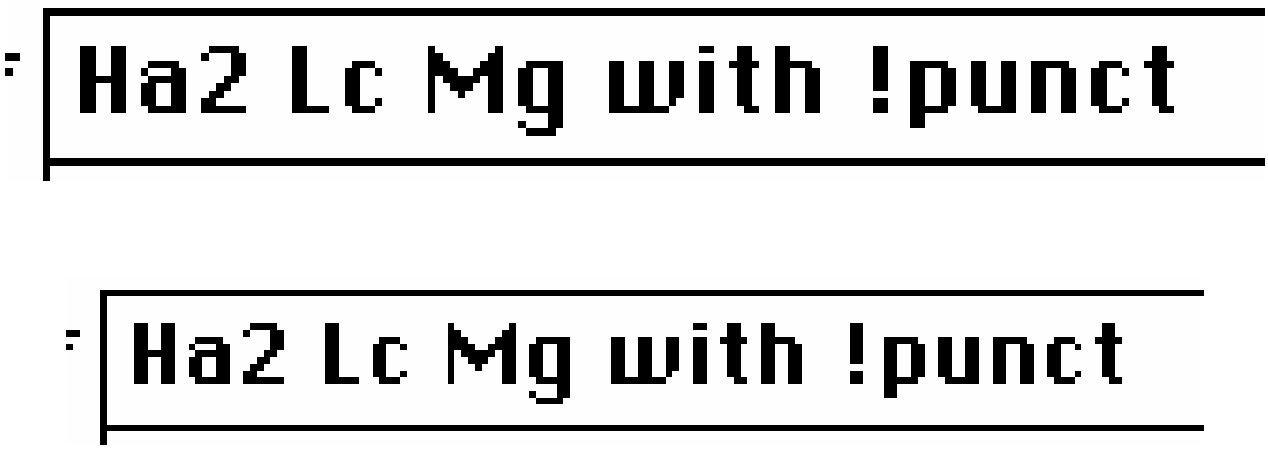


Figure 10: filtering out punctuation variants

- Press **Search All** again: this time, the query returns 17 readings. (again, you can do this search by clicking here)

Through queries such as this, the Collate variant database is used to identify the readings characteristic of the fundamental groups. Thus, after SplitsTree and PAUP have identified the manuscripts which appear to belong to a fundamental group (e.g. Ha^2 Lc Mg), we then use the database as above, to identify the group of variants characteristically found only in the witnesses of that group and therefore likely to have been introduced by the common ancestor of that group. In the example above, we would identify the seventeen variants found in the last search as characteristic of the group Ha^2 Lc Mg in lines 1-250 of The General Prologue.

2.6 Variant group profiles

Once these fundamental groups are identified, we are able to use this information to explore the whole tradition: to allocate further manuscripts to the fundamental groups; to determine manuscripts which use more than one exemplar, either by shift of exemplar or by contamination; to judge the exact relationship of particular manuscripts with each other.

We use the variant database to allocate manuscripts to the fundamental groups and to manage the problems of manuscripts using multiple

groups and to manage the problems of manuscripts using multiple exemplars as follows. Presume that we identify 150 variants as being characteristic of manuscripts which belong to group A, by using the variant database as shown above. These 150 variants will then likely have been introduced in the common exemplar of the manuscripts of group A, and a significant number of these A variants will then have been inherited by the manuscripts descended from this common exemplar (A itself). Presume that we have a separate group of manuscripts, group C (descended from C), characterized by a separate 200 variants: we would expect that manuscripts belonging to this group would have a significant number of these. We want to identify in the tradition the manuscripts which are:

1. descended from A
2. descended from C
3. descended from A up to line 500 and from C after that
4. descended from A, but also using readings from C
5. descended from neither A nor C

We use the variant database to make this identification, through what we call 'variant group profiles'. Each 'variant group profile' consists of a statement of how many variants from each fundamental group are present in a particular manuscript. Further, we break the fundamental groups into different sections for different parts of the text: in this example, presume there are 90 A variants and 80 C variants up to line 500, and 110 A variants and 70 C variants after line 500.

Thus, the variant group profiles for the five manuscripts **v w x y z** might appear as follows (numbers in brackets are the numbers of variants in that group available in that sequence of lines):

v

A	variants to line 500 (90)	70
A	variants after line 500 (110)	80
C	variants to line 500 (80)	6
C	variants after line 500 (70)	7

w

A	variants to line 500 (90)	5

		8
C	variants to line 500 (80)	50
C	variants after line 500 (70)	45

x

A	variants to line 500 (90)	75
A	variants after line 500 (110)	6
C	variants to line 500 (80)	7
C	variants after line 500 (70)	55

y

A	variants to line 500 (90)	50
A	variants after line 500 (110)	55
C	variants to line 500 (80)	25
C	variants after line 500 (70)	20

z

A	variants to line 500 (90)	5
A	variants after line 500 (110)	6
C	variants to line 500 (80)	6
C	variants after line 500 (70)	8

In various articles I have suggested that if over half the variants characteristic of a fundamental group are present in a particular manuscript, then that manuscript is a member of that group (for example, in my study with Robert O'Hara, Robinson and O'Hara, 1993.) If between a quarter and a third of the variants are present in a manuscript, then it is likely that the manuscript is contaminated by that group. If less than ten per cent of the variants characteristic of a fundamental group are present, this is probably simple accidental agreement and means nothing at all. We can use this pragmatic rule-of-thumb to allocate these five manuscripts, as follows:

v: more than half the A variants in both sections so a member of the A group

w: more than half the C variants in both sections and so a member of the

C group

x: more than half the A variants in the first part, and more than half the C variants in the second half: a member of A in the first half, then shifts exemplar and joins C in the second half

y: more than half the A variants throughout, though rather fewer than in other manuscripts. The number of C variants throughout (45 of 150) is too high for chance agreement and suggests deliberate import of readings from a C manuscript

z: the low number of variants from both A and C suggests that these agreements are simple chance, and that z is unrelated to manuscripts of both the A and C groups.

This system works well when the numbers of variants are reasonably unambiguous, as in this example. Indeed, for the majority of cases (as we will see) the numbers are indeed unambiguous and manuscripts clearly do belong or do not belong to a particular group. However, there are also borderline cases where the numbers of variants fall in the margin between accident and contamination, or between contamination and group membership. These have to be treated on an individual basis, and in some cases (as we will see) there is just not sufficient information to form a judgement.

---

3. The analysis workshop

Our analysis, as we have described it in the previous section, proceeds through the following stages:

identification of the fundamental groups through SplitsTree, and refinement of this identification through PAUP (Section 2.4 above)

use of the variant database to identify the distinct groups of readings characteristically present in each fundamental group, and so likely to have been introduced by the common ancestor of each group (Section 2.5 above)

use of the variant group profiles to allocate the manuscripts to groups and to identify manuscripts which show evidence of contamination and of shift of exemplar (Section 2.6 above)

This section will take you through the stages of this analysis, for The General Prologue. The main tools we use (SplitsTree and the Variant

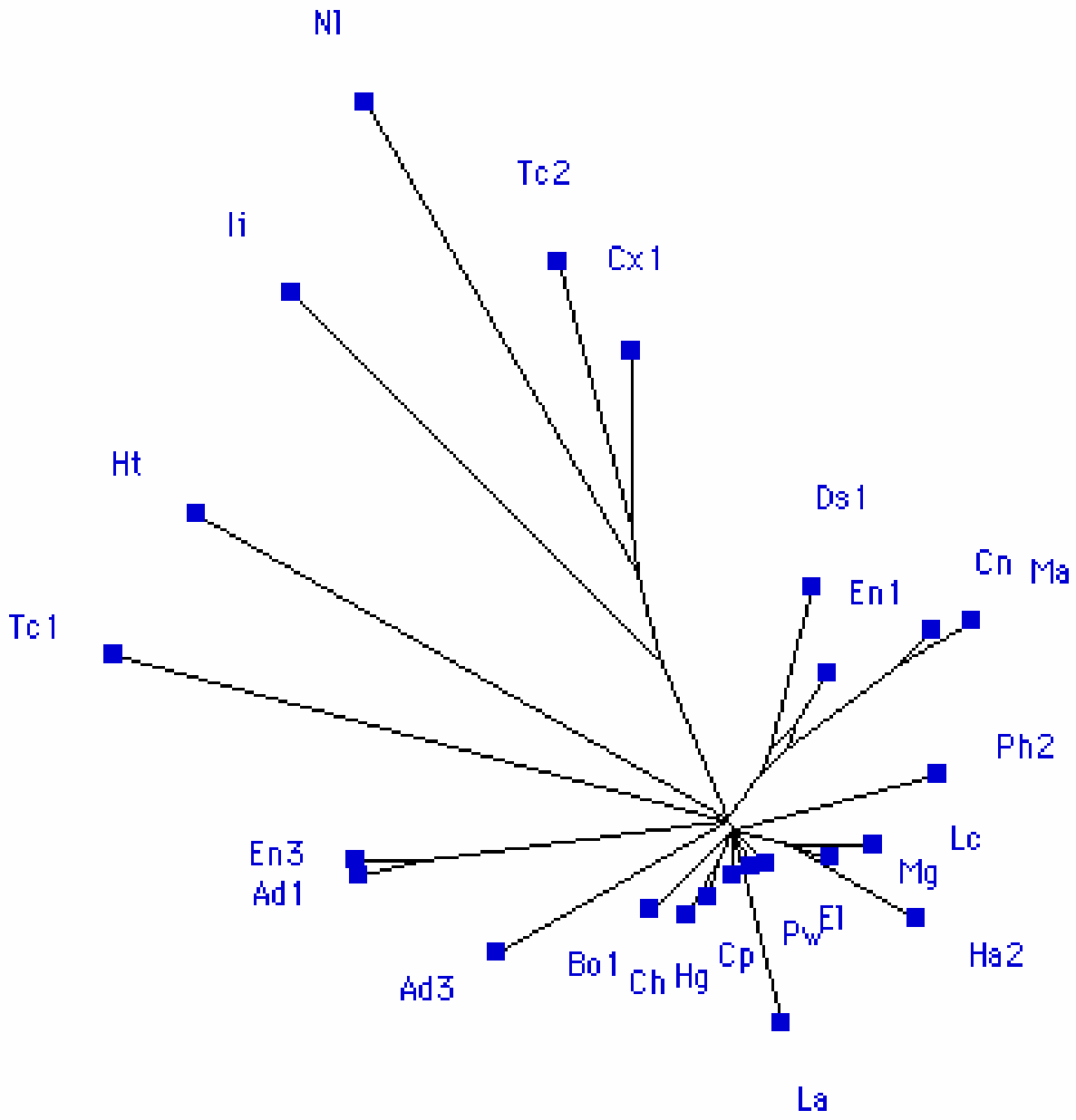
Database) are provided with this CD-ROM and exercises for you to use them are also provided. In the separate Stemmatic Commentary, I show how what we have learnt of the manuscript relations might be used in discussion of particular readings.

3.1 Identifying the fundamental groups

In Figure 6 above, we gave a SplitsTree graph based on the data of the agreements and disagreements for 42 witnesses to lines 1-250 of The General Prologue: all the manuscripts which contain reasonably complete texts of these first 250 lines. You can have SplitsTree recreate this graph, by clicking here. We have found that when dealing with large numbers of manuscripts in SplitsTree, the program may sometimes fail to show relations which become apparent when some of the manuscripts are removed. It appears that the program can become confused by a high proportion of accidental agreement or of contamination in particular manuscripts, and so be unable to distinguish relationships in other manuscripts. By removing certain manuscripts from the analysis, then, it will be possible to see more clearly the relations in the manuscripts which remain. Thus: analysis of the 42 manuscripts showed only seventeen which appeared to belong to groupings. Reducing the number analysed might actually increase the number which belonged to groupings.

3.1.1 Lines 1-250, and alpha

I experimented with the SplitsTree analysis of lines 1-250 in this manner. It appeared that the clearest picture of the manuscript relations was given if we removed eighteen manuscripts from the forty-two analysed in Figure 6. You can carry out this analysis by clicking here:


Nl
Tc2
Ii
Cx1
Ht
Ds1
Cn
Ma
En1
Tc1
Ph2
Lc
En3
Ad1
Mg
El
Pw
Ha2
Ad3
Bo1
Ch
Hg
Cp
La

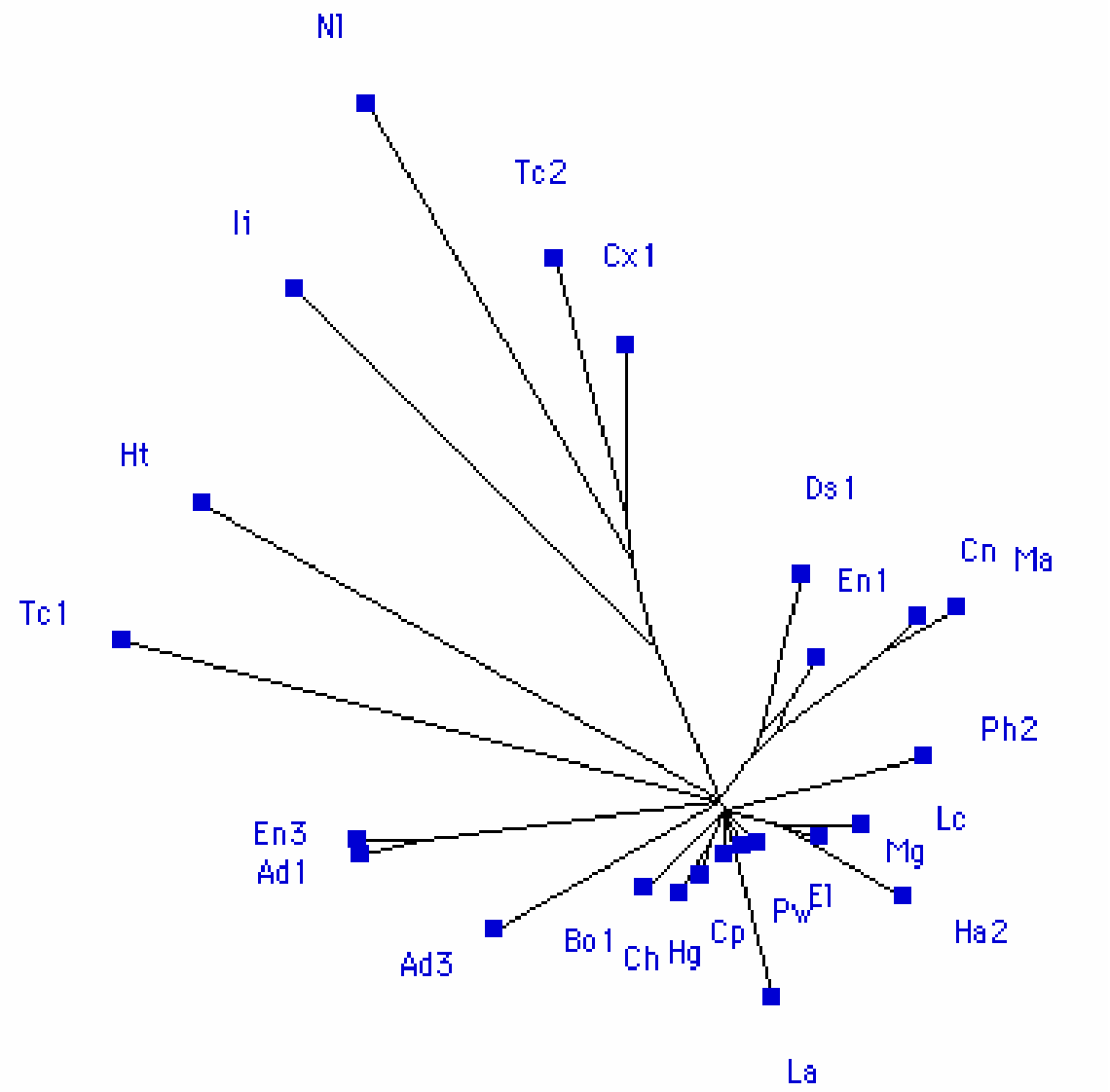


Figure 11: a SplitsTree graph for lines 1-250, giving more detail of manuscript relations

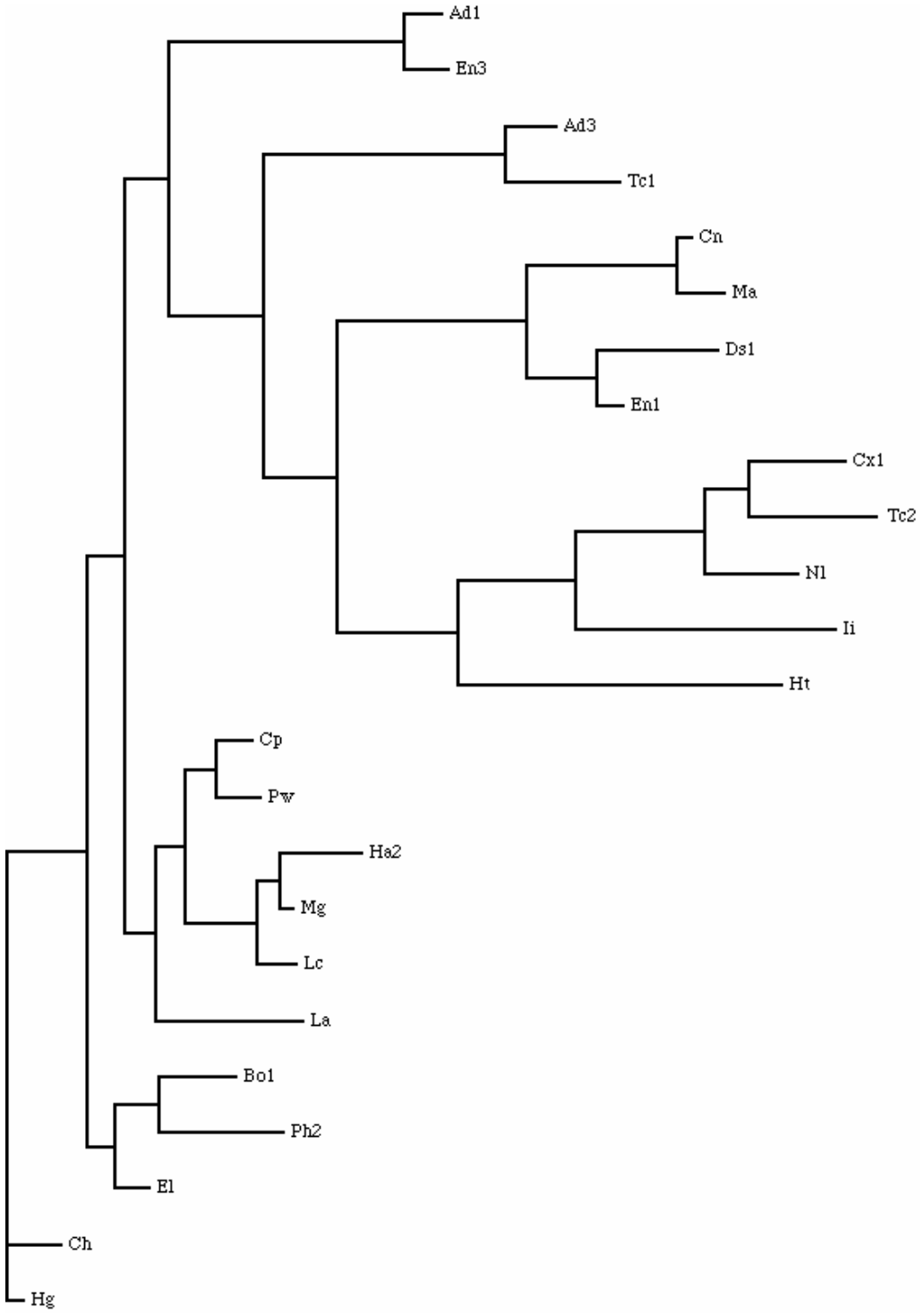
The manuscripts here excluded and present in the earlier tree are as follows: Bo^{2} Cx^{2} Fi Gg Ha^{3} Ha^{4} Ld^{1} Ld^{2} Mm Pn Py Ps Se Ry^{1} Ry^{2} Sl^{2} To^{1} Wy. This tree suggests a tree-like relationship for twenty-four manuscripts. We can add to this twenty-four the three manuscripts (actually, printed editions) Cx^{2} Pn Wy. These were excluded from the

tree because it is known from Caxton's preface to the second edition that he used a 'better manuscript' to correct his first edition (Cx^1) while making Cx^2: thus, contamination. Inclusion of Cx^2 (and its descendants Pn Wy) might therefore obscure the tree. However, the earlier tree in Figure 6 showed, contamination aside, that Cx^2 Pn Wy are related to the manuscripts in the top of this tree, Ii Nl Cx^1 Tc^2 (Manly and Rickerts b group). Thus, we now have a total of 27 manuscripts for which SplitsTree suggests a 'tree-like' relationship.

The picture of relations suggested by this tree has several notable features. Firstly, the three major manuscript groupings found by Manly and Rickert are apparent in this tree: a (Cn Ds^1 En^1 Ma); b (Ii Nl Cx^1 Tc^2); cd (Ha^2 Lc Mg; also Cp La Pw come from the same node). Secondly, the tree suggests a fundamental cleavage in the tradition between the manuscripts in the top half of the tree (the a and b groups; but also Ad^3 Ht En^3 Ad^1 Tc^1) and those in the bottom half (including the two usually considered most authoritative, El Hg; also the cd group and Bo^1 Ch Ph^2). This cleavage is represented by the distance between the node from which all the top group radiate and the node from which all the bottom group radiate. In stemmatic terms, this suggests that at each of these nodes there is a group of variants characteristically found only in the manuscripts descended from each node. Further, this cleavage corresponds to the hypothesis, presented in my analysis of the manuscripts of The Wife of Bath's Prologue, that there was a very early exemplar of the *Tales* which we call (following Dan Mosser) alpha (Robinson 1997.) Alpha seems to have been the ancestor of Manly and Rickert's a and b groups throughout, but also seems to have been the ancestor of other manuscripts at other points: thus, here, Ad^3 En^3 Ht Ad^1 Tc^1 also. Thus, alpha would correspond to the node from which all the manuscripts in the top half of the graph radiate: it will be a distinct exemplar, representing a line of descent distinct from that of Hg El and other manuscripts in the bottom half of the graph. At various points in this workshop, we will return to analysis of alpha: to isolate exactly the variants it may have introduced to the tradition; to assess their character; to determine its influence across the tradition.

I remarked above that while SplitsTree is good at estimating whether or not certain manuscripts may be seen in a tree-like relationship, PAUP may be more useful in determining the exact relationships between the

manuscripts than is SplitsTree. Thus, we gave PAUP exactly the same data for these 250 lines in these 24 manuscripts, and arrived at this tree:


Ad1
En3
Ad3
Tc1
Cn
Ma
Ds1
En1
Cx1
Tc2
Nl
Ii
Ht
Cp
Pw
Ha2
Mg
Lc
La
Bo1
Ph2
El
Ch
Hg

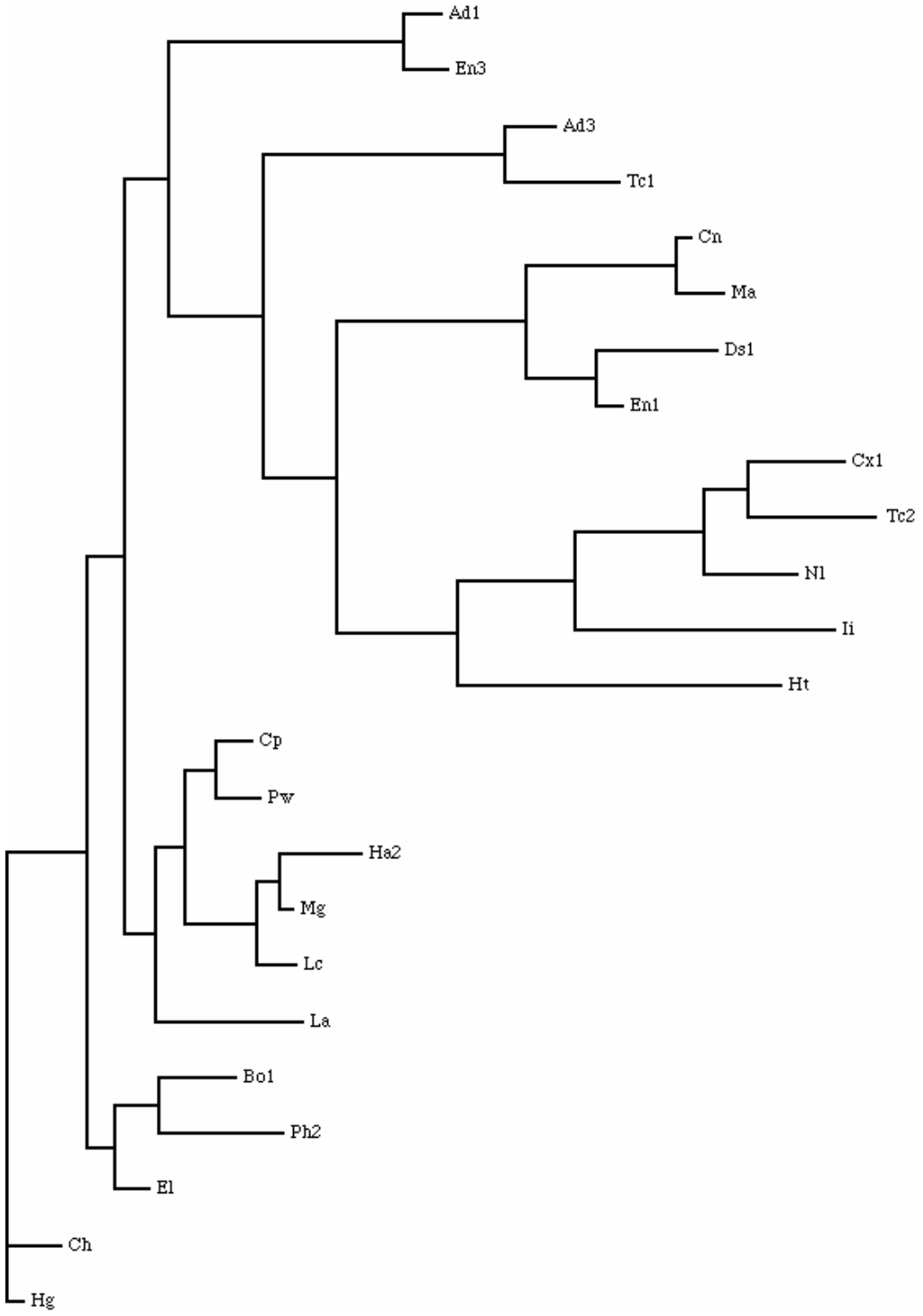

Ad1
En3
Ad3
Tc1
Cn
Ma
Ds1
En1
Cx1
Tc2
Nl
Ii
Ht
Cp
Pw
Ha2
Mg
Lc
La
Bo1
Ph2
El
Ch
Hg

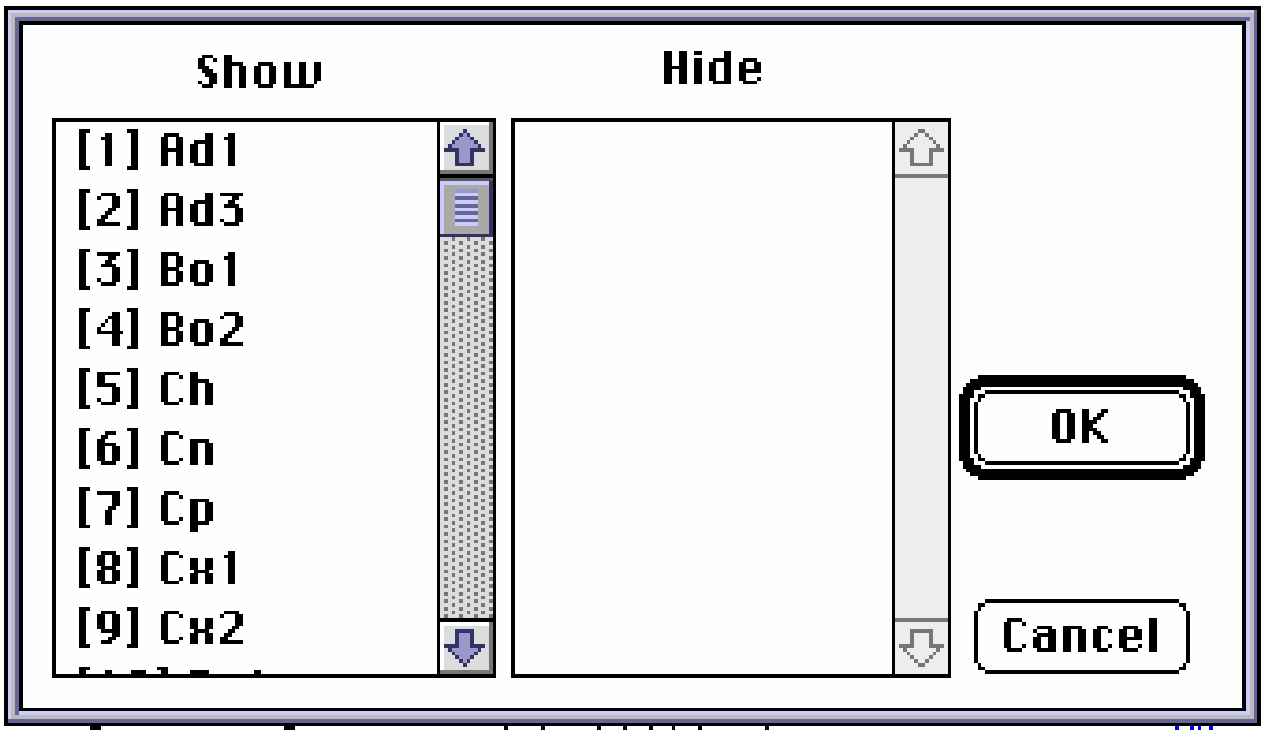
Figure 12: the tree ('cladogram') for lines 1-250 in 24 manuscripts, produced by PAUP (file GPWS2.nex). [Note: the NEXUS source file for this tree is GPWS2.nex, in the 'docs' directory. If you have PAUP, start PAUP, open the file and execute it].

Indeed, this line shows several further relationships among the manuscripts which SplitsTree could not show. The pair Bo^1 Ph^2, separated in the SplitsTree graph though descended from the one node, are here shown as a pair, as they are elsewhere in the *Tales*. The cd manuscripts Cp La Pw are here brought together with the triplet Ha^2 Lc Mg, again as they are elsewhere in the *Tales*. The groups a and b are shown here as descended from a single exclusive common ancestor below alpha, again as they are elsewhere in the *Tales*. The pair Ad^3 and Tc^1 are brought together, as they are elsewhere in the tales. Most interesting of all, PAUP confirms the cleavage between the two halves of the tradition, with the manuscripts in the top half of the cladogram all deriving from a single node corresponding precisely to the same group of manuscripts (ab plus Ad^3 Ht En^3 Ad^1 Tc^1) shown in the SplitsTree graph as descending from a single node. Later in this workshop, we will use the variant database to explore these relations further.

Exercises: now, try these for yourself..

1. You can go through the processes by which we moved from the tree shown in Figure 6 to that shown in Figure 11, as follows:

- Load SplitsTree and the file GPWS1.vm by clicking here. SplitsTree will then give you the tree given in Figure 6.
- To move from this tree to that given in Figure 11: you need to delete 18 manuscripts ('taxa' in evolutionary biology terms) from the analysis. These manuscripts are: Bo^2 Cx^2 Fi Gg Ha^3 Ha^4 Ld^1 Ld^2 Mm Pn Ps Py Ry^1 Ry^2 Se Sl^2 To^1 Wy. To delete them, choose the **Taxa...** command from the **Options** menu. You will see a dialogue box as follows:


Show
Hide
[1] Ad1
[2] Ad3
[3] Bo1
[4] Bo2
[5] Ch
[6] Cn
[7] Cp
[8] Cx1
[9] Cx2
OK
Cancel

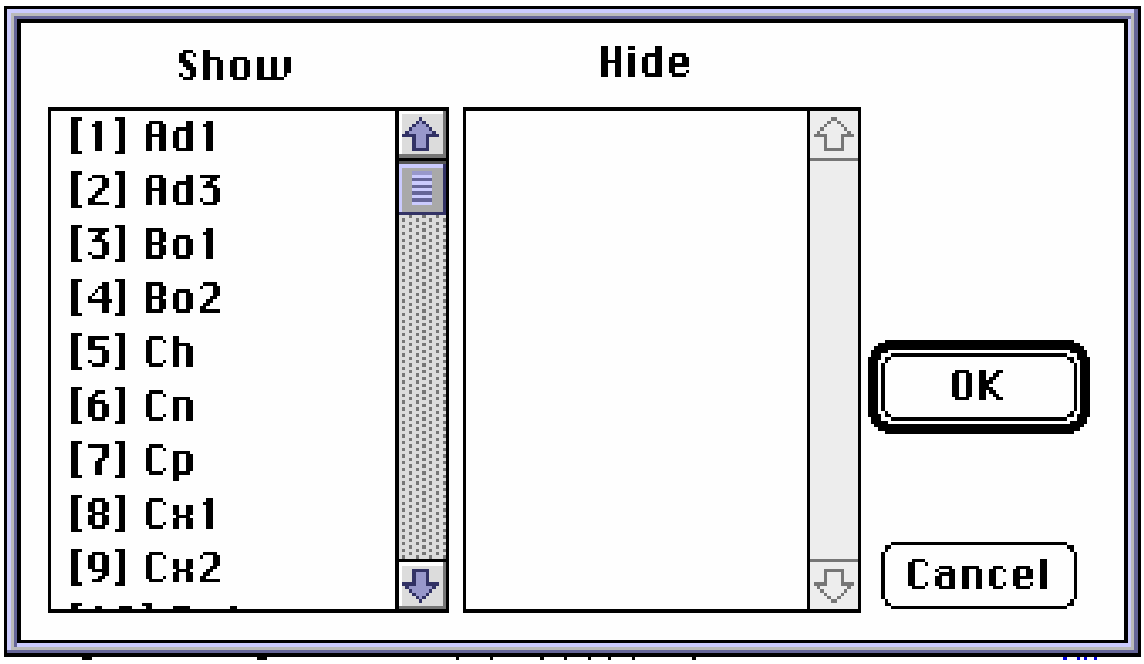
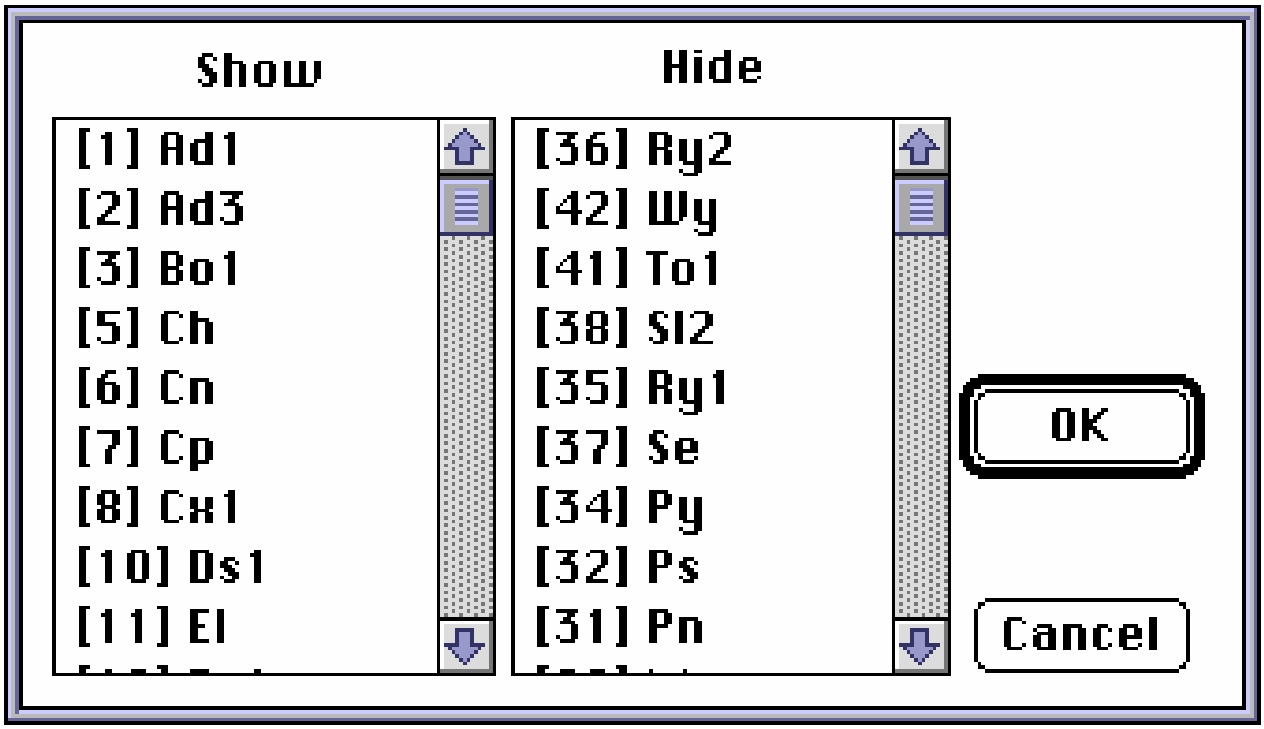


Figure 13: the Taxa dialogue box in SplitsTree

You can remove manuscripts from the analysis simply by clicking on them in the **Show** box, so that they move over to the **Hide** box. Do this now for the eighteen manuscripts Bo^2 Cx^2 Fi Gg Ha^3 Ha^4 Ld^1 Ld^2 Mm Pn Ps Py Ry^1 Ry^2 Se Sl^2 To^1 Wy. The box should now appear as follows, or similar, depending on the order in which you have chosen the manuscripts:


Show
Hide
[1] Ad1
[2] Ad3
[3] Bo1
[5] Ch
[6] Cn
[7] Cp
[8] Cx1
[10] Ds1
[11] El
[36] Ry2
[42] Wy
[41] To1
[38] Sl2
[35] Ry1
[37] Se
[34] Py
[32] Ps
[31] Pn
OK
Cancel

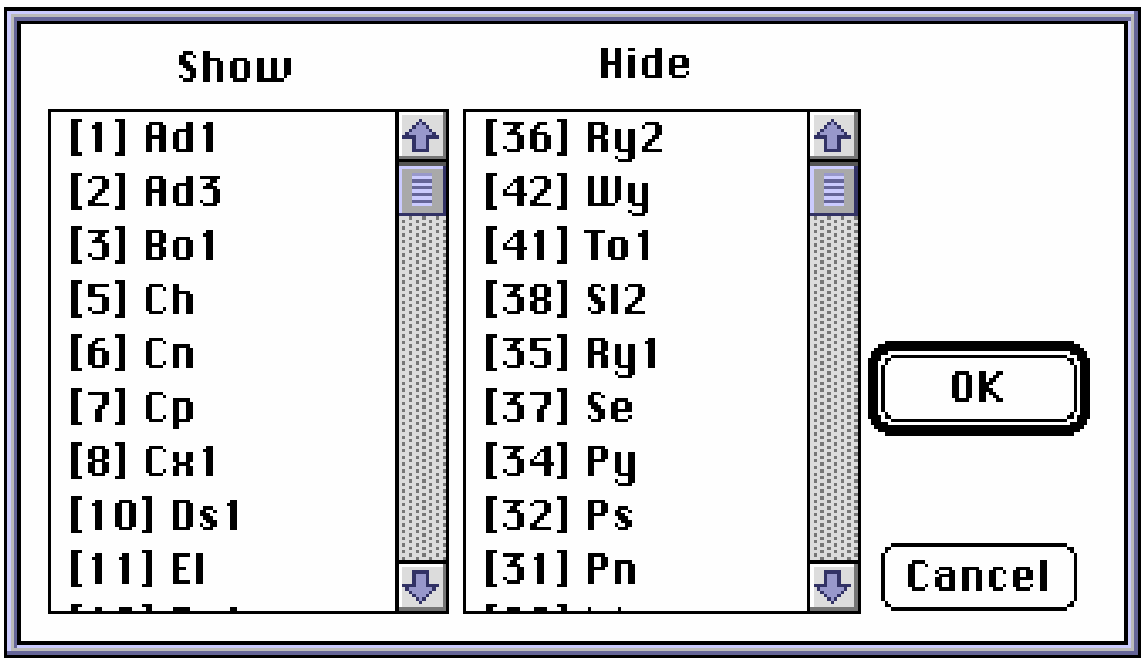
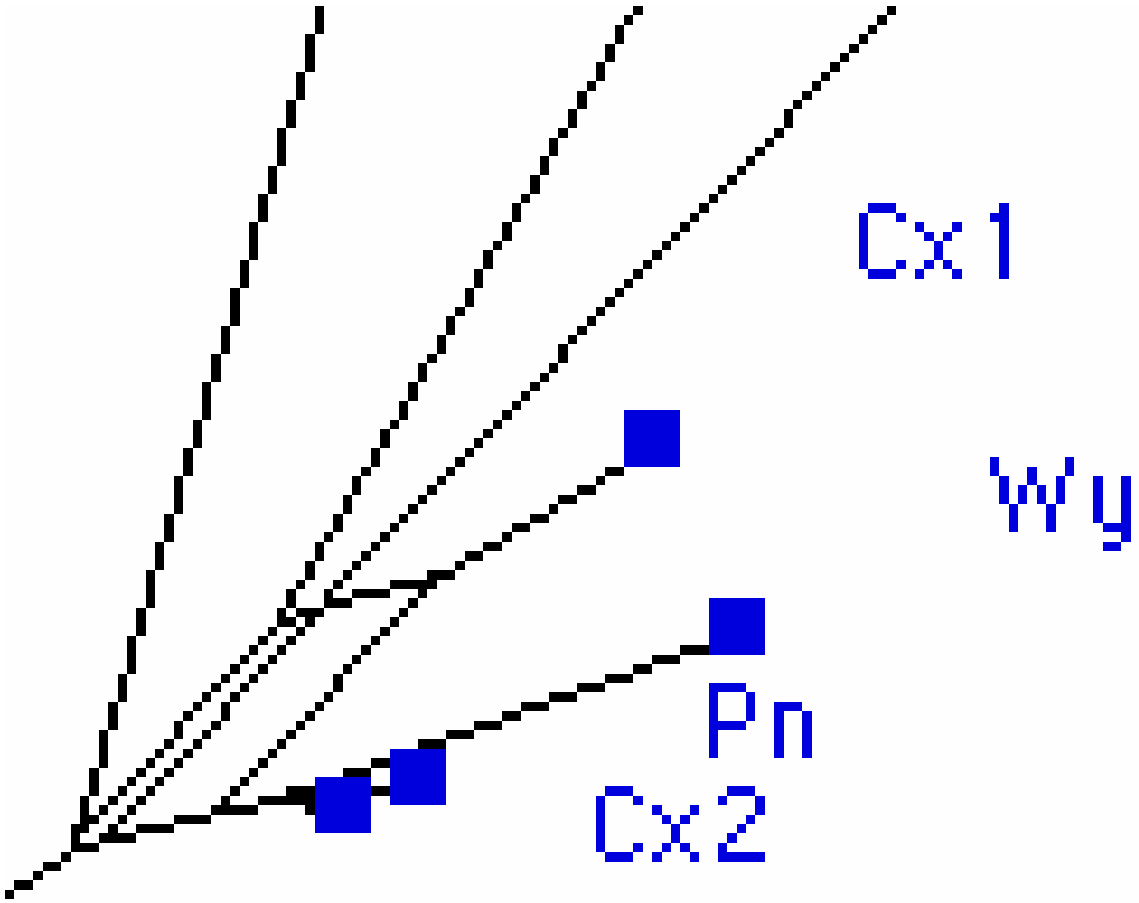


Figure 14: selecting manuscripts for SplitsTree analysis

Now, run the analysis again by clicking the **OK** button. You should now get the same tree as given in Figure 11. You can also get this tree by clicking here.

2. Once you have derived the same tree as Figure 11, but still working with file GPWS1.vm, put the three printed editions Cx^2 Pn Wy back in the analysis, again by choosing the **Taxa...** command from the **Options** and then returning these three to the **Show** box by clicking on their names in the **Hide** box, and rerun the analysis.

You should see two differences between this tree and that in Figure 11. The first difference is that the trio Cx^2 Pn Wy do not come cleanly from a single point, but from a 'box' adjacent to Cx^1 as follows:


Cx1
Wy
Pn
Cx2

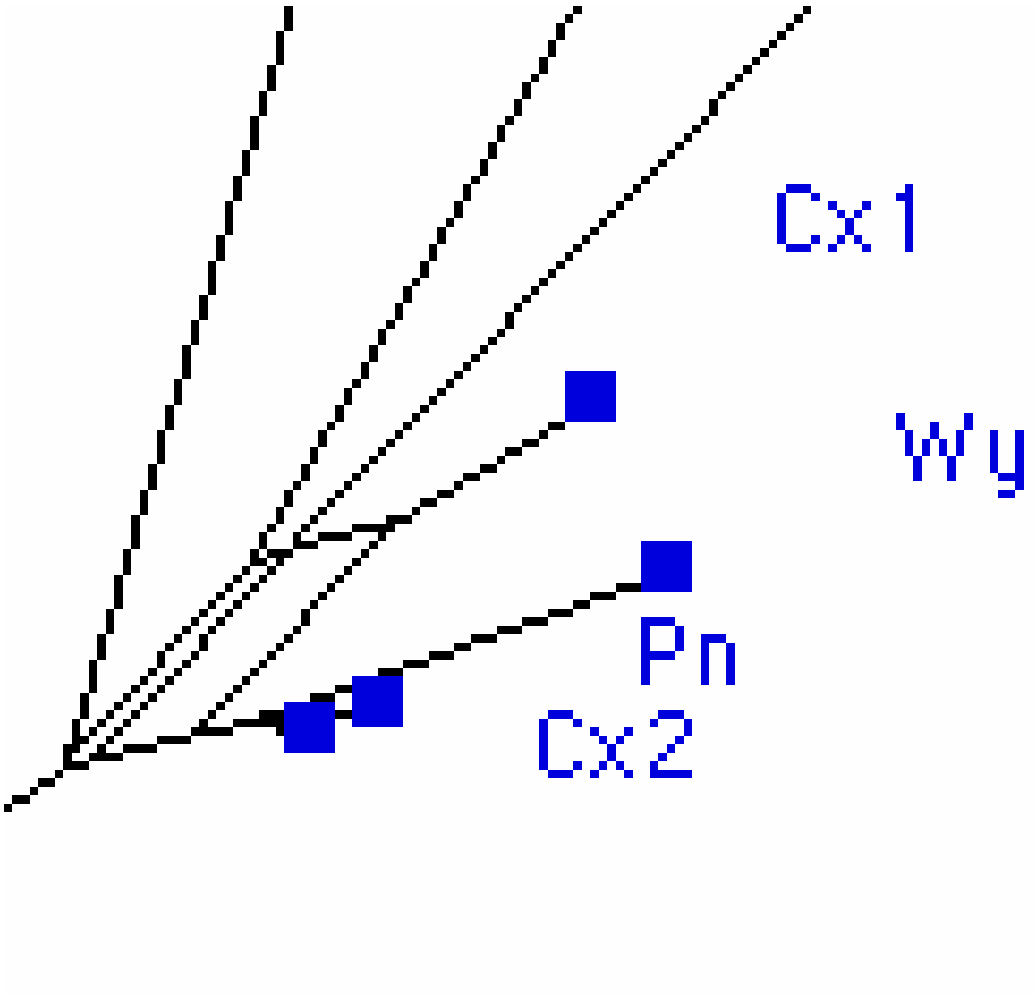


Figure 15: the relationship of Cx^2 and others

The second difference is that the separation between the two nodes defining the whole tradition in Figure 11 (the top representing what we call alpha; the bottom the remainder of the manuscripts) disappears in this analysis. The two phenomena are actually related. As I observe above, we have Caxton's own statement that he ' corrected' the text of Cx^1 when preparing Cx^2, by using a 'better manuscript'. The characteristics of this 'better manuscript' have been studied by several scholars, and in my analysis of the Wife of Bath's Prologue (Robinson 1997.) This analysis has suggested that this '

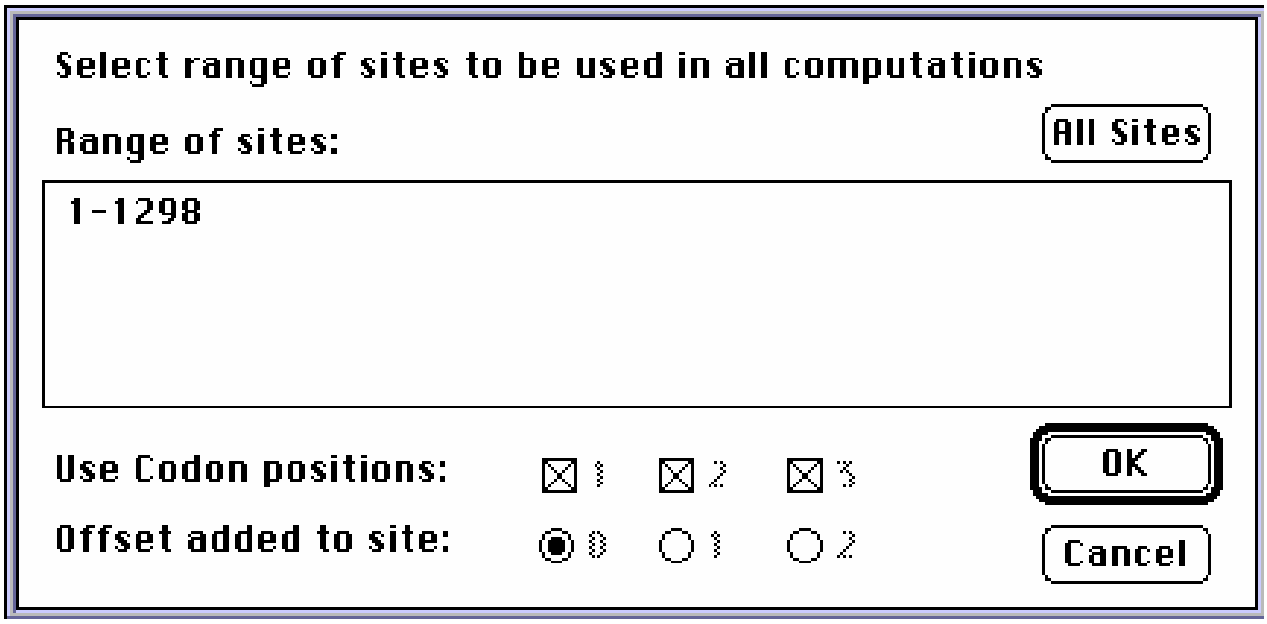
better manuscript' was indeed a manuscript very close to the original of the whole tradition. Thus, it contains many readings not found in Cx^1 and other b group manuscripts, but found in (for example) Ch Hg El — manuscripts otherwise widely separated from Cx^2 — and in the bottom half of the tree given by Figure 11. SplitsTree tries to express this by the box in Figure 15, at the junction of the lines linking Cx^1 Wy Cx^2 and Pn. This box suggests that the point of attachment of Cx^2 to the tree cannot be precisely defined by SplitsTree: it is near Cx^1, but that is all the program is able to say with certainty. Further, this presence in Cx^2 of variants elsewhere found only in the manuscripts in the bottom half of the table has the effect of closing up the space between the two nodes. SplitsTree is not able to see that these are two distinct nodes because there are variants found in both the manuscripts of the bottom half and also in Cx^2.

This case is a particularly clear instance of how contamination in part of the tradition can affect SplitsTree's ability to show the relations of manuscripts in other parts of the tradition.

3. You can now try and improve the tree by deleting other manuscripts from the analysis. For example: PAUP suggested that the manuscripts Cp La Pw are related to the group Ha^2 Lc Mg, in Manly and Rickert's cd group. Figure 11 does not show this relationship. Can you have SplitsTree show this relationship by removing other manuscripts from the analysis? (hint: remove Ii, another b manuscript which shows signs of contamination; also remove some of the manuscripts in the bottom half e.g. Ch; perhaps Hg El too). The file GPWS3.vm shows my attempt at this. Can you do better?
4. To gain a clearer picture of the manuscripts in each half of the tree, try removing all the manuscripts from the other half. Thus: to see the relationships of the manuscripts in the bottom half, remove all those in the top half of the tree (this is done in file GPWS4.vm). To see the relationships in the top half, remove those in the bottom half (file GPWS5.vm). You will see that for these manuscripts, SplitsTree now gives relationships very close to those suggested by PAUP in Figure 12.
5. You may have noticed the figures in the bottom of the tree screen SplitsTree gives when it has finished processing. The most important of these is 'fit', expressed as a percentage. This tells you the

proportion of data which SplitsTree has found consistent with the tree which it offers you: the higher the fit, the better that tree describes the data. As the data becomes more and more 'tree-like', as we look at fewer and more related manuscripts, you will see the 'fit' figure rise. For the 42 manuscripts in Figure 6, the fit figure is just 43.8%. In Figure 11 this rises to 50.5. For the 11 manuscripts only in the bottom half of Figure 11 it is 79.8 per cent, suggesting a high degree of consistency in the data for these manuscripts.

6. All the analyses in the previous exercises are based on lines 1-250. You could run any of these analyses, using any of the files in the directory 'docs', on different selections of data by using the **Sites...** command on the **Options** menu. This dialogue will appear:



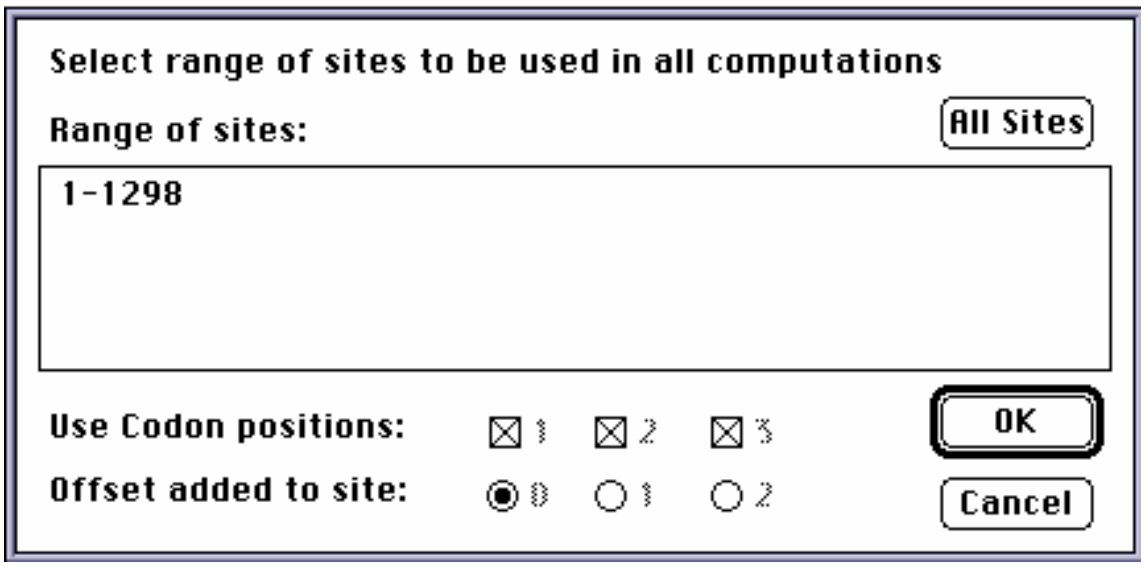


Figure 16: the Sites dialogue box in SplitsTree

Sites 1-1298 correspond to the variants in lines 1 to 250. Typing different values into this box will give you an analysis over different selections of lines, as follows:

Lines 1-50: sites 1-243

Lines 51-100: sites 244-489

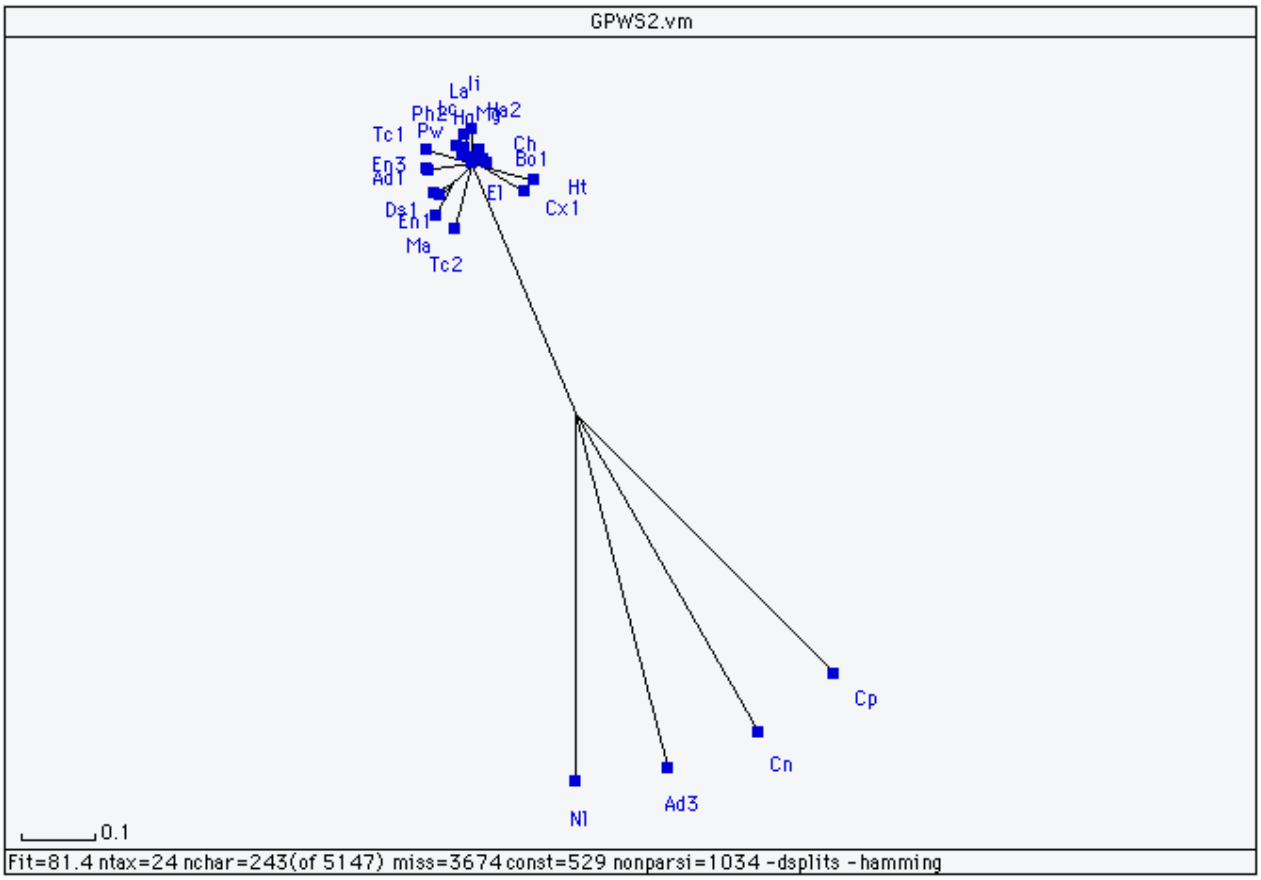
Lines 101-150: sites 490-744

Lines 151-200: sites 745-1027

Lines 201-250: sites 1028-1298

You can calculate the boundaries of other segments of text by reading the file 'GP.vm' in a text-only text editor. This file contains the SplitsTree source for the whole of The General Prologue. The section labelled '[VARIANT KEY]' in this file gives a key to the 'sites ' of all the variants.

WARNING: not all the manuscripts are extant for these smaller sections of the text. If you run SplitsTree over sections of the data where a manuscript is 'out' you will get bizarre results. For example: running SplitsTree on the file GPWS2a.vm for lines 1-50 (by typing 1-243 into the Sites dialogue box) will give the following result:


GPWS2.vm
La
li
Ph2
Ha
Ma2
Tc1
Pw
Ch
Bo1
En3
Ad1
Ht
El
Cx1
Ds1
En1
Ma
Tc2
Cp
Cn
Ad3
Nl
0.1
Fit=81.4 ntax=24 nchar=243(of 5147) miss=3674 const=529 nonparsi=1034 -dsplits -hamming

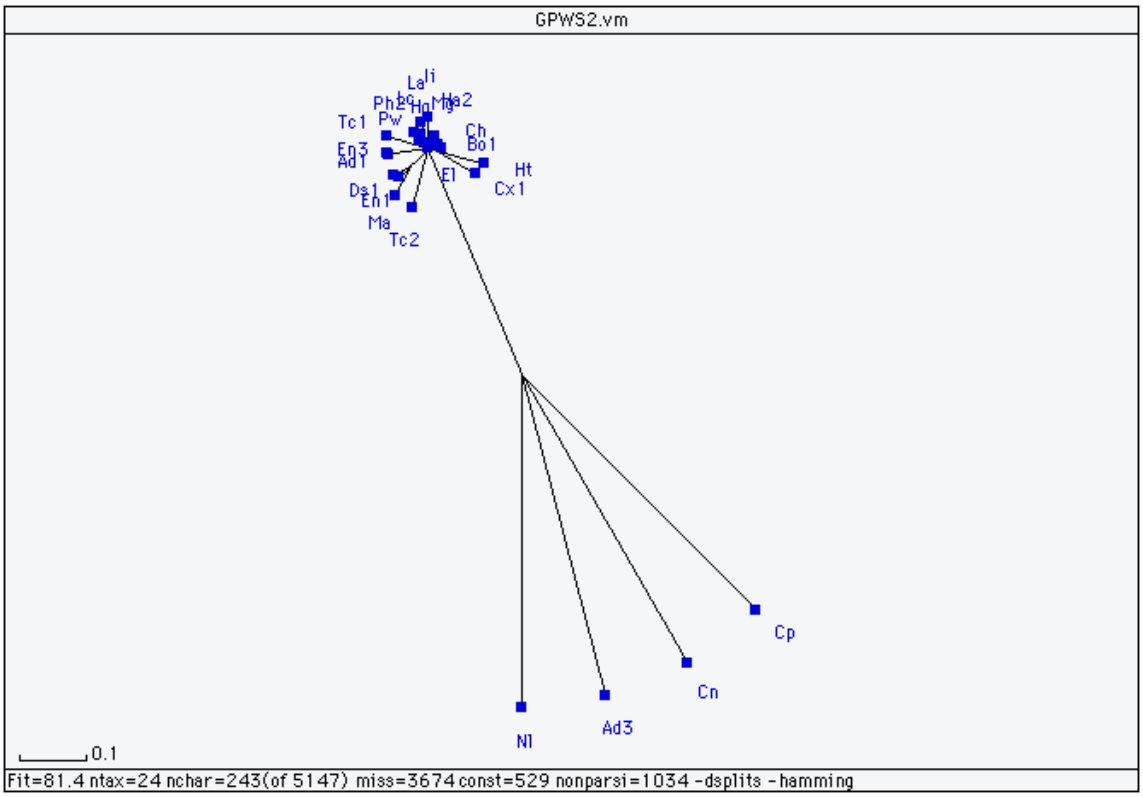


Figure 17: Using SplitsTree where manuscripts have no data

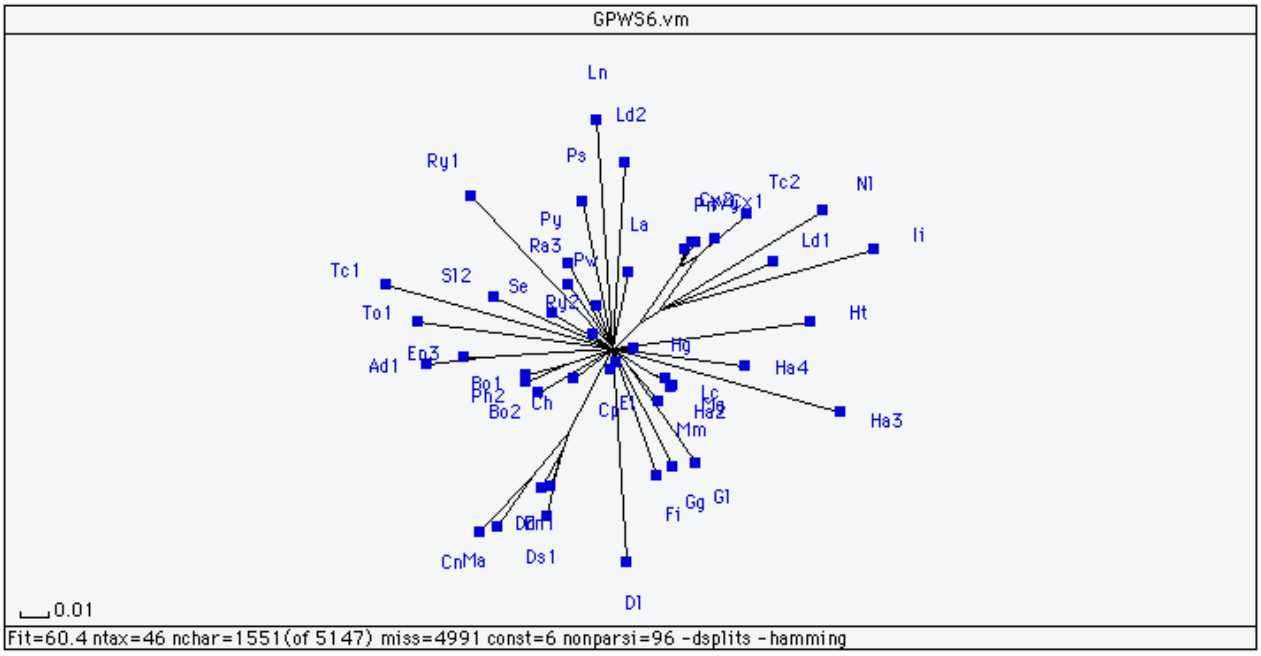
SplitsTree has produced this result because the four manuscripts Nl Ad3 Cn Cp are all missing lines 1-50, and so have no data. Therefore, SplitsTree sees these four as coming from a single node, but also is unable to see relations in other parts of the data. Remove these four manuscripts from the analysis by using the **Taxa** command and proceed.

Compare the different trees produced by the analyses over different segments of text. How are they similar; how are they different? How can you explain the differences?

We have found that 100 line segments usually provide sufficient evidence of variation to make analysis meaningful. You may like to test this hypothesis for yourself.

3.1.2 Lines 251-500

The next stage of analysis is to test (and, where possible, extend) the picture of fundamental groupings given by SplitsTree for the first 250 lines. If this picture of fundamental groupings is correct, then it will be broadly true for the whole text, with local differences explicable by shift of exemplar, or by manuscripts which are 'out'. Thus, we proceed in the same manner as before. First, we make a SplitsTree graph of all the manuscripts which have lines 251-500, in file GPWS6.vm: that is, for all manuscripts except Ad^{3} Ad^{4} Bw Do Ne Ra^{2} Sl^{1}, in sites 1299-2849. This gives the following tree:



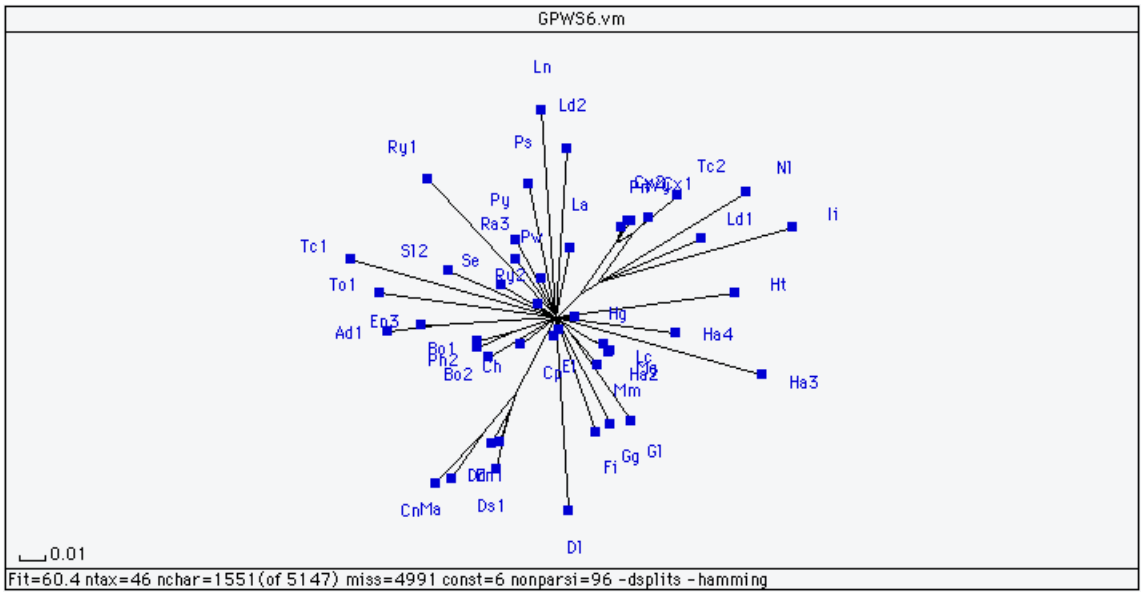
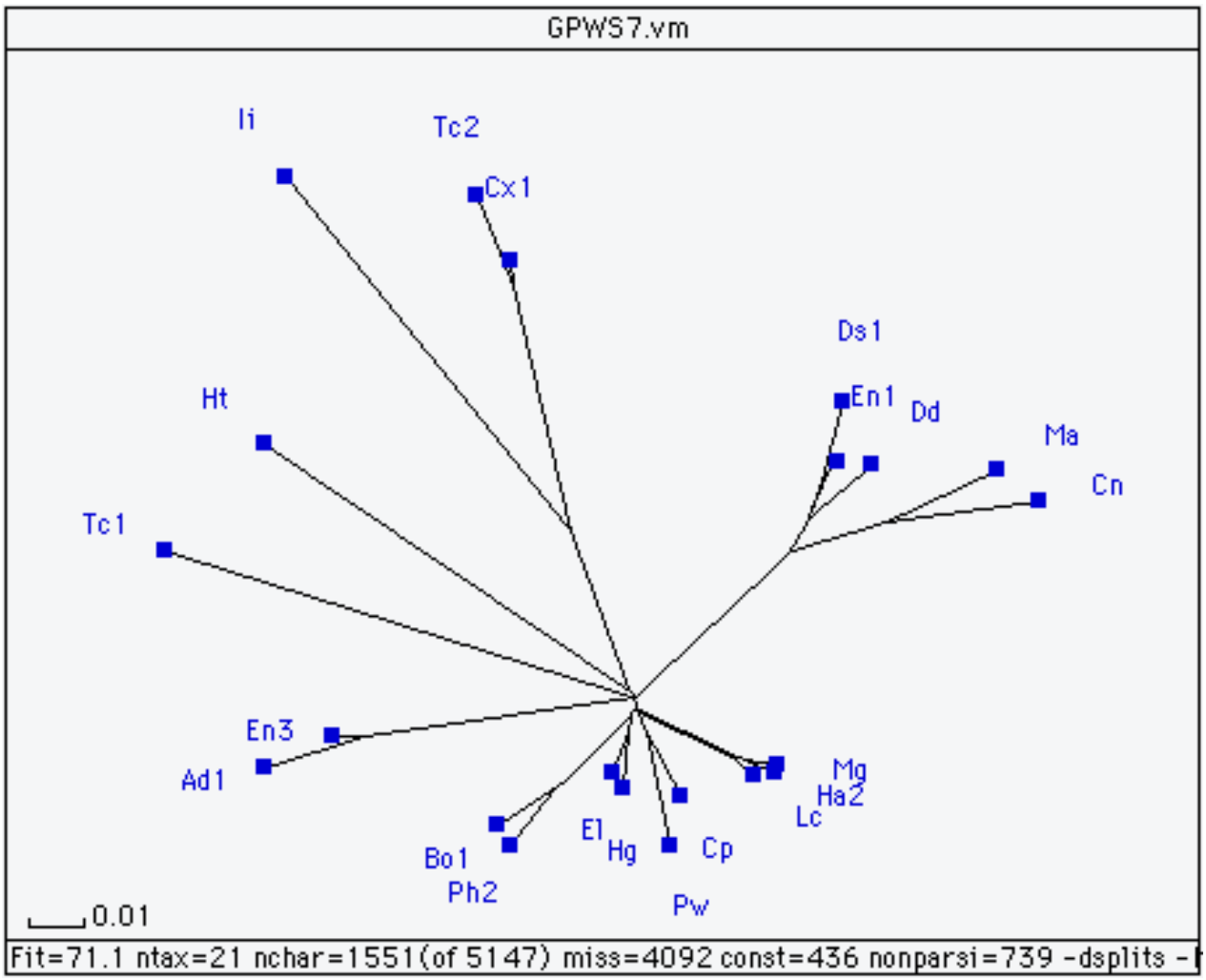


Figure 18: the SplitsTree graph for lines 251-500, in 46 manuscripts

As in the corresponding Figure 6 for lines 1-250, one can see certain basic groupings (indeed, essentially the same groupings as for 1-250), but again most manuscripts show no tree-like behaviour and simply radiate from the centre. Thus, as in the last section, we seek to improve the picture of relationships SplitsTree gives by removing 'non-tree-like' manuscripts from the analysis. Removing Bo^2 Cx^2 Ch Dl Fi Gg Gl Ha^3 Ha^4 La Ld^1 Ld^2 Ln Mm Nl Pn Ps Py Ra^3 Ry^1 Ry^2 Se Sl^2 To^1 Wy in addition to those removed for Figure 18 gives the following tree (file GPWS7.vm):


GPWS7.vm
Ii
Tc2
Cx1
Ds1
En1
Dd
Ma
Cn
Ht
Tc1
En3
Ad1
Mg
Ha2
Lc
El
Hg
Cp
Bo1
Ph2
Pw
0.01
Fit=71.1 ntax=21 nchar=1551(of 5147) miss=4092 const=436 nonparsi=739 -dsplits -

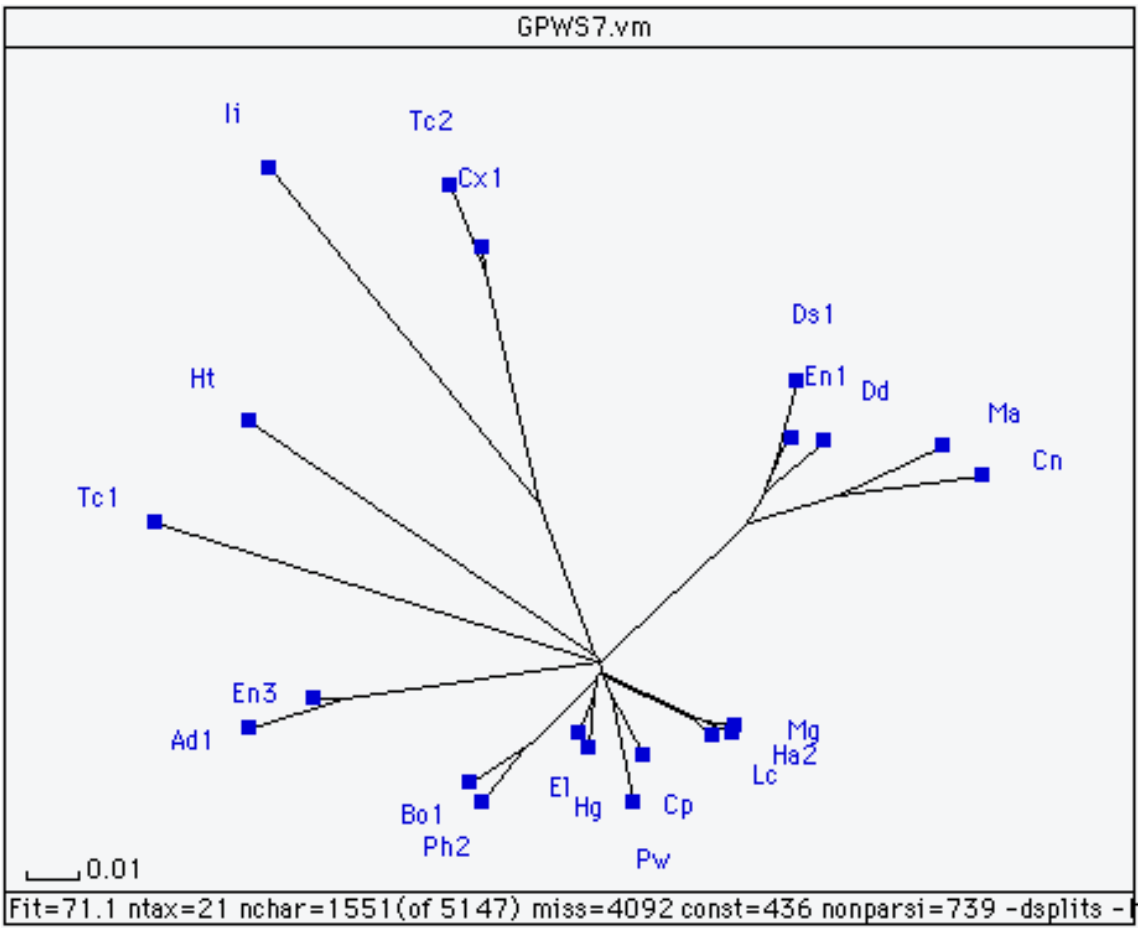
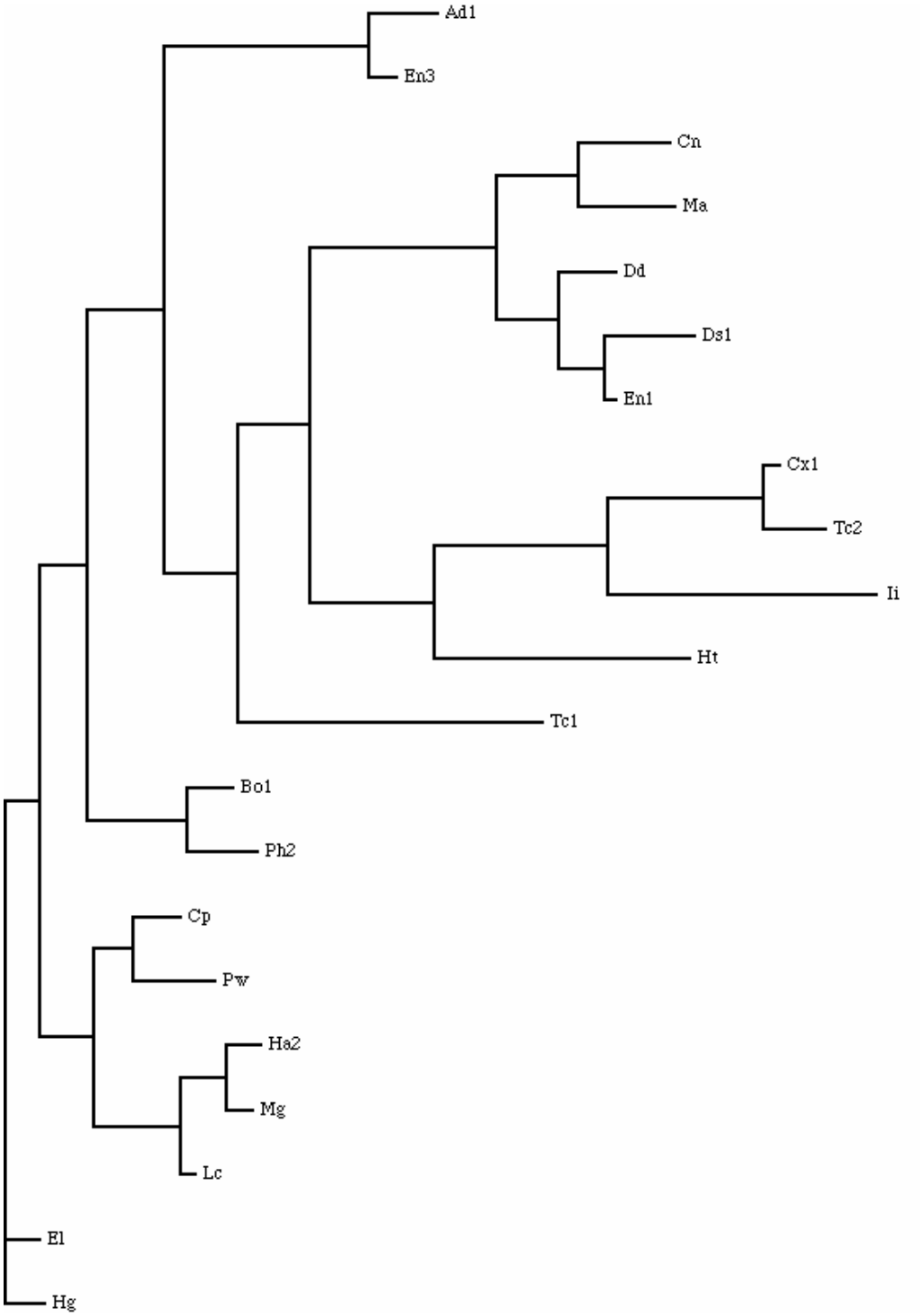


Figure 19: a SplitsTree graph for lines 251-500, giving more detail of manuscript relations

The closeness of this graph to that in Figure 11, for lines 1-250, is remarkable. The same groups <u>abcd</u> appear, indeed in almost exactly the same positions as they are for 1-250, with the same manuscripts associated with them. Especially, we see the same basic cleavage in the manuscript tradition, between those associated with alpha in the top half, and others (including El Hg Bo^{1} Ph^{2} and <u>cd</u>) in the bottom half. The most significant differences between this and Figure 11 are the absence of Ad^{3} from this figure (OUT for most of lines 251-500) and the presence of Dd (OUT for all of lines 1-250). Dd joins the tree in the midst of Manly and Rickert's <u>a</u> group, precisely as it does elsewhere in the *Tales*. The

same datafile was given to PAUP for processing, to see if it could further refine the analysis. PAUP produced this cladogram for these manuscripts in lines 251-500 (from file 'GPWS7.nex'):


Ad1
En3
Cn
Ma
Dd
Ds1
En1
Cx1
Tc2
Ii
Ht
Tc1
Bo1
Ph2
Cp
Pw
Ha2
Mg
Lc
El
Hg

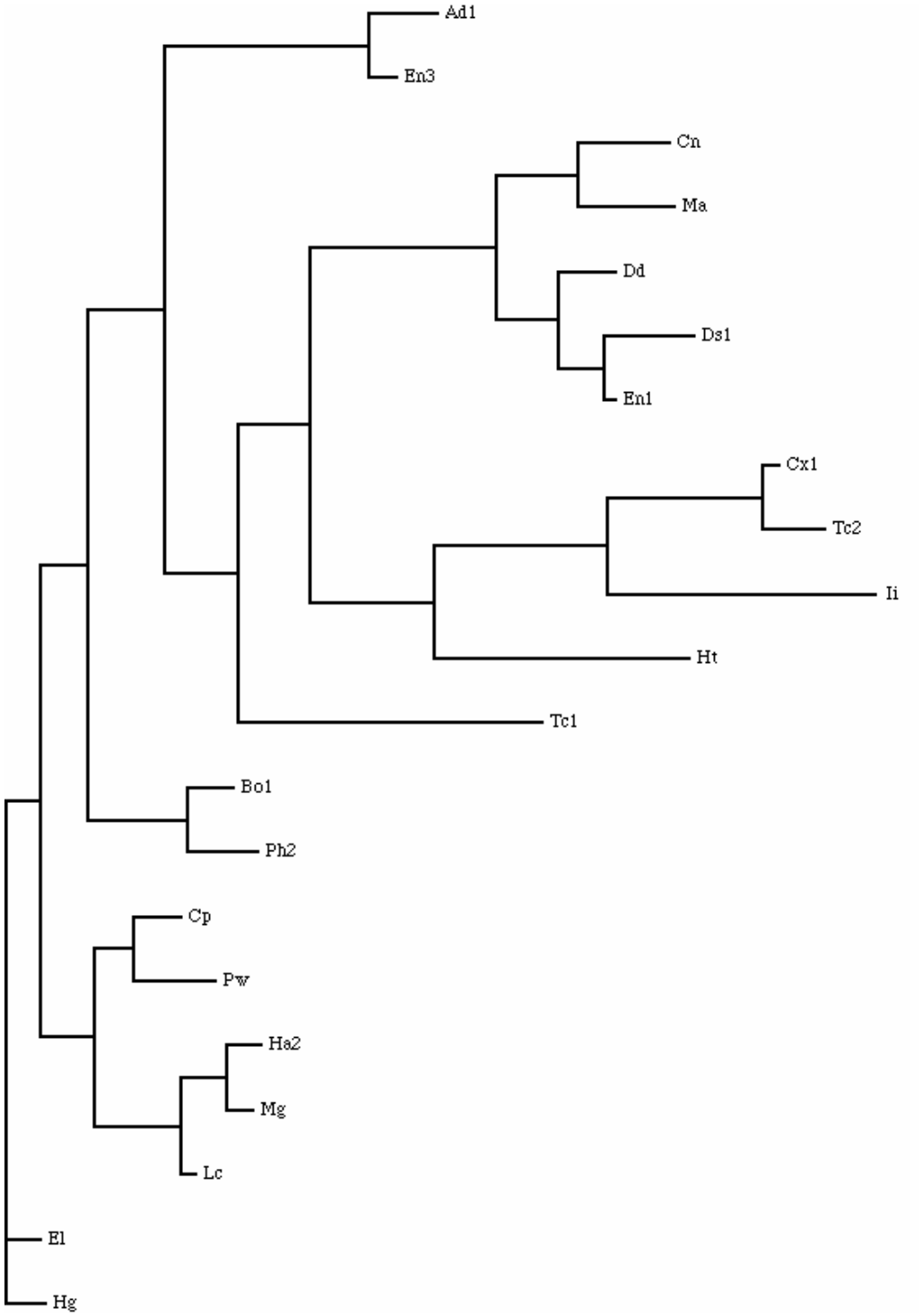

Ad1
En3
Cn
Ma
Dd
Ds1
En1
Cx1
Tc2
Ii
Ht
Tc1
Bo1
Ph2
Cp
Pw
Ha2
Mg
Lc
El
Hg

Figure 20: cladogram for 21 manuscripts, lines 251-500

Again, this is nearly identical to the cladogram in Figure 12. The agreement of the results of both programs, for both segments of texts, suggests that the view of fundamental relations here given is good throughout lines 1-500.

Exercises: now, try these for yourself..

1. You can go through the processes by which we moved from the tree shown in Figure 18 to that shown in Figure 19, as follows:
   - Load SplitsTree and the file GPWS6.vm by clicking HERE. SplitsTree will then give you the tree given if Figure 18.
   - To move from this tree to that given in Figure 19: you need to delete 25 manuscripts ('taxa' in evolutionary biology terms) from the analysis. These manuscripts are: Bo^2 Cx^2 Ch Dl Fi Gg Gl Ha ^{3} Ha^4 La Ld^1 Ld^2 Ln Mm Nl Pn Ps Py Ra^3 Ry^1 Ry^2 Se Sl^2 To^1 Wy. See exercise 1 in 3.1.1 for instructions on how to delete manuscripts from the analysis.

You should now get the same tree as given in Figure 19, file GPWS7.vm

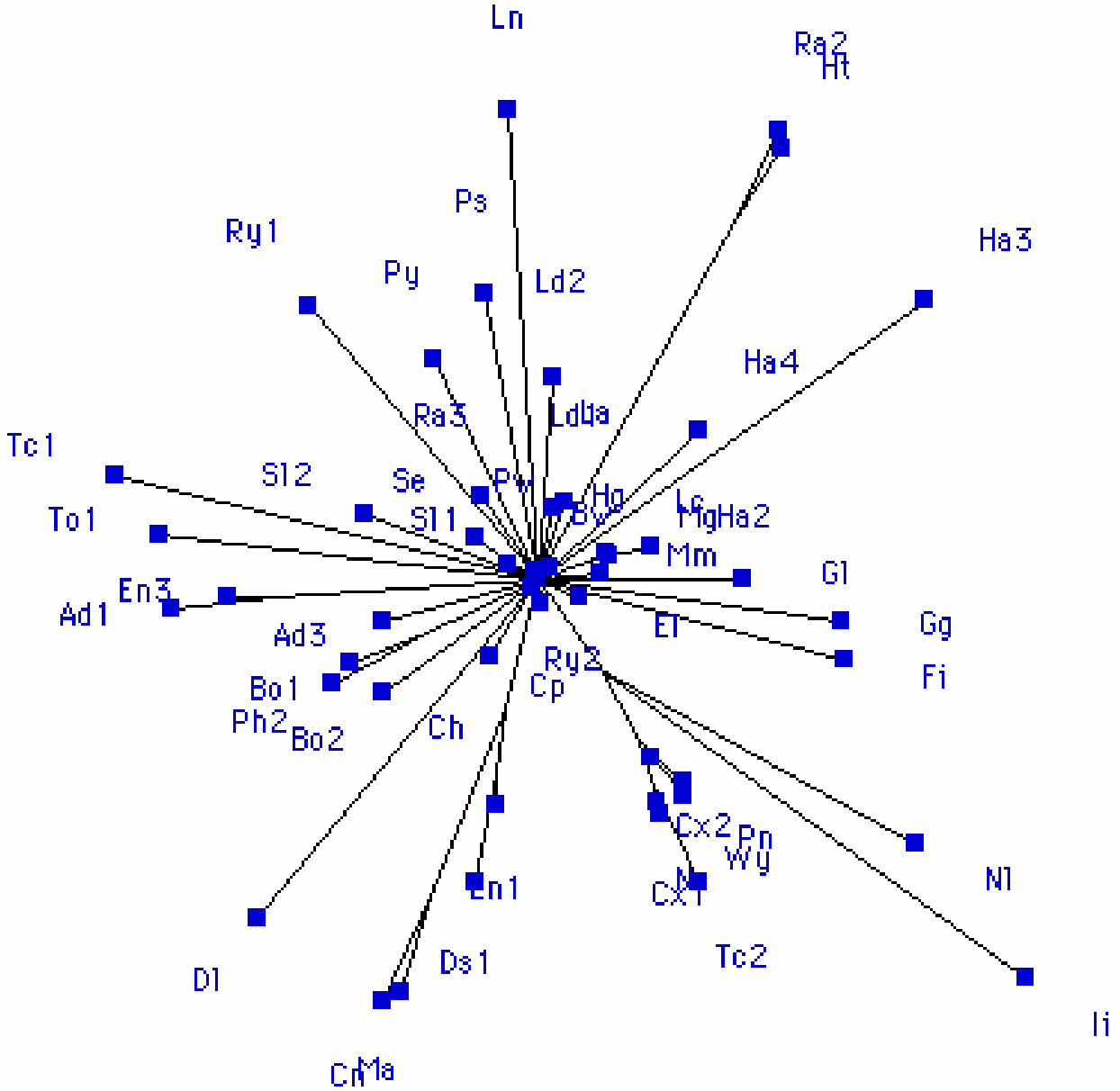
2. Once you have derived the same tree as Figure 19, but still working with file GPWS6.vm, put the three printed editions Cx^2 Pn Wy back in the analysis, again by choosing the **Taxa...** command from the **Options** and then returning these three to the **Show** box by clicking on their names in the **Hide** box, and rerun the analysis. This time, the 'box' connecting Cx^2 to Cx^1 returns; but this time the space between the two nodes at the centre of the tradition remains, as it does not for lines 250-500. Can you explain this difference?
3. Try and improve the tree by deleting other manuscripts from the analysis, as in exercise 3 of 3.1.1.
4. Gain a clearer sense of the relations in each half of the tree by removing all the manuscripts in the other half, as in exercise 4 of 3.1.1.
5. Observe the different values of the 'fit' in each SplitsTree graph (see exercise 5 in 3.1.1). How would you explain these different values?
6. Run the analysis over different segments of the text, as described in exercise 6 of 3.1.1. For lines 251-500 the different segments are:

Lines 251-300: sites 1299-1564
Lines 301-350: sites 1565-1862
Lines 351-400: sites 1863-2160
Lines 401-450: sites 2161-2509
Lines 451-500: sites 2510-2849

Compare the different trees produced by the analyses over different segments of text. How are they similar; how are they different? How can you explain the differences?

3.1.3 Lines 501-end

We now test the analysis offered for lines 1-500 over the remainder of the text of The General Prologue. First, we make a SplitsTree graph of all the manuscripts which have lines 501-end, in file GPWS8.vm: that is, for all manuscripts except Ad^4 Dd Do, in sites 2850-5147. This gives the following tree:


Ln
Ra2
Ht
Ps
Ry1
Ha3
Py
Ld2
Ha4
Ra3
LdUa
Tc1
Sl2
Se
Pw
Hg
Bw
Mg
Ha2
To1
Sl1
Mm
Gl
En3
Ad1
Et
Gg
Ad3
Ry2
Fi
Bo1
Cp
Ph2
Bo2
Ch
Cx2
Pn
Wy
Nl
En1
Cx1
Tc2
Ds1
Dl
Ma
Cn
Ii

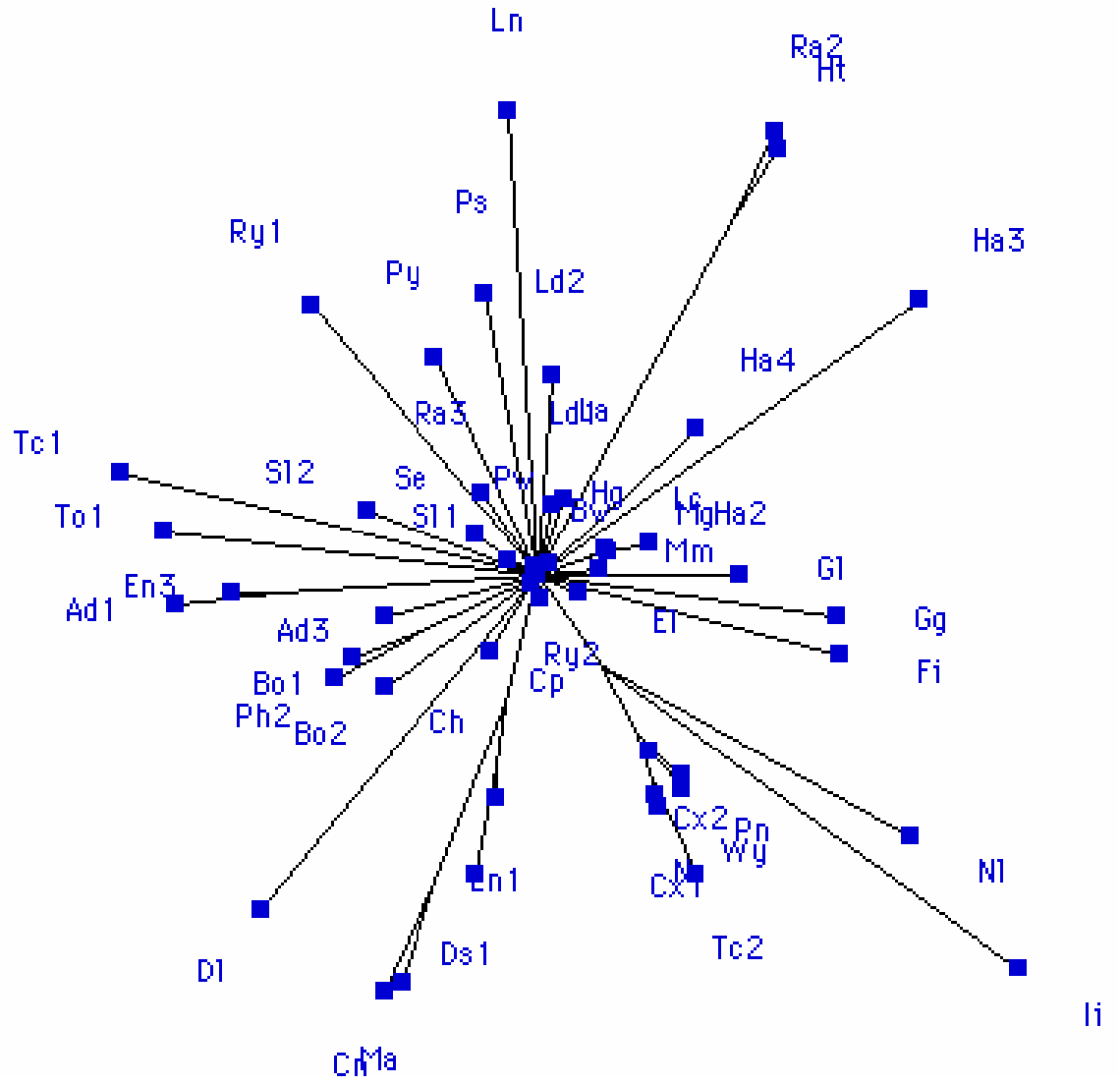
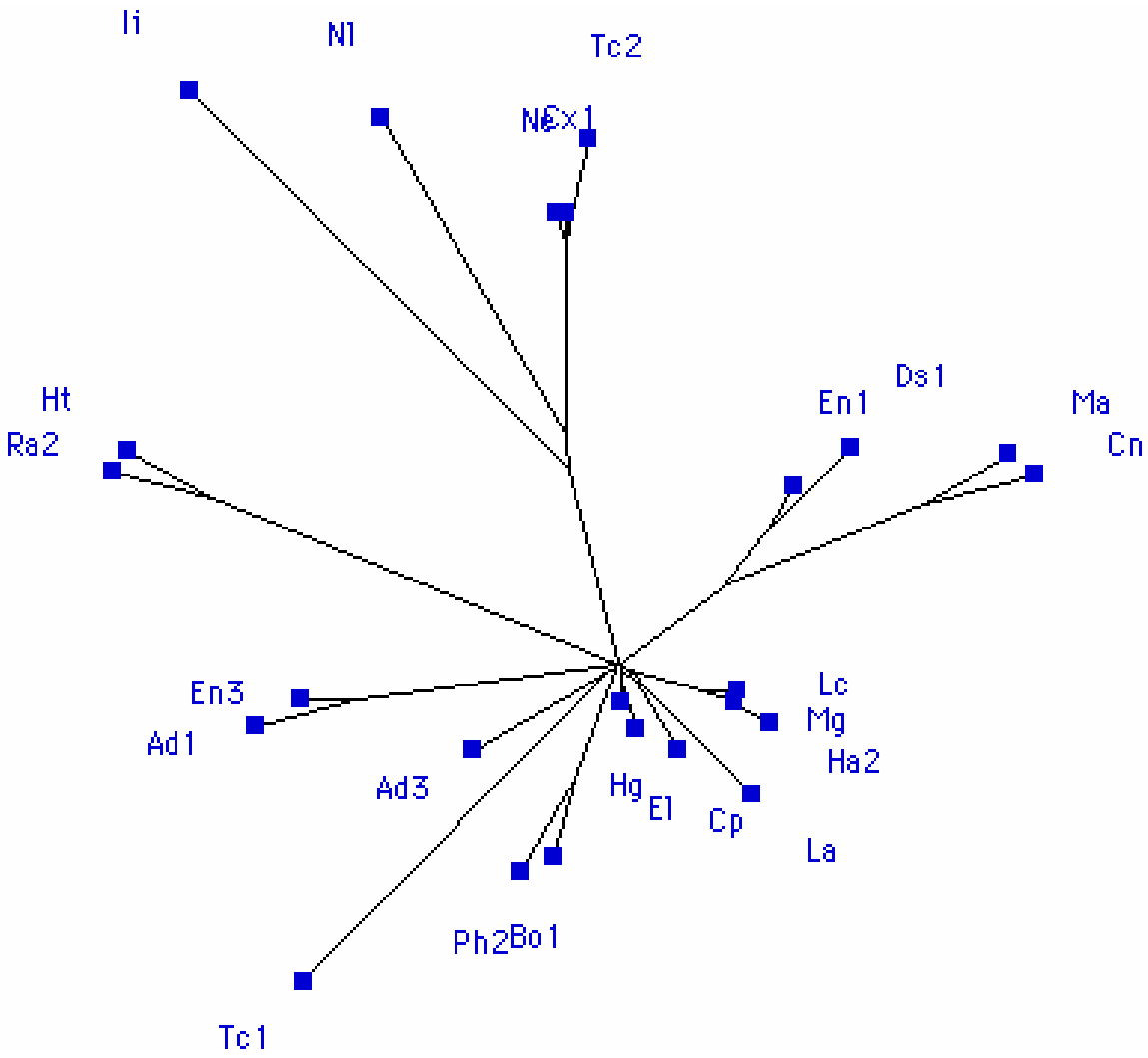


Figure 21: the SplitsTree graph for lines 501-end, in 50 manuscripts

Again, we seek to optimize the tree by removal of manuscripts from the analysis: the manuscripts Bo^2 Bw Ch Cx^2 Dl Fi Gg Gl Ha^3 Ha^4 Ld^1 Ld^2 Ln Mm Pn Ps Pw Py Ra^3 Ry^1 Ry^2 Se Sl^1 Sl^2 To^1 Wy. This gives the following SplitsTree graph (in file GPWS9.vm):


Ii
Nl
Tc2
N&x1
Ht
Ra2
Ds1
En1
Ma
Cn
En3
Ad1
Ad3
Lc
Mg
Ha2
Hg
El
Cp
La
Ph2
Bo1
Tc1

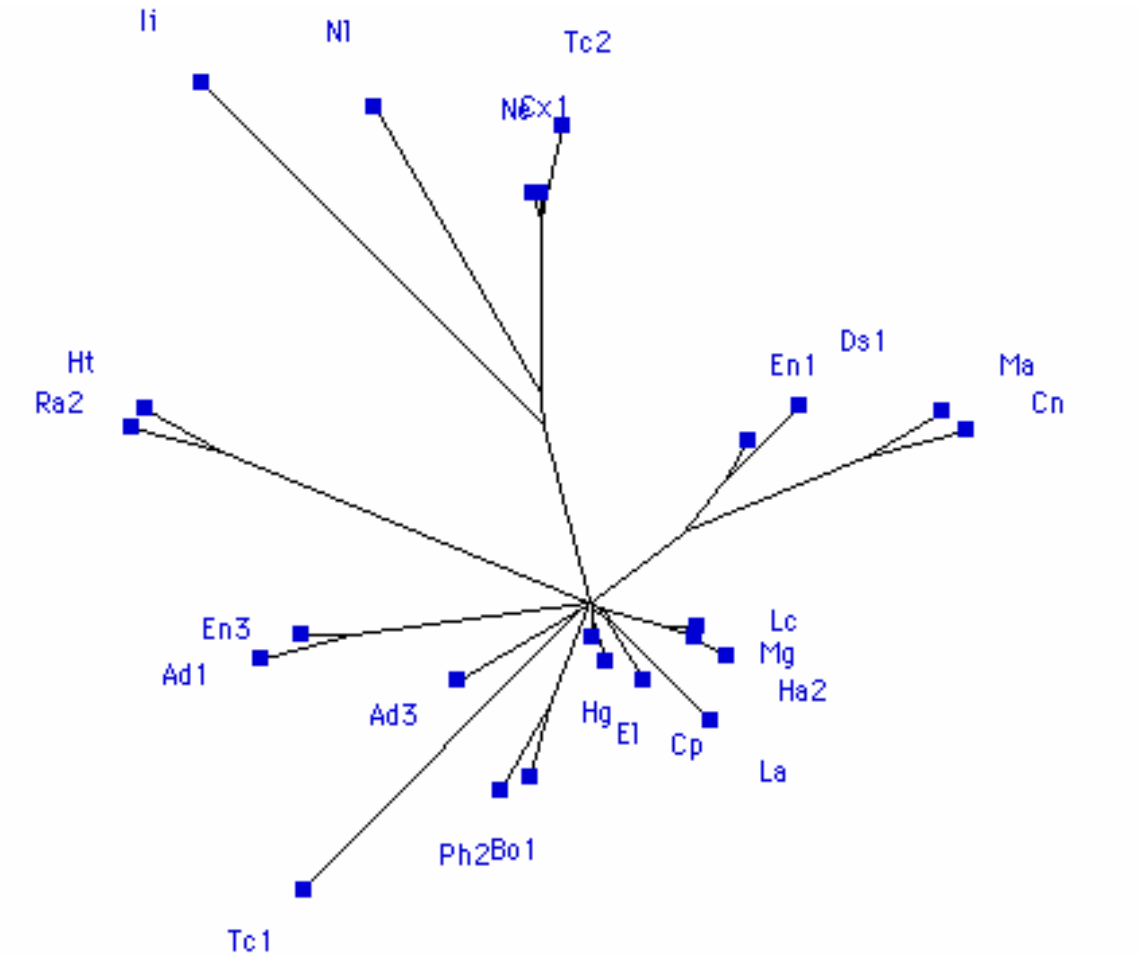
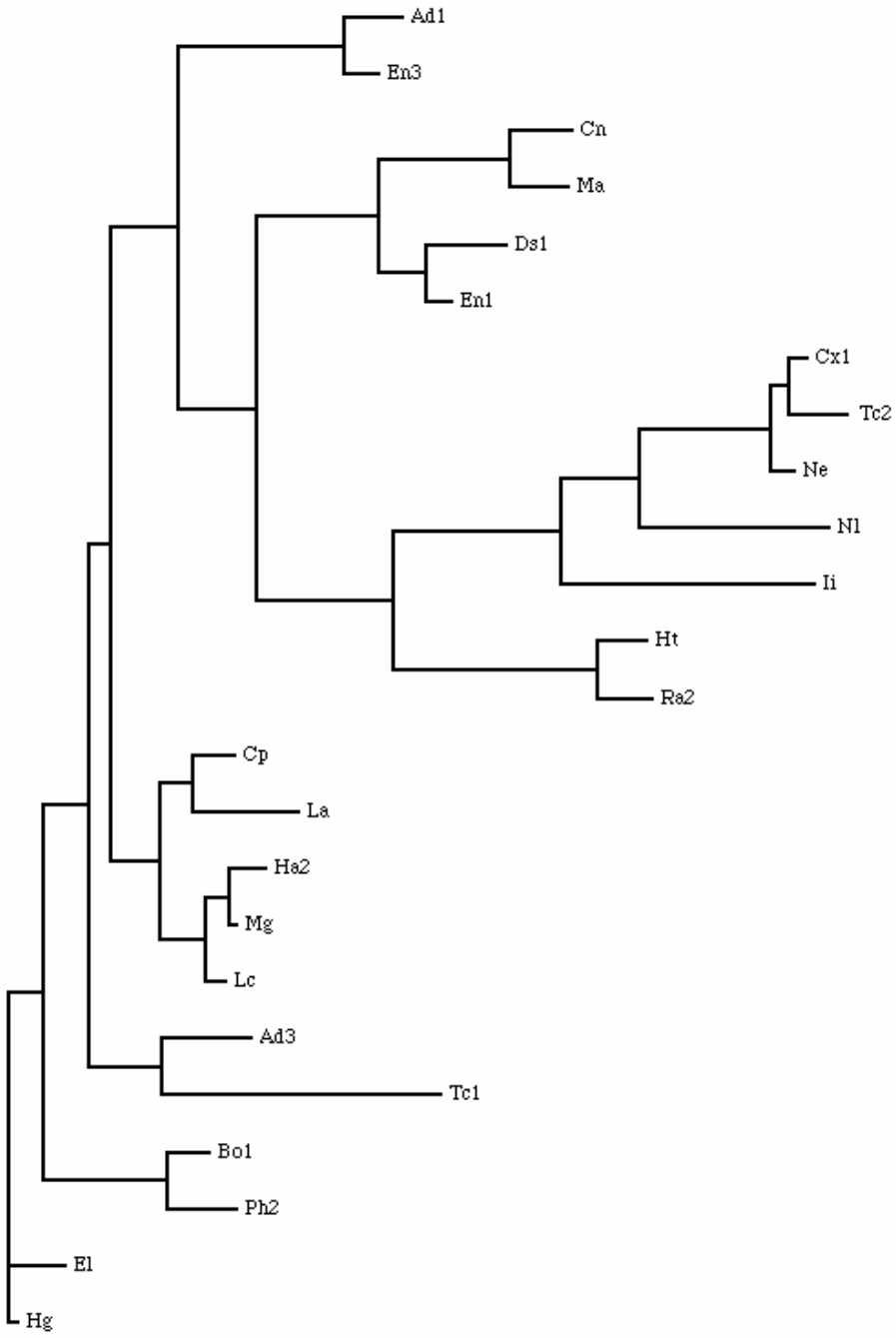


Figure 22: a SplitsTree graph for lines 501-end, giving more detail of manuscript relations

To create a more detailed view of the manuscript relations, exactly the same datafile was given to PAUP for processing (in file 'GPWS9.nex'). PAUP produced the following cladogram:


Ad1
En3
Cn
Ma
Ds1
En1
Cx1
Tc2
Ne
Nl
Ii
Ht
Ra2
Cp
La
Ha2
Mg
Lc
Ad3
Tc1
Bo1
Ph2
El
Hg

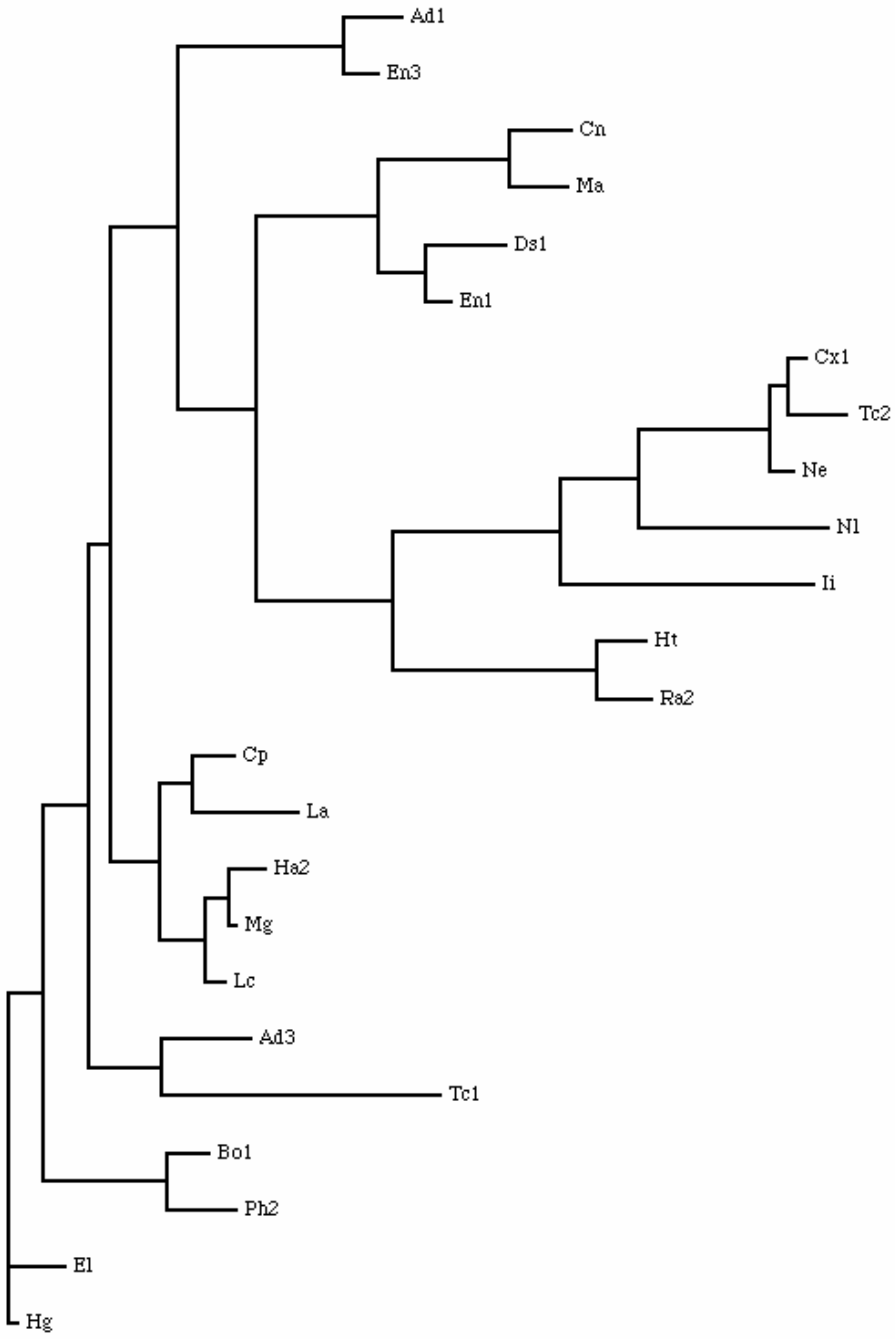

Ad1
En3
Cn
Ma
Ds1
En1
Cx1
Tc2
Ne
Nl
Ii
Ht
Ra2
Cp
La
Ha2
Mg
Lc
Ad3
Tc1
Bo1
Ph2
El
Hg

Figure 23: cladogram for 24 manuscripts, lines 501-end

The SplitsTree graph and PAUP cladogram show, with one exception, the same manuscript groupings for these 350 lines as for the two segments of 250 lines preceding. Thus, in all three segments the Manly-Rickert groups a b cd are distinct, as are the pairs Bo^1/Ph^2 Ad^1/En^3. The difference between this 350 lines and the earlier 500 lines is that in this 350 lines there appears to be no separation between the manuscripts in the top half of the earlier SplitsTree graphs — the ab manuscripts and Ad^1/En^3 Ht Ad^3/Tc^1, which we suggest are descended from alpha — and the manuscripts in the bottom half of the SplitsTree graphs, including Hg El. Thus, Figure 22 shows that both these top and bottom groups derive from a single point, and not from two distinct points, as they derive in Figures 11 and 19.

There are several possible explanations for this difference:

1. The identification of a separate alpha exemplar in the first 500 lines could be mistaken. As you will by now have discovered if you have done the SplitsTree exercises earlier in this workshop, if you include slightly different selections of manuscripts then you can make this and other relationships appear and disappear almost at will. The separation of the tradition into two branches suggested by Figures 11 and 19 could be an illusion. We consider this possibility further in 3.3.4 below.
2. There could indeed have been a separate alpha exemplar, but about line 500 certain manuscripts (Ad^1 Ad^3 En^3 Ht etc.) stopped using that exemplar. Thus: alpha would exist throughout and ab would be descended from it throughout; however the other manuscripts shown by Figures 11 and 19 as descended from it (Ad^1/En^3 Ht Ad^3/Tc^1) move away from alpha to a manuscript nearer to, or identical with, the ancestor of Hg El and the other manuscripts in the bottom half of those figures. This shift would account for the differences between Figures 11 and 19 and Figure 22.
3. However, this would require the distinct ancestors of at least three different groups of manuscripts (Ad^1/En^3 Ad^3/Tc^1 Ht/Ra^2, on this analysis) all to change their exemplars at the same time. This seems very unlikely. One might postulate yet another exemplar, shared by these groups below alpha; but one should not multiply exemplars

unless really necessary. A simpler explanation is this: that the alpha exemplar itself changed character about line 500. Up to that point, one might suppose that alpha introduced some forty changes into its copy of O (about one every twelve lines), and that a significant number of these changes were inherited by its descendants ab Ad[1] /En[3] Ht Ad[3]/Tc[1]. However, after line 500 alpha may have become a slightly more faithful copy of O, and introduced only some fifteen changes (one every twenty lines). Not surprisingly, SplitsTree and PAUP might fail to determine the existence of a separate exemplar from such slender evidence, and this would explain the failure of the two programs to distinguish alpha in lines 500-end.

We will see evidence in favour of this latter hypothesis when we come to work with the variant database. Analysis with this (see 3.2.3) suggests that the alpha group of variants does persist throughout, but that alpha variants are rather less frequent in the second half of the Prologue than in the first. A similar shift in the character of an exemplar may be seen in the behaviour of the cd ancestor. Analysis in The Wife of Bath's Prologue suggested that the cd ancestor introduced some 200 changes into the text; about one every four lines. However, as we will see below it appears that the cd ancestor in The General Prologue introduced only some sixty changes, so that some cd manuscripts (notably, Cp) appear almost identical with O for long stretches of the text.

Exercises: now, try these for yourself..

1. You can go through the processes by which we moved from the tree shown in Figure 21 to that shown in Figure 22, as follows:
   - Load SplitsTree and the file GPWS8.vm. SplitsTree will then give you the tree given if Figure 21.
   - To move from this tree to that given in Figure 22: you need to delete 26 manuscripts ('taxa' in evolutionary biology terms) from the analysis. These manuscripts are: Bo[2] Bw Ch Cx[2] Dl Fi Gg Gl Ha[3] Ha[4] Ld[1] Ld[2] Ln Mm Pn Ps Pw Py Ra[3] Ry[1] Ry[2] Se Sl[1] Sl[2] To[1] Wy. See exercise 1 in 3.1.1 for instructions on how to delete manuscripts from the analysis.

   You should now get the same tree as given in Figure 22, file GPWS9.vm
2. Try and improve the tree by deleting other manuscripts from the analysis, as in exercise 3 of 3.1.1.

3. Figures 11 and 19 suggest a cleavage in the tradition, between the manuscripts in the top (alpha) and bottom (Hg, El) halves of the trees. Can you achieve a similar cleavage, for at least some of the manuscripts for some of the text of lines 500-end? (hint: start with GPWS9.vm, remove Tc^1 and select sites 4000-5147, lines 670-end). What do you notice about the position of Ad^3 in this tree?
4. Run the analysis over different segments of the text, as described in exercise 6 of 3.1.1. For lines 500-end the different segments are:

   Lines 501-550: sites 2850-3168

   Lines 551-600: sites 3169-3520

   Lines 601-650: sites 3521-3888

   Lines 651-700: sites 3889-4197

   Lines 701-750: sites 4198-4509

   Lines 751-800: sites 4510-4790

   Lines 801-850: sites 4791-5147

Compare the different trees produced by the analyses over different segments of text. How are they similar; how are they different? How can you explain the differences?

3.2 Identifying the variants characteristic of each group

The SplitsTree and PAUP analysis described in section 3.1 has suggested the existence of the following manuscript groupings, throughout the General Prologue:

Dd Cn/Ma Ds^1/En^1: Manly and Rickert's a

Cx^1 Cx^2 Ii Ne Pn Wy: Manly and Rickert's b. PAUP analysis (Figures 12, 20, 23) suggested that the groups ab have a single ancestor.

Cp La Mm, with the triplet Ha^2 Mc Lg: Manly and Rickert's cd

In addition: the pairs Ad^3/Tc^1 Ht/Ra^2 Ad^1/En^3 Bo^1/Ph^2 also appear. The first three pairs were classified as O by me in analysis of The Wife of Bath's Prologue, and the pair Bo^1/Ph^2 as e.

According to the account given above, if these groupings are valid, then there will be a group of variants introduced by the common ancestor of each of these groups. The members of each group can then be identified,

by assessing the proportion of the variants introduced by that ancestor present in each witness.

The first step, then, is to isolate the groups of variants introduced by the ancestor of each group: what we call the 'fundamental variant groups'. We do this by taking the manuscripts which the evolutionary biology programs have suggested are members of each group. If they do indeed constitute a group of manuscripts, related by descent from a common ancestor below the archetype, then there will be a group of variants which were introduced by this common ancestor and which have been inherited by the members of this group. These variants will be characteristically found together in significant numbers only in the members of this group, and not elsewhere in the tradition. For example: if the a group (Dd Cn/Ma Ds^1/En^1) really exists, we would expect to find a group of variants which are commonly:

1. found in at least one of Ds^1/En^1, and
2. found in at least one of Cn/Ma, or
3. found in a combination of Dd with any one of the pairs Cn/Ma Ds^1/En^1, and
4. NOT commonly found in manuscripts elsewhere: particularly, not in manuscripts close to the archetype of the whole tradition
5. NOT commonly found in manuscripts representing separate lines of descent (for example, the b group, which must have its own ancestor distinct from the direct ancestor of a; but also the cd group and the pair Bo^1 Ph^2)

To identify such a group by hand, one would have to go through the entire corpus of variants, looking for readings which have variant groups such as 'Cn Ds^1 Ma' and 'Cn Dd Ds^1'. There are over 16000 variant groups in The General Prologue so this would be time-consuming, and probably inaccurate.

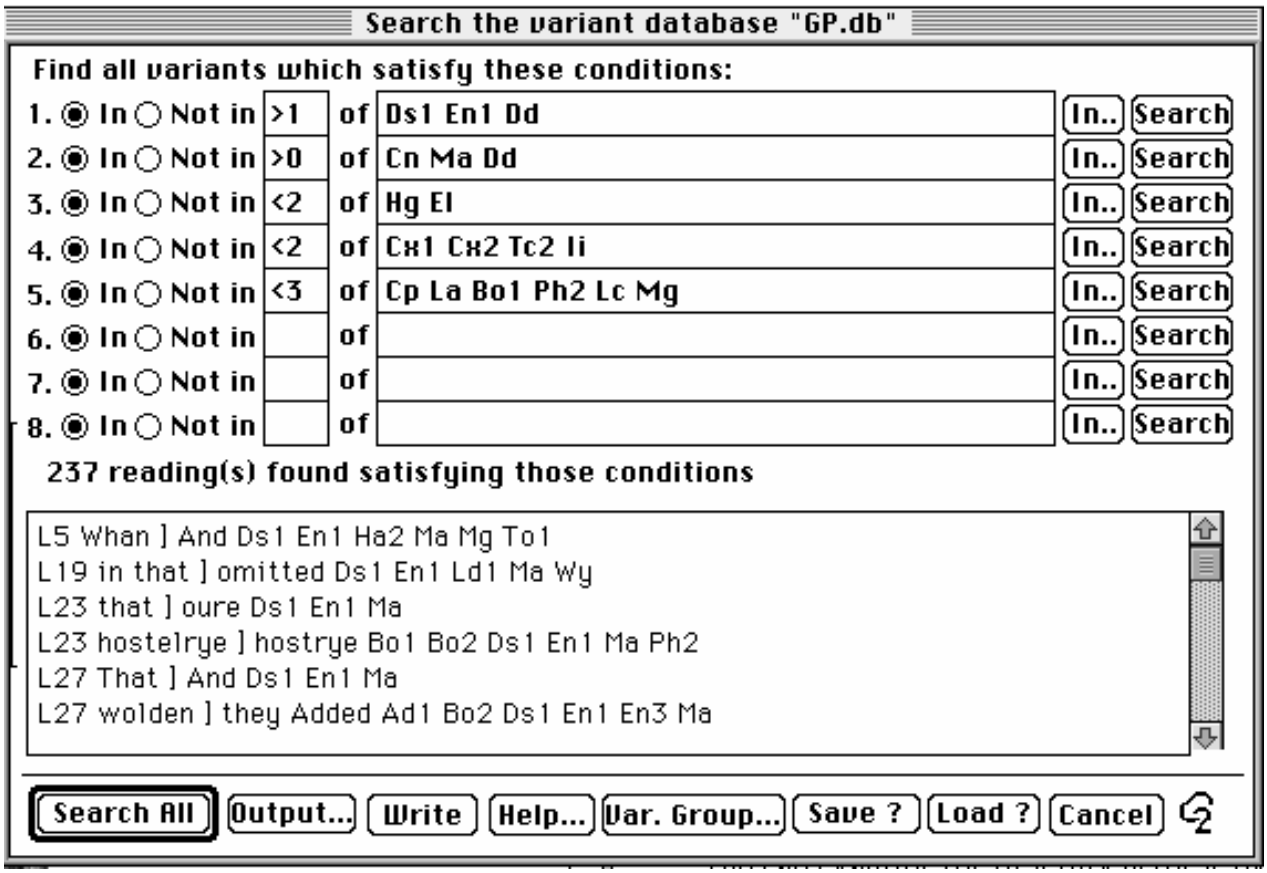
3.2.1 Using the variant database to identify the groups of variants: the a group

Therefore, we use a variant database to carry out such searches. Using *Collate*, we translate all the variants into a database format, and use the application VBase to search for the variants which satisfy such complex queries.

In the last section, I suggested that variants introduced by the common ancestor of the group Cn Dd Ds^1 En^1 Ma could be identified as satisfying five conditions. These can be translated into a VBase search as follows:

1. The variants must be in at least two of Ds^1 En^1 Dd, and
2. The variants must be in at least one of Cn Ma Dd (both Cn and Dd are defective at the beginning, hence the recasting of the query), and
3. Taking Hg and El as nearest the archetype of the whole tradition: the variants must not be in both Hg and El, and
4. Not in more than one b group manuscript, and
5. Not in more than two manuscripts from groups cd and the pair Bo^1 Ph^2

To see how VBase does such a search, click here. The VBase application should appear, load the variant database 'GP.db', and present the database search screen with the query already loaded, as follows (the exact appearance of the screen may differ, depending on your computer platform and the version of VBase installed):


Search the variant database "GP.db"
Find all variants which satisfy these conditions:
1. In Not in >1 of Ds1 En1 Dd In.. Search
2. In Not in >0 of Cn Ma Dd In.. Search
3. In Not in <2 of Hg El In.. Search
4. In Not in <2 of Cx1 Cx2 Tc2 Ii In.. Search
5. In Not in <3 of Cp La Bo1 Ph2 Lc Mg In.. Search
6. In Not in of In.. Search
7. In Not in of In.. Search
8. In Not in of In.. Search
237 reading(s) found satisfying those conditions
L5 Whan] And Ds1 En1 Ha2 Ma Mg To1
L19 in that] omitted Ds1 En1 Ld1 Ma Wy
L23 that] oure Ds1 En1 Ma
L23 hostelrye] hostrye Bo1 Bo2 Ds1 En1 Ma Ph2
L27 That] And Ds1 En1 Ma
L27 wolden] they Added Ad1 Bo2 Ds1 En1 En3 Ma
Search All
Output...
Write
Help...
Var. Group...
Save ?
Load ?
Cancel

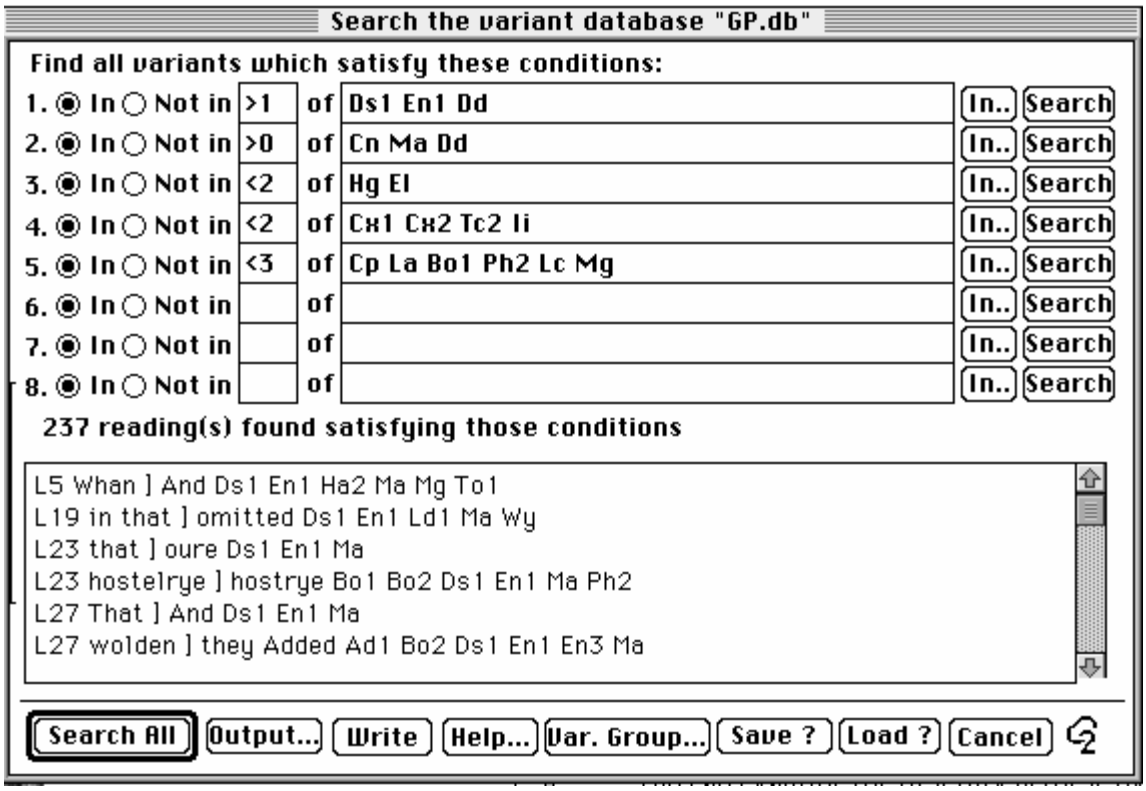
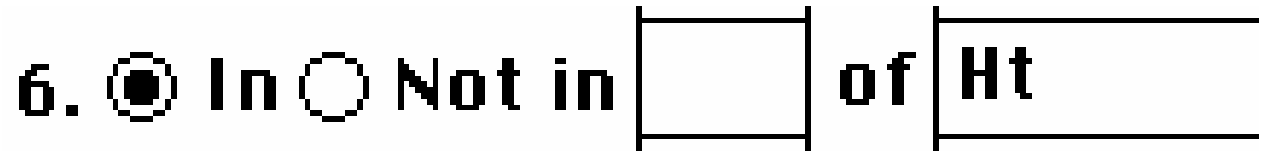
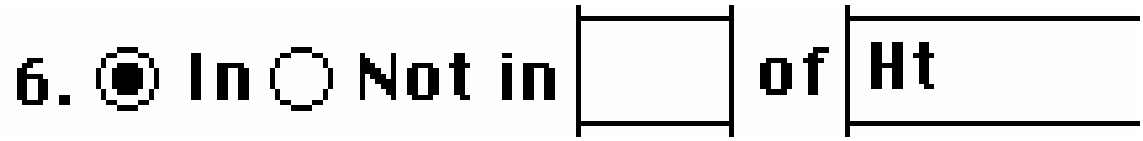


Figure 24: the VBase search dialogue

The five lines of the query correspond to the five conditions set out above, as having to be satisfied by a variant characteristically found only in these manuscripts. This query has returned 237 readings. You can examine these by scrolling through the box at the base of the screen. You can save them to a window by clicking on the **Write** button. This will write all the variants to a window; from there you can save the variants to a file.

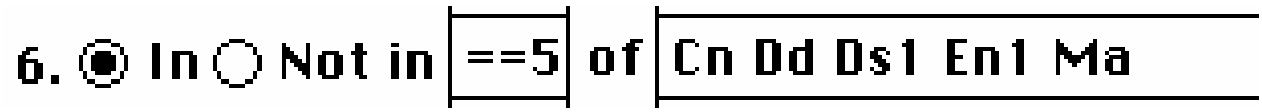
You can edit the variant search conditions, and run the search yourself. You could reduce the number of 'outside' manuscripts the variant might be in, by changing the '<2' in line 3 of the query to '<1', and similarly in lines 4 and 5, and then press the **Search All** button to run the new search. Or, you could add another line to the query. For example, you could discover how many of these variants are in Ht by typing Ht into the sixth line of the query, so this appears as follows:

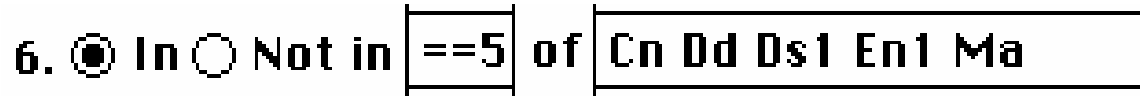




and press **Search All** again.

You can also do a search on a single condition by pressing the **Search** button to the right of that query. For example, you could search for all variants in the five Cn Ds[1] Dd En[1] Ma, and only in that five, by typing into a search box as follows:



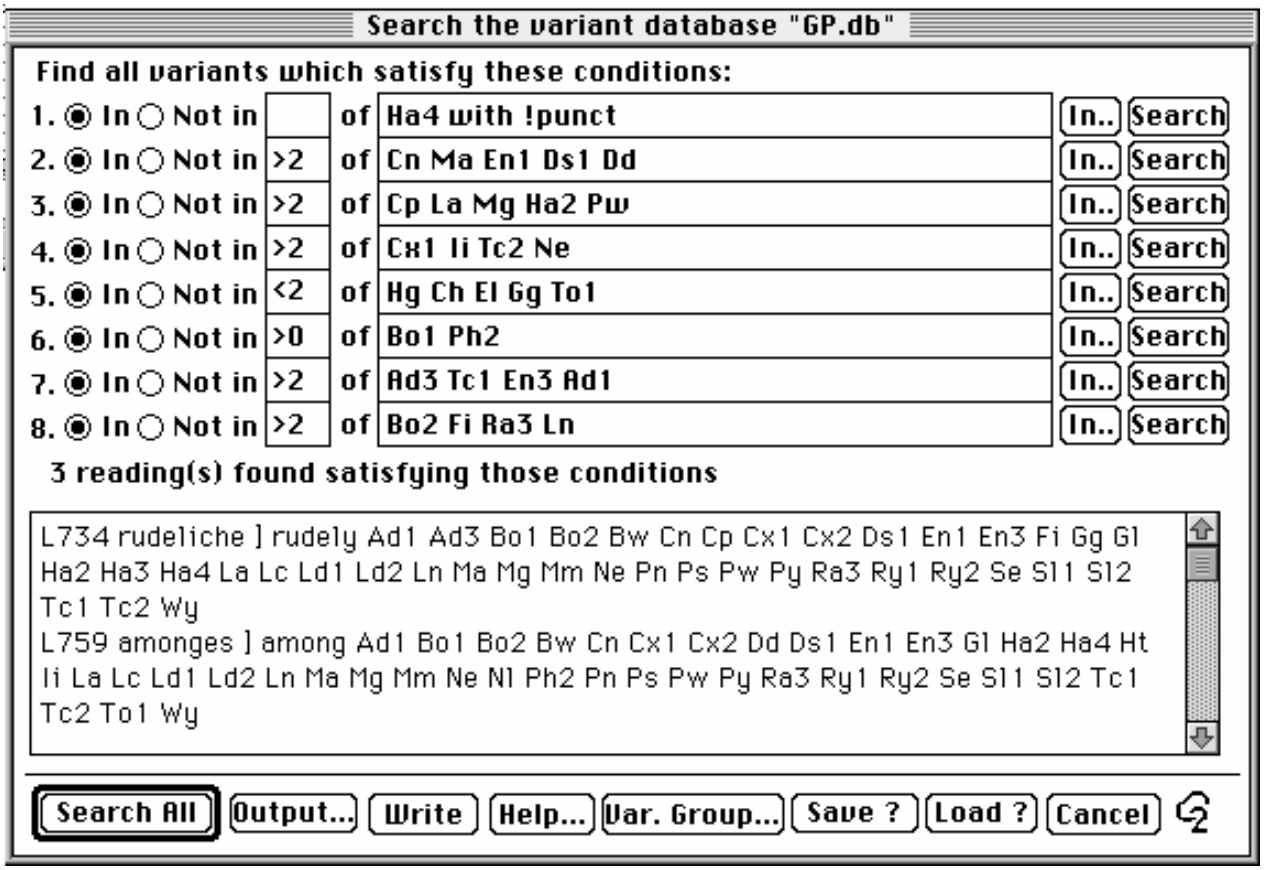


Pressing the **Search** button beside this query should return 35 hits.

The number of variants returned by this search for the a variants — over 200 in all — suggests that this is indeed a distinct family group, and that there was a separate a ancestor which introduced over 200 variants into the tradition. Significant numbers of these a variants are (of course) present in the five manuscripts Cn Dd Ds[1] En[1] Ma. But some other manuscripts also have a number of these. We will discuss later the possible significance of these occurrences of a variants in 'outside' manuscripts.

It may be objected that this procedure appears circular. We think that manuscripts XYZ may form a group; we search then for readings found characteristically only in XYZ; and then we declare that XYZ are members of this group because they have these readings. In fact, it is not circular for a simple reason. If the group XYZ does not in fact exist, then the search for readings characteristically found only in XYZ will fail to find a significant number of readings.

One can illustrate this with respect to the suggestion by Manly and Rickert (II 78) that all the manuscripts of The General Prologue except six (Hg El Ch Gg Do To^{1}) are derived from a single common ancestor. This 'composite group', according to Manly and Rickert, is made up the separate sub-groups a b cd with Ha^{4}; it also contains Bo^{2} Fi Ra^{3} Ln; it must (if Manly and Rickert are correct) also contain the pairs Bo^{1}/Ph^{2} Ad ^{1}/En^{3} Ad^{3}/Tc^{1}; it does not contain Hg Ch El. Manly and Rickert's attempts to prove the existence of this group, and thus that all these had a distinct exclusive common ancestor, are rather tortuous. With VBase, one can search for evidence supporting this hypothesis. The query MRComp.cdb loads this query:


Search the variant database "GP.db"
Find all variants which satisfy these conditions:
1. In Not in of Ha4 with !punct In.. Search
2. In Not in >2 of Cn Ma En1 Ds1 Dd In.. Search
3. In Not in >2 of Cp La Mg Ha2 Pw In.. Search
4. In Not in >2 of Cx1 Ii Tc2 Ne In.. Search
5. In Not in <2 of Hg Ch El Gg To1 In.. Search
6. In Not in >0 of Bo1 Ph2 In.. Search
7. In Not in >2 of Ad3 Tc1 En3 Ad1 In.. Search
8. In Not in >2 of Bo2 Fi Ra3 Ln In.. Search
3 reading(s) found satisfying those conditions
L734 rudeliche] rudely Ad1 Ad3 Bo1 Bo2 Bw Cn Cp Cx1 Cx2 Ds1 En1 En3 Fi Gg Gl Ha2 Ha3 Ha4 La Lc Ld1 Ld2 Ln Ma Mg Mm Ne Pn Ps Pw Py Ra3 Ry1 Ry2 Se Sl1 Sl2 Tc1 Tc2 Wy
L759 amonges] among Ad1 Bo1 Bo2 Bw Cn Cx1 Cx2 Dd Ds1 En1 En3 Gl Ha2 Ha4 Ht Ii La Lc Ld1 Ld2 Ln Ma Mg Mm Ne Nl Ph2 Pn Ps Pw Py Ra3 Ry1 Ry2 Se Sl1 Sl2 Tc1 Tc2 To1 Wy
Search All
Output...
Write
Help...
Var. Group...
Save ?
Load ?
Cancel

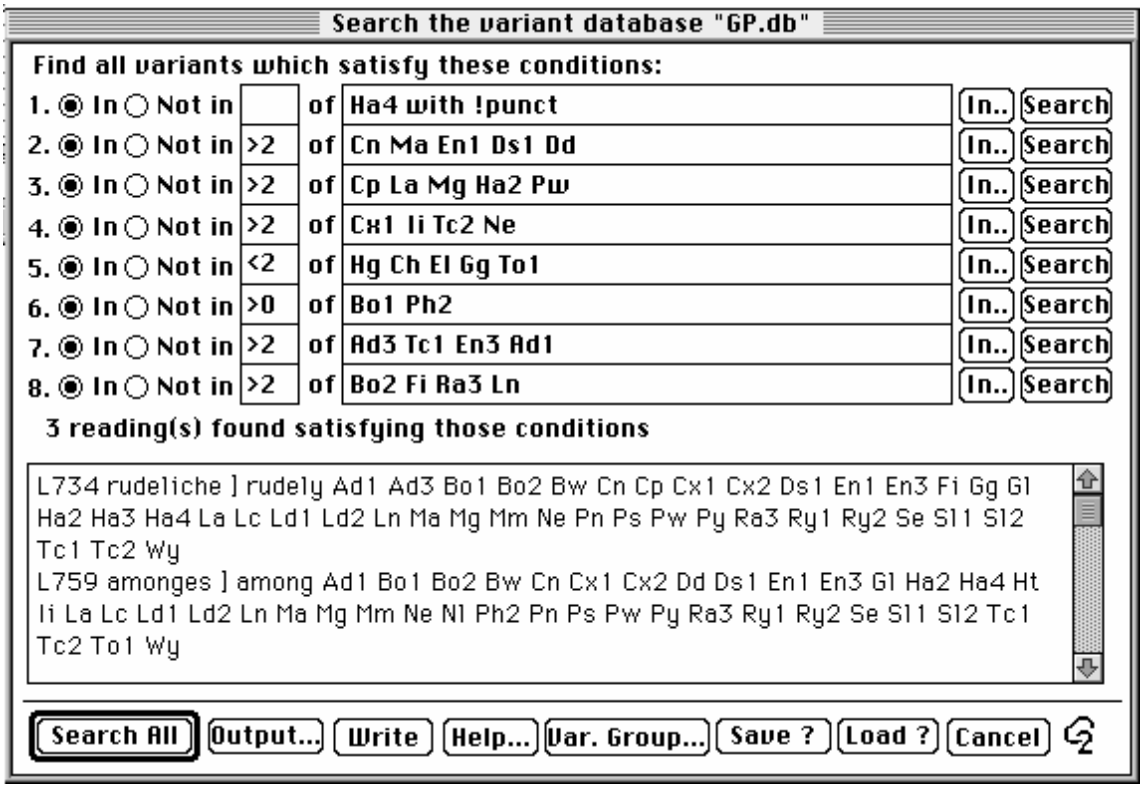


Figure 25: searching for Manly and Rickert's 'large composite group'

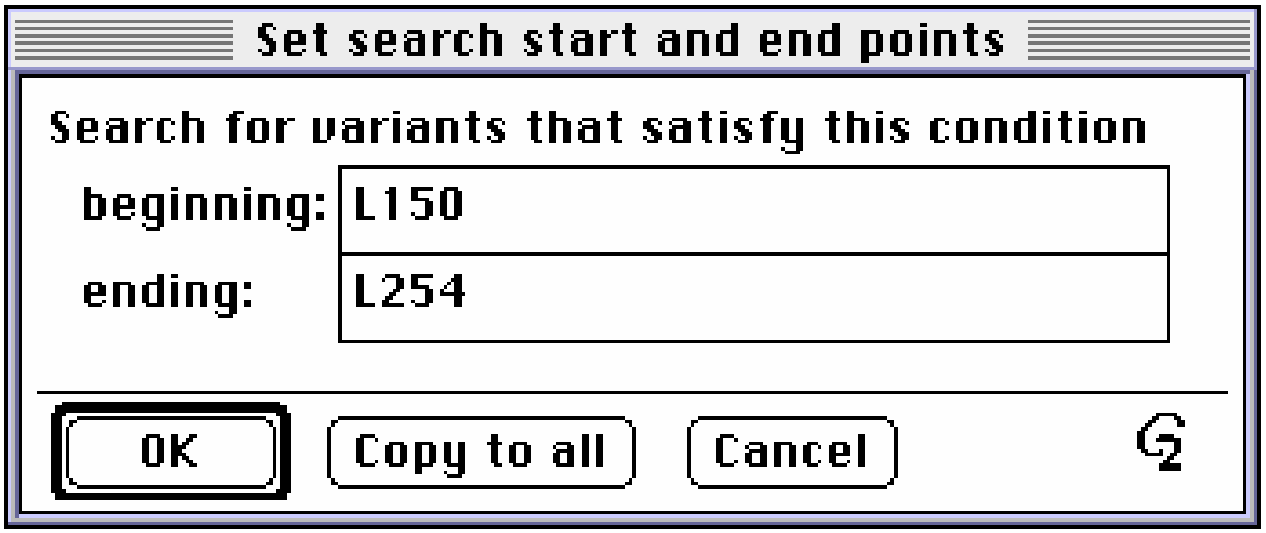
This search produces just three readings. So small a number, and the unpersuasive character of the readings themselves (which would not exist as variants at all if we had chosen to regularize rudely/rudeliche amonges/among ferrer/ferther, as we did consider doing) suggests that there is no such 'composite group'. Indeed, if one alters the fifth line of the query, so that none of Hg El Ch Gg To1 should have a variant belonging to this group (as one might) then the search finds no readings whatever. There is no evidence for Manly and Rickert's assertion that all these manuscripts had a single common ancestor. Hence their assertion, which is the basis of their whole analysis of this part of the *Tales*, that the tradition of The General Prologue divides into two (Hg El Ch Gg To1, and this 'composite group') is false.

This failure to find a group where none exists should give confidence that where our methods do find a group, it does indeed exist. There is a

further method of validating the existence of a hypothetical group. If indeed it does exist, then 'counter-groups' should not exist. That is, one should not be able to find evidence for other groups whose composition would contradict the hypothetical group. Thus, for this a group one should not find a group made up of some members of a, some members of cd (e.g. Cp La Bo^1 Ph^2 Lc Mg Ha^2), Bo^1/Ph^2, but excluding other members of a. The queries GPnota1.cdb and Gpnota2.cdb attempt to find such countergroups, and fail. Try these for yourself: they return just three and two readings each. Of these five readings, three are found in Hg (though not El) and are therefore likely to be ancestral to the whole tradition, and so cannot be taken as evidence of anything. This leaves just one reading only in each 'counter-group', opposed to the 200-plus in the a group itself.

Exercises: now, try these for yourself..

1. After loading Gpavars.cdb, try further combinations of the search conditions. Use the **In** and **Not in** buttons to include or exclude manuscripts, singly or in groups. Use the **In..** button, to the right of the box, to set a range of lines for the search. Note that you must precede the line number with L. Thus, typing the following values into the dialogue box will have VBase search only between lines 150 and 254. The **Copy to all** button will have the range apply to all the search conditions.



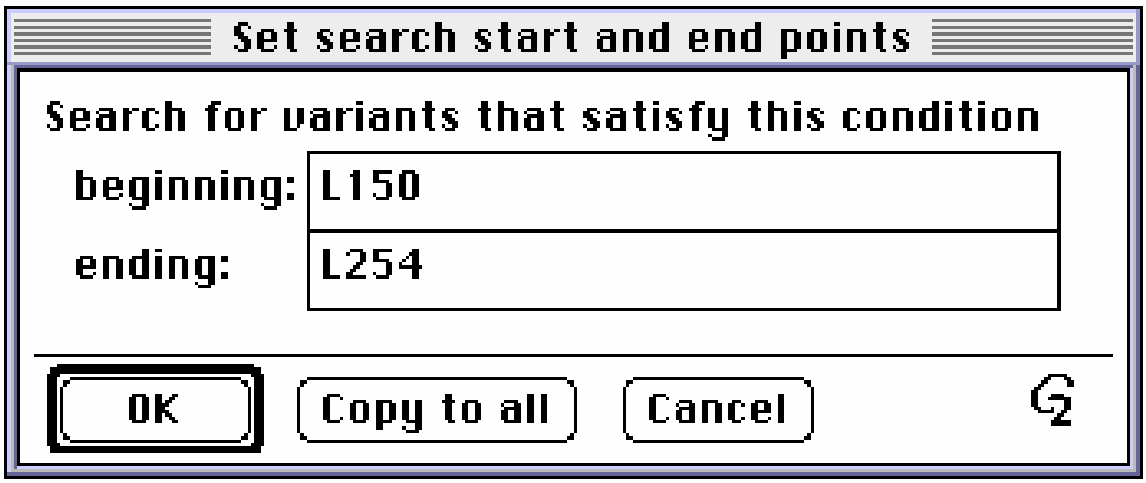


Figure 26: the start and end points dialogue

Note that the keywords '\start' and '\end' have the search begin and end at the first and last variant respectively (you can achieve the same result by typing L1 and L858.)

2. Load the file Mrcomp.cdb. You will see that the first line is **Ha[4] with !punct**. In the database, the type of every variant is marked, as one of:

   missblk blkorder: missing or differently ordered lines

   agreeword: agreements with the base

   repword omword addword: variants on single words

   repphr omphr: variants on phrases

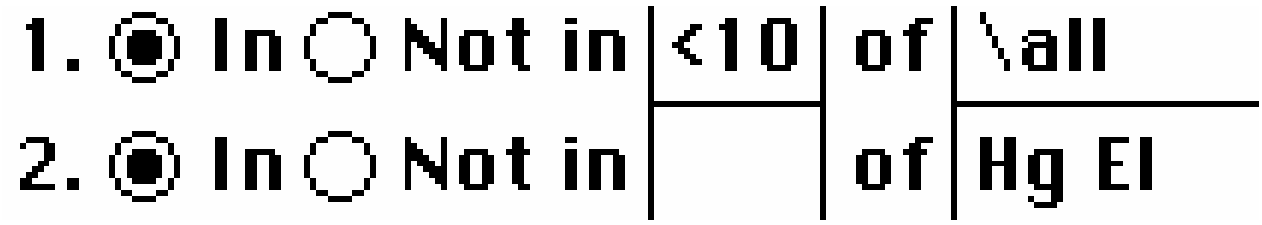
   reppunct addpunct ompunct: variants on punctuation (virgules, paraphs, punctus elevatus)

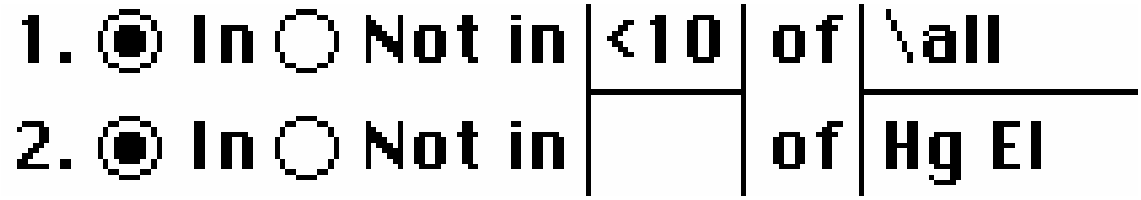
Thus, adding the qualification 'with repword' would search only for variants which are replacements of single words. The qualification ' with !repword' would search for variants which are NOT replacements of single words. VBase also searches for combinations of letters within the variant types. Thus, the qualification 'with rep' will look for all instances of replacements, whether in words, phrases, or punctuation. Similarly, the qualification 'with !punct', as here, will exclude all punctuation variants from the search.

Try various combinations of these searches for yourself. You can discover (for example) that El has 1010 punctuation variants, by typing 'El with punct' into the text box and pressing the **Search** button to the right.

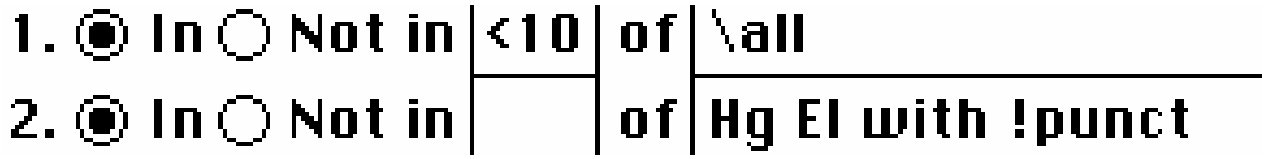
3. A 'Help' file is provided with VBase, and is accessible through pressing the **Help...** button on the Search Database dialogue. This gives details of other features of VBase. A particularly useful facility is the keyword '\all'. Typed in the manuscript specification box, this specifies that a particular reading should be found in all manuscripts (note that the type qualification feature does not work with the \all keyword.) In combination with the number specification box, one can use this to find (for example) all readings in more than thirty manuscripts:

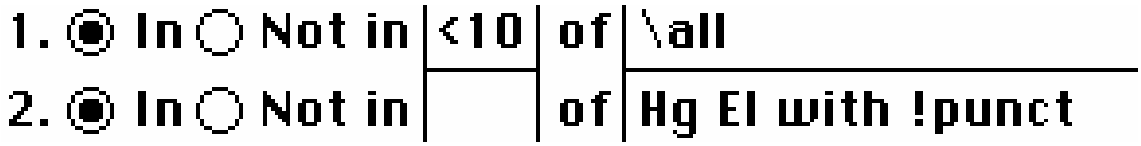
Explore this for yourself. For example, you could find all variants found in both El and Hg and found in more than thirty manuscripts (hint: type El Hg into the second query line, with the above query as the first, and press **Search All**.) This query returns over 6000 hits: basically, these are the readings of the archetype, preserved in El Hg and everywhere else. However, if we ask for the readings in El Hg and in less than 10 manuscripts of \all, as follows, matters get more interesting:





This returns 550 queries, and a glance at them shows that almost all of them are punctuation variants. This is a reflection of the special punctuation system used by the El/Hg scribe (see Solopova forthcoming). We can exclude the punctuation variants, as follows:





This query returns just seventeen hits. This is remarkably few, given the closeness of Hg and El throughout. You might like to consider why there are so few variants found only in El Hg and less than ten manuscripts. This is discussed further in 3.4.1 below.

4. A crucial question is the nature of the variants in each group. If Chaucer did indeed revise the text of The General Prologue, one would expect that one of the fundamental manuscript groups might

descend from this revision. Therefore, the variants characteristic of that group would be those introduced by Chaucer in that revision. Accordingly, when I carried out the analysis of The Wife of Bath's Prologue, I studied each group of variants to assess whether they were scribal in nature (and so likely to have been introduced not by Chaucer, but by a scribe) or authorial in character (and so likely to have been introduced by Chaucer himself). Only one of the groups of variants in The Wife of Bath's Prologue appears to have been authorial in character: those found characteristically in what I called the O manuscripts (Robinson 1997). It appears these variants were in **O** itself and so were ancestral to the whole tradition, and thus not evidence of revision but of the closeness of these manuscripts to the archetype.

Look more closely at some of the a variants found by VBase: for example the replacement of Whan by And in Whan Zephirus eek with his sweete breeth 5; the omission of in that in Bifel that in that sesoun on a day 19; the replacement of that by oure in At nyght was come / into that hostelrye 23. Do you think it likely that Chaucer was responsible for these changes?

With over fifty manuscripts, sixteen thousand variants, and many different combinations of variant type and range of lines, there is no limit to the number of searches you could run through VBase. Be warned: use of VBase is highly addictive.

3.2.2 Using VBase to identify other fundamental groups: b cd ab e

We now use VBase to identify the groups of variants characteristic of each group of manuscripts suggested by SplitsTree and PAUP, in the same way we used it to identify the a group.

The b group of variants is identified by this query (in GPbvars.cdb)

in >2 of Cx^{1} Tc^{2} Ii with !punct (the b group itself)

AND in <4 of Cp La Mm Ch Lc Mg Ha^{2} Hg El Ad^{3} Ad^{1} Bo^{1} Ph^{2} (cd and the archetype)

AND in <2 of Cn Ma Ds^{1} En^{1} Dd (the a group)

This query returns 215 hits. On this evidence, the b archetype was as distinctive as the a archetype, introducing a new reading about every four lines.

The cd group of variants is identified by this query (in GPcdvars.cdb)

in >2 of Cp La Pw Mm with !punct (the cd group itself)
AND in <2 of Ad^3 Ad^1 Bo^1 Ch El Hg Ph^2 (the archetype)
AND in <2 of Cn Ma Ds^1 En^1 Dd Cx^1 Cx^2 (ab)

This query returns 48 hits. On this evidence, the cd archetype was considerably less distinctive than the a or b archetypes. It introduced a new variant only every seventeen lines or so. Interestingly, this is a considerably lower frequency of variation than shown by the cd archetype in The Wife of Bath's Prologue, where there is a distinct cd variant about every four lines. Thus, the character of the cd ancestor changed between The General Prologue and The Wife of Bath's Prologue: the cd ancestor is considerably closer to **O** in the first than it is in the latter. This supports, by analogy, the hypothesis offered in 3.1.3, that the alpha exemplar changed character around line 500 of The General Prologue, introducing less variation after 500 than before.

PAUP (though not SplitsTree) suggested that groups a and b descend from a single ancestor. This was also Manly and Rickert's view. VBase was used to identify the variants which might have been present in this ancestor of ab. This is the ab group, and it is identified by this query (in GPabvars.cdb):

in >1 of En^1 Ds^1 Dd with !punct (in branch of a)
AND in >0 of Cn Ma Dd (in other branch of a)
AND in >2 of Cx^1 Tc^2 Ii Cx^2 Ne (b)
AND in <4 of Ad^1 Ad^3 En^3 Tc^1 Cp La Mm Ch Lc Mg Ha^2 Bo^1 Ph^2 En^3 (cd; other groupings likely to be distinct, thus Bo^1/Ph^2 Ad^1/En^3 Ad^3 /Tc^1)
AND NOT in Hg El (the archetype)

This query returned 67 hits. This suggests that Manly and Rickert, and PAUP, are correct in asserting that a and b share a common ancestor, ab, below the archetype for the whole tradition. This ab archetype introduced approximately the same number of variants as the cd archetype, at the rate of one approximately every twelve lines.

In the analysis of The Wife of Bath's Tradition, I identified a manuscript

group which I called e. This group contained the manuscripts Bo^1 Ph^2 Gg Si. Si does not have the General Prologue and Gg is very close to **O** in The General Prologue. However, Bo^1/Ph^2 are a strongly marked pair and their variants, constituting what remains of the e group, are likely to be of interest. The e group of variants is identified by this query (in GPevars.cdb):

in <2 of En^1 Ds^1 Dd (not in branch of a)
AND in <2 of Cn Ma Dd (not in other branch of a)
AND in <2 of Cx^1 Tc^2 Ii Cx^2 Ne (not in b)
AND in >1 of Bo^1 Ph^2 with !punct (in e)
AND in <2 of Ad^1 Ad^3 En^3 Tc^1 Ht (not in alpha)
AND in <3 of Hg El Ch Cp La Mm Lc Mg Ha^2 (not in archetype, cd)
AND in >2 of \all (in other manuscripts beside Bo^1/Ph^2)

The last condition was introduced because of the large number of unique variants found only in Bo^1/Ph^2. Because these variants are unique, they are of no help in classifying other manuscripts. They might however be useful in determining the nature of the e common exemplar, and you can identify them through the query in GPeonly.cdb

Exercises: now, try these for yourself..

1. For each, or any one, of the queries identifying the a b ab cd groups, try altering the numbers of manuscripts sought in parts of the queries, and the composition of the manuscript groups themselves. For example, in the fourth line of the query GPabvars.cdb, alter the number specification to '<3' or to '<5' from '<4'. Notice how the number of variants found alters as you vary the query. You could also add a query line '<x' of '\all', as a further way of limiting the query: restricting the ab search by having it look for variants in '<20' of '\all' reduces the number of variants found to 53 from 67.

   From this, you can appreciate that determining the exact composition of the variants characteristic of each group is something of an art. If the query is framed too restrictively, you will not identify variants which are indeed part of the group. For example: seeking variants only in the manuscripts Cn Dd Ds^1 En^1 Ma, and in all these manuscripts, when seeking the a group brings back just 35 readings:

see GPaonly.cdb. However, if the query is framed too broadly, many variants will be returned which are ancestral to the whole tradition, or which are the result of accidental agreement only.

2. For each, or any one, of the queries identifying the a b ab cd groups, seek to construct a 'counter-group': a group whose existence contradicts any one of the groups a b ab cd. Can you do this? (Warning: variants likely to have been present in the archetype are NOT evidence of a counter-group, or indeed of any grouping whatever below the archetype of the whole tradition.)
3. Manly and Rickert, at various points, argue for the existence of c and d as distinct groups. Broadly: they argue that only Cp La Sl^2 are c, and all other cd manuscripts are properly d. Thus, among the manuscripts here seen as forming the core of cd, Mm Pw and the triplet Ha^2 Lc Mg are (according to Manly and Rickert) d not c. Using VBase, can you find any evidence of the separate existence of c and d? The files GPconly.cdb and GPdonly.cdb contain my attempt to do this: see if you can do better. The evidence for the division of c and d is rather ambiguous. Although the query GPdonly.cdb returns some 105 hits, study of these shows that these are made up almost entirely of variants present in the triplet Ha^2 Lc Mg: 89 of the 105 variants are present in at least two of these three (GPcdLc.cdb). It is possible that what Manly and Rickert see as a separate group d is really only the result of a scattering of the variants introduced in the cd ancestor, with Ha^2 Lc Mg forming a subgroup within cd. Thus, the trio Cp La Sl^2 only appear separate from the rest of the large cd group because they are closer to the ancestor of cd, and so preserve variants lost in later descendants of cd: these are the variants found by GPconly.cdb

---

3.2.3 Using VBase to identify the alpha group

Perhaps the single most important hypothesis advanced by the Project's work on the *Canterbury Tales* tradition is our argument for the existence of a key early copy of the whole *Tales*. Following Dan Mosser's work on the tradition, we call this copy alpha. Mosser suggested that alpha was the ultimate exemplar of the ab groups. Manly and Rickert themselves at various points suggest the existence of such an exemplar (for example, in their suggestion that El in the second half of The Wife of Bath's Prologue used the ancestor of ab) but were not able to explore this further. Alpha

also appears to have been very close to the 'better manuscript' used by Caxton in preparing his second edition of the *Tales*. It also appears to have had the *Tales* in the a order familiar from El and most modern editions. I argue elsewhere (following Cooper and Benson; Robinson forthcoming) that this order was Chaucer's own order, and so that alpha was a copy made before the fragments which made up **O**, Chaucer's working draft, became disordered. Thus, alpha was an earlier copy than Hg or its ancestor, which was a copy made after the fragments had become disordered.

In my analysis of The Wife of Bath's tradition I suggested that alpha contained the so-called 'added passages', but did not contain many features of ab (notably, the renumbering of the husbands, and many other readings) which were introduced by the descendant of alpha which became, in turn, the ab ancestor. This argues that alpha and the ab ancestor are distinct, with alpha standing very close to **O** and ab being descended from a copy of alpha and introducing many new variants. Is this also the case for The General Prologue? If so, we should be able to find a group of manuscripts which descend from alpha and not from ab. I argue that El and Cx^2 in the second half of The Wife of Bath's Prologue have the 'added passages' from alpha and not from the ab ancestor; this is probably also true of both Ad^3 and Gg which appear to have links to alpha throughout.

The SplitsTree and PAUP analyses of the first 500 lines of The General Prologue suggest that five manuscripts Ad^1/En^3 Ad^3/Tc^1 Ht may derive from an ancestor lying between **O** and ab: alpha. Accordingly, we should be able to use VBase to determine the variants likely to have been introduced in the common ancestor of these five manuscripts and the ab group: alpha itself. The alpha group of variants is identified by this query (in file GPalpha.cdb):

in >1 of En^1 Ds^1 Dd with !punct (branch of a)
AND in >0 of Cn Ma Dd (other branch of a)
AND in >2 of Cx^1 Tc^2 Ii Cx^2 Ne (b)
AND in <3 of Cp La Mm Ch Lc Mg Hg El Ha^2 Bo^1 Ph^2 En^3 (archetype; cd and others)
AND in >1 of Ad^1 Ad^3 En^3 Tc^1 Ht (the alpha manuscripts)

This query returns 35 readings. This makes alpha a considerably less

distinctive group than the other fundamental groups identified in this analysis: we should expect this, given the closeness of alpha to **O**.

The importance of the existence of this group (and indeed, the question whether this really does constitute a group at all) can hardly be overstated for our analysis. If alpha is shown to have existed for both The General Prologue and The Wife of Bath's Prologue, then we have established that there was a very early exemplar (possibly, the earliest of all) in existence for both these widely-separated parts of the *Tales*. If so, then this would be strong evidence for both these parts of the *Tales* having the same underlying textual history. This, in turn, would argue against the long-held belief that the *Tales* circulated in separate fragments, possibly during Chaucer's lifetime.

This alpha group is discussed further in 3.3.4 below.

Exercises: now, try these for yourself..

1. Considerable effort has gone into framing the query in GPalpha.cdb which produced the alpha group of variants. Try, for yourself, to refine this further, by altering the number of manuscripts sought for each part of the query, and the composition of the manuscript groupings themselves.
2. I have sought to validate the existence of alpha by looking for counter-groups: groups whose existence cannot be consistent with alpha. One such group would be the agreement of manuscripts from the ab group with manuscripts from the cd group against any of the five manuscripts Ad^1 Ad^3 En^3 Tc^1 Ht. This query, in GPabc.cdb, returned just one variant. Another such group would be the agreement of manuscripts from the ab group with the pair Bo^1/Ph^2 against any of the five manuscripts Ad^1 Ad^3 En^3 Tc^1 Ht. This query, in GPabBo1.cdb returned just three variants. Try, for yourself, to construct other counter-groups. (Warning, again: variants likely to have been present in the archetype are NOT evidence of a counter-group, or indeed of any grouping whatever below the archetype of the whole tradition.)

3.3 Using Variant Group Profiles to classify manuscripts

The identification of the fundamental groups of variants apparently introduced at key points within the tradition permits us to proceed to the next stage of analysis. As described earlier, in 2.6, we use these fundamental groups to allocate manuscripts to the groups; to determine manuscripts which use more than one exemplar, either by shift of exemplar or by contamination; and to judge the exact relationship of particular manuscripts with each other. This is done through Variant Group profiles. In essence, we take each of the fundamental groups of variants found in the last section and then, for each manuscript, we count how many of these variants are found in different parts of the manuscripts.

3.3.1 Making the Variant Group profiles

The first stage of this process is to create, for each fundamental group of variants, a Variant Group file containing all the variants of that group. This is done through VBase, as follows (using the example of the a group):

1. Load the variant query for the a group

2. You should see the VBase search dialogue box, with the a query loaded and the information that 237 variants satisfy the query

3. Press the button **Var. group...** in this dialogue. You will be asked for the name of the file in which you want to save this variant set (conveniently, VBase will offer the name 'A variants'). You could press **Cancel** in the file dialogue, or you could save the file on your computer (not on the CD-ROM!)

The Variant Group file is a plain text file, and you can read it if you wish. You could use VBase itself to open and read the file, through **Open** on the **File** menu, or any text-only editor. You should be careful not to edit the file, as it is structured for further processing by VBase. The beginning of the file appears as follows:

```
CollateDB="GP.db" font=" Geneva" size="12"
A Variants -- Variant Set, 237 hits, resulting from Multiple Query
 in >1 of Ds1 En1 Dd: FROM \start TO \end
      AND in >0 of Cn Ma Dd: FROM \start TO \end
      AND in <2 of Hg El: FROM \start TO \end
      AND in <2 of Cx1 Cx2 Tc2 Ii: FROM \start TO \end
      AND in <3 of Cp La Bo1 Ph2 Lc Mg: FROM \start TO \end


63 L5 Whan ]
0 And Ds1 En1 Ha2 Ma Mg To1

277 L19 in that ]
0 omitted Ds1 En1 Ld1 Ma Wy
```


```
CollateDB="GP.db" font=" Geneva" size="12"
A Variants -- Variant Set, 237 hits, resulting from Multiple Query
 in >1 of Ds1 En1 Dd: FROM \start TO \end
      AND in >0 of Cn Ma Dd: FROM \start TO \end
      AND in <2 of Hg El: FROM \start TO \end
      AND in <2 of Cx1 Cx2 Tc2 Ii: FROM \start TO \end
      AND in <3 of Cp La Bo1 Ph2 Lc Mg: FROM \start TO \end


63 L5 Whan ]
0 And Ds1 En1 Ha2 Ma Mg To1

277 L19 in that ]
0 omitted Ds1 En1 Ld1 Ma Wy
```


Figure 27: a Variant Group File

A Variant Group file was made for each of the fundamental groups identified in 3.1 above: that is, for the a b ab cd e and alpha groups. These are in the files AVars, BVars, etc., in the 'docs' folder on the CD-ROM.

These files were then split into separate files, corresponding to the two-hundred line sections 1-200, 201-400, 401-600, 601-end. The alpha

variants file was divided into two, up to and after line 500: there are so few alpha variants that further division would have been meaningless. Dividing the files in this way will make it possible to observe changes in affiliation within the length of the text. These files are named A0002000, A200400, etc., for the variants from 1-200, 201-400.

We are now ready to generate the Variant Group profiles, for each manuscript. Each Variant Group profile is a list of how many variants in each Variant Group is present in each manuscript. In order to make the Variant Group profiles, we need a further file, listing all the Variant Group files to be used while making the profiles. This is a plain-text file, containing only the names of the Variant Group files with each file-name appearing on a line to itself. The file VSetList.txt, in the 'docs' directory, is such a file. This file must be in the same directory as the Variant Group files it references. The first lines of this file are:

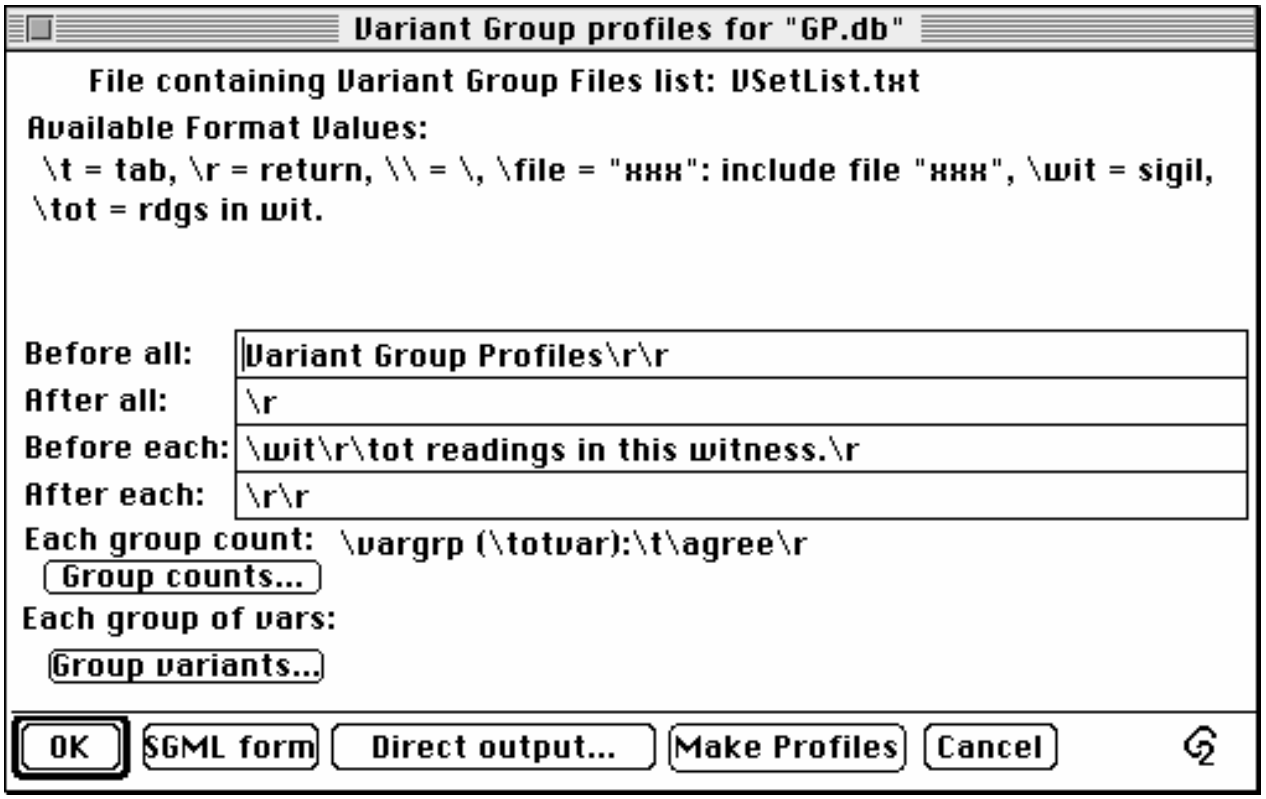
Alphvars
Alph500
AlphEnd
ABVars
AB000200
AB200400

Figure 28: a file listing the Variant Group Files

This directs VBase to use the files Alphvars, Alph500 (alpha up to line 500), AlphEnd (the rest of Alpha), ABVars, AB0002000, AB200400 to make the Variant Group profiles.

To make the profiles (note: this feature may not have been implemented in the version of VBase provided for your computer platform):

- choose **Make Variant Group Profiles...** from the **Database** menu of VBase (you should already have the database file 'GP.db' loaded. If you do not, then load it through the command **Load Variant Database...** on the same menu.)
- you will be asked to select the file containing the list of Variant Group files. Select the file 'VSetList.txt'
- the Variant Group profiles dialogue will appear:



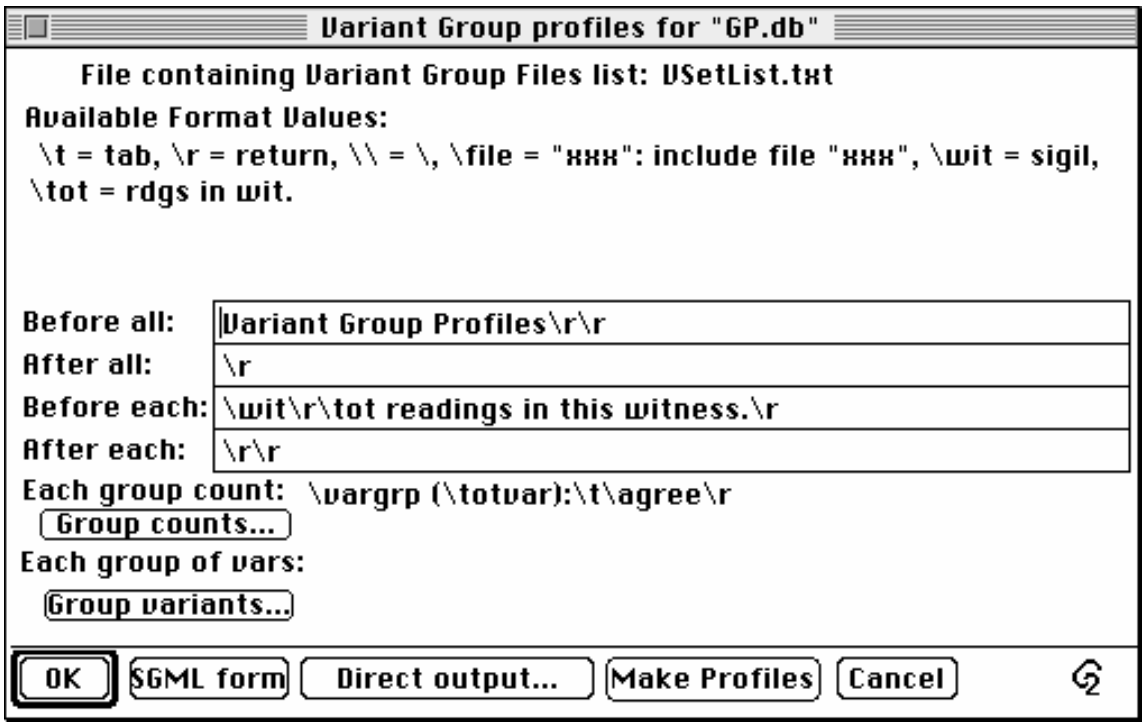


Figure 29: the Variant Groups dialogue box

- Click on the button **Make Profiles** at the base of the dialogue. VBase will open a new window and will write the Variant Group profiles into this window.

The Variant Group profiles for the manuscripts made of The General Prologue made by this process are in the file 'VarProfs.txt', in the 'docs' directory on the CD-ROM. You can open this file with VBase, or a text-only word processor. The first lines of it contain the Variant Group profile for Ad[1], thus:

Ad[1]

7798 readings in this witness.
Alphvars (35): 18
Alph500 (25): 14
AlphEnd (10): 4
ABVars (67): 5

AB000200 (9): 1
AB200400 (19): 2
AB400600 (19): 1
AB600END (20): 1
AVars (237): 17
A000200 (49): 6
A200400 (52): 3
A400600 (70): 4
A600END (66): 4
BVars (215): 15
B000200 (38): 3
B200400 (43): 2
B400600 (67): 3
B600END (67): 7
CDVars (48): 6
CD000200 (3): 0
CD200400 (11): 4
CD400600 (17): 0
CD600End (17): 2
EVars (55): 1
E000200 (4): 1
E200400 (12): 0
E400600 (16): 0
E600End (23): 0

Exercises: now, try these for yourself..

1. Try making your own Variant Profiles using the instructions given above, for different selections of the Variant Group files provided, or your own Variant Group files.
2. You can have VBase output a list of all the variants in a particular group in a particular witness. You do this through pressing the

Group Variants... button in the Group Profiles dialogue box (Figure 29) and typing the following into the box which then appears (note: this feature may not have been implemented in the version of VBase provided for your computer platform):

\vargrp (\totvar): \agree readings in this witness\vargrpcont\r

This is an instruction, in the VBase programming language, for VBase to output for each group the name of the group (\vargrp), the total number of variants in that group (\totvar); the total number of those variants in this manuscript (\agree) and the text of all those agreements (\vargrpcont) followed by a carriage return (\r). You could use this programming language (which is a subset of the *Collate* programming language) to produce output in different formats.

3.3.2 Reading the Variant Group profiles

Taking the example of Ad^1 given above, we may read this group profile as follows: Ad^1 is likely to be descended from alpha, as it has 18 of the 35 alpha variants. There are no signs of affiliation with any other group: typically, Ad^1 has fewer than 10% of the variants characteristic of any group (for example, 5 of the 65 ab variants; 17 of the 237 a variants, and so on.)

The division of the profiles into segments should permit us to see instances of shift of exemplar in any one manuscript, from an exemplar of one group to an exemplar of another. There is a clear case of this in Ld^1. The Variant Group profile for Ld^1 is as follows:

Ld^1

7557 readings in this witness.

Alphvars (35):	24
Alph500 (25):	22
AlphEnd (10):	2
ABVars (67):	36
AB000200 (9):	8
AB200400 (19):	14
AB400600 (19):	13

	1
AVars (237):	13
A000200 (49):	3
A200400 (52):	3
A400600 (70):	6
A600END (66):	1
BVars (215):	109
B000200 (38):	23
B000200 (43):	38
B400600 (67):	48
B600END (67):	0
CDVars (48):	26
CD000200 (3):	2
CD200400 (11):	0
CD400600 (17):	8
CD600End (17):	16
EVars (55):	4
E000200 (4):	1
E200400 (12):	1
E400600 (16):	2
E600End (23):	0

Up to about line 500, Ld[1] is clearly a member of the b group: it has 61 of the 81 b variants in lines 1-400 (the lines B000200 and B000200); it has 22 of the 28 ab variants in lines 1-400 (AB000200 and AB200400); it has 22 of the 25 alpha variants up to line 500 (Alph500). This would make it a perfectly regular member of the b group: you will see near identical patterns of numbers in Cx[1] Ii Cx[2] and other b group members. But Ld[1] has no b variants after line 600 at all; only one ab variant; and no alpha variants (the two listed in AlphEnd both appear before line 600). However, after line 600 Ld[1] has an extremely high proportion of the cd variants, containing 16 of the 17 cd variants in the file CD600END. The obvious explanation is that at some point between lines 400 and 600, the copyist of Ld[1] changed from a b exemplar to a cd exemplar.

The Variant Group profiles may be used in a similar manner to identify possible cases of contamination: the use by the scribe of more than one exemplar in one stretch of text. Indeed, it appears that this is what happens between lines 400 and 600 in Ld[1]. There is not a single point in the text, before which there are only b variants and after which there are only cd variants. Examination of the files BVars and CDVars shows that the last b variant in Ld[1] occurs in line 521, and the first cd variant occurs in line 431 (regarding cd variants in Ld[1] in lines 94 and 131 as accidental agreement.) Between lines 431 and 521 the scribe mixes variants from the two exemplars, before moving exclusively to the cd exemplar from 521 on.

We can use the Variant Group profiles, also, to validate our analysis. If our definition of the fundamental manuscript groups is correct, then the Variant Group profiles should be consistent with one another. That is:

- No manuscript should belong to two unrelated groups. For example, a manuscript which is a member of the a or b groups cannot also be a member of the cd or e groups
- If a manuscript belongs to a group which is itself descended from another group then the manuscript must be a member of the parent group also. Thus: b manuscripts are descended from ab which is descended from alpha. Therefore, a b manuscript must have a significant proportion of ab and alpha readings as well as b readings.

Exercises: now, try these for yourself..

1. Examine the Variant Group profiles in the file VarProfs.txt. Can you find any instances of manuscripts which offend the principles set out above, that no manuscript should belong to unrelated groups and that manuscripts must inherit groupings?
2. Try your hand at classifying the manuscripts into groups yourself from this file, and compare them with the classifications suggested in the next section.

3.3.3 Provisional classification of the manuscripts by their Variant Group profiles

In this section, I examine the Variant Group profile for each manuscript and suggest how it might be classified. Of particular interest are the

twenty-five (or so) manuscripts for which SplitsTree and PAUP could not determine affiliation.

For incomplete manuscripts, I also give in square brackets a scaled figure for the number of readings in the group. The total number of readings in the manuscript gives some idea of how fragmentary the manuscript is. A full manuscript has around 7500 readings. Thus, Ad^3 with 4873 readings has about two thirds of the text. Therefore, the 10 alpha readings found in Ad^3 scales to 15 and I express this as follows: 10 [15] of 35.

Ad^1	alpha (19 of 35); no other affiliations
Ad^3	alpha (10 [15] of 35, but the manuscript is incomplete (4873 readings), and this may understate its affiliation. Its close relation Tc^1, which is complete for The General Prologue, has 25 of the 35 alpha variants)
Ad^4	too incomplete to be classifiable
Bo^1	e; no other affiliations
Bo^2	a proportion of cd variants (16 of 48) suggests that it may descend from the ultimate ancestor of cd; contamination is also possible
Bw	incomplete (2595 readings): the extant text (lines 570-end) is cd (13 of 17 cd variants line 600-end)
Ch	no affiliations. See 3.4.2 below
Cn	a (168 [212] of 237), and so also ab (55 [67] of 67) and alpha (29 [35] of 35). Incomplete (6292 readings); no other affiliations
Cp	cd (44 of 48); no other affiliations
Cx^1	b (215 of 215); and so also ab (66 of 67) and alpha (35 of 35) ; no other affiliations
Cx^2	b (127 of 215); and so also ab (62 of 67) and alpha (33 of 35) ; no other affiliations. The strikingly lower proportion of b readings than Cx^1 is likely to result from Caxton's correction of the text by a 'better manuscript'
Dd	a (118 [237] of 237), and so also ab (31 [62] of 67) and alpha (17 [35] of 35). Incomplete, 3147 readings.
Dl	cd (20 [30] of 48). Incomplete (4742 readings); no

	other affiliations
Do	incomplete. Apparently a copy or close relative of Gg
Ds^1	a (226 of 237), and so also ab (66 of 67) and alpha (35 of 35)
El	no affiliations. See 3.4.1 below.
En^1	a (234 of 237), and so also ab (66 of 67) and alpha (34 of 35)
En^3	alpha (19 of 35). No other affiliations
Fi	appears to combine alpha/ab readings and cd readings throughout: the proportion of cd variants (24 of 48) suggests that the base exemplar was cd, but that readings from a manuscript close to ab were deliberately imported into the text (10 of 35 alpha and 23 of 67 ab).
Gg	no affiliations. See 3.4.4 below
Gl	cd (26 [44] of 48). Incomplete (4550 readings); no other affiliations
Ha^2	cd (23 of 48). The lower than normal proportion of cd variants found in descendants of cd is likely to be the result of 'variant drift'[2] and replacement of the cd variants by those characteristic of the triplet Ha^2/Lc/Mg
Ha^3	cd (23 of 48). The lower than normal proportion of cd variants found in descendants of cd is likely to be the result of 'variant drift' (see 2.2)
Ha^4	There are signs that an alpha manuscript lies behind Ha^4: thus the 12 of 35 alpha variants found in Ha^4. However, the scribe (or that of the exemplar) appears to have introduced a considerable amount of variants into the text. Many of these coincide with variants in other manuscripts: thus the scattering of readings throughout Ha^4 which properly belong to other groups, and the alpha agreements may only be the result of such coincidence. It is suggested below that Ha^4 descends from a manuscript close to Hg: see 3.4.2
Hg	no affiliations. See 3.4.1 below
Ht	ab (31 of 67) and alpha (24 of 35). See the next section for an explanation of the number of a (30 of 237) and b

(54 of 215) variants present in Ht.

Ii — b (215 of 215); and so also ab (47 of 67) and alpha (30 of 35) ; no other affiliations

La — cd (43 of 48); no other affiliations

Lc — cd (25 of 48); no other affiliations

Ld1 — b up to around line 500 and cd thereafter, with a mixture of b and cd around lines 450-550. See the last section

Ld2 — The proportion of cd variants (20 of 48) suggests a cd exemplar lies behind Ld2. The manuscript was corrected against a printed edition, and this may account for the proportion of alpha, ab and b variants found in Ld2

Ln — Slight signs of affiliation to ab and alpha (12 [20] of 67 ab, 6 [10] of 35 alpha); this is difficult to judge as the manuscript is incomplete (4287 readings). No other affiliations

Ma — a (203 of 237), and so also ab (63 of 67) and alpha (33 of 35); no other affiliations

Mg — cd (27 of 48); cf. on Ha2 above

Mm — cd (42 of 48); no other affiliations

Ne — b (76 [215] of 215), and so also ab (20 [60] of 67) and alpha (8 [24] of 35). Incomplete (2621 readings); no other affiliations

Nl — b (155 [200] of 215), and so also ab (44 [55] of 67) and alpha (24 [30] of 35). Incomplete (6096 readings); no other affiliations

Ph2 — e. No other affiliations

Pn — b (122 of 215); and so also ab (56 of 67) and alpha (32 of 35); no other affiliations. The strikingly lower proportion of b readings than Cx1 is due to inheritance of the corrections made by Caxton in Cx2

Ps — Slight signs of affiliation to ab (15 of 67) and hence alpha (13 of 35). This is a heavily edited manuscript: it is difficult to tell whether these ab/alpha readings arose

	through descent or were introduced in the correction process
Pw	cd (40 of 48); no other affiliations
Py	ab (24 of 67) and alpha (19 of 35). See the next section for an explanation of the number of a (22 of 237) and b (30 of 215) variants
Ra^2	ab (17 [38] of 67) and alpha (9 [21] of 35). See the next section for an explanation of the number of a (17 [38] of 237) and b (23 [50] of 215) variants. Incomplete (3320 readings)
Ra^3	no affiliations; some suggestion of relationship with cd (14 [18] of 48). See 3.4.4
Ry^1	Probably ab (21 of 67) and alpha (19 of 35). No other affiliations
Ry^2	cd (44 of 48); no other affiliations
Se	cd (38 of 48); no other affiliations
Sl^1	cd (8 [29] of 48). Incomplete (1920 readings); no other affiliations
Sl^2	cd (24 of 48). A markedly higher degree of cd affiliation in lines 600-end, with 13 of 17 variants in that section and only 3 of 17 in lines 400-600. No other affiliations
Tc^1	alpha (25 of 35). The proportion of ab variants in Tc^1 (18 of 67) may indicate that some of these ab variants were actually present in alpha
Tc^2	b (215 of 215); and so also ab (66 of 67) and alpha (35 of 45); no other affiliations
To^1	no affiliations. See 3.4.4
Wy	b (125 of 215); and so also ab (60 of 67) and alpha (33 of 35) ; no other affiliations. The strikingly lower proportion of b readings than Cx^1 is due to inheritance of the corrections made by Caxton in Cx^2

This classification accordingly groups manuscripts as follows:

No affiliation or	Bo^2 Ch Do El Gg Ha^4 Hg Ln Ps Ra^3 To^1 (11)

Unclassifiable	Ad^4 (1)
alpha	Ad^1 Ad^3 En^3 Tc^1 (4)
ab	Ht Py Ra^2 Ry^1 (4)
a	Cn Dd Ds^1 En^1 Ma (5)
b	Cx^1 Cx^2 Ii Ld^1*[cd] Ne Nl Pn Tc^2 Wy (9)
cd	Bw Cp Dl Fi* [ab] Gl Ha^2 Ha^3 La Lc Ld^1*[b] Ld^2 Mg Mm Pw Ry^2 Se Sl^1 Sl^2 (18)
e	Bo^1 Ph^2 (2)

* indicates evidence of contamination or shift of exemplar. Ld^1 appears twice, as it shifts exemplar.

As the three groups ab a b are all descended from alpha, the six groups alpha ab a b cd e actually represent three independent lines of descent from the original, **O**: the manuscript which we suggest was Chaucer's working copy and the ancestor of the whole tradition. These lines of descent are through alpha (including ab a b: 21 manuscripts, excluding Ld^1), cd (18 copies), e (2 copies).

Exercises: now, try these for yourself..

1. Which of these classifications do you disagree with? Can you see ways to refine any of them?
2. For some eleven of these manuscripts, the analysis is unable to suggest any affiliation. How do you think we might assess these manuscripts? Compare your thoughts with 3.4.4 below.

3.3.4 Refining the definition of ab

This work revealed a difficulty with the identification of the ab group. Defining the ab group purely as the agreement of the a and b groups (that is, the variants found in both a and b) as in the query file GPabvars.cdb meant that a variant could have been present in ab, but would be identified as an a variant if it were not inherited by b, or as a b variant if it were not inherited by a. As a result, manuscripts such as Ht and Py which appear to descend from ab show a higher than usual number of a and b variants: variants which should properly have been identified as ab rather than either a or b.

The file GPabHtPy.cdb modifies the query used in file GPavars.cdb to identify the ab variants, by seeking to identify variants in either a or b and present also in Ht Py. This query increases the number of ab variants from 67 to 85: 52 of these 85 are present in Ht, 45 in Py.

We can now use this reshaped query to re-examine manuscripts tentatively assigned to ab, or apparently contaminated by ab. These are:

Fi	29 of 85 (23 of 67)
Ht	52 of 85 (31 of 67)
Ln	12 of 85 (12 [20] of 67)
Ps	20 of 85 (15 of 67)
Py	45 of 85 (24 of 67)
Ra^2	24 [50] of 85 (17 [38] of 67)
Ry^1	29 of 85 (21 of 67)

In fact, this redefinition does little to clarify the relationships of these manuscripts further. More work might be done on redefining the alpha and a and b queries also, in relation to these and other manuscripts. However, we are dealing with rather few variants and with manuscripts which may be separated by many intervening copies from their ancestors, and so have lost many of the readings present in those ancestors.

3.4 Manuscripts not grouped by this classification

In 3.3.3 above,we suggest that 41 of the 52 classifiable manuscripts of The General Prologue represent three lines of descent from the original, **O**, the manuscript which we suggest was Chaucer's working copy and the ancestor of the whole tradition. These lines of descent are:

through alpha (including ab a b):	21 manuscripts
through cd:	18 manuscripts
through e:	2 manuscripts

The eleven remaining manuscripts are Bo^2 Ch Do El Gg Ha^4 Hg Ln Ps Ra^3 To^1. Of these, Do may be disregarded as fragmentary, and closely related to Gg. The ten remaining manuscripts include the two manuscripts

related to Gg. The ten remaining manuscripts include the two manuscripts usually regarded as most important to the establishment of the text, El Hg, and two others in Gg Ha4 which have heavily influenced past editors. It appears that the only relationship these ten have with one another may be that they are all descended from **O**. In the Wife of Bath's Prologue analysis I labelled manuscripts whose only relationship was that they were all descended from **O** as the Q manuscripts. Eight of these ten (all except Gg Ln) were among those classified as Q manuscripts in the Wife of Bath's Prologue analysis.

It is improbable that these ten represent ten further lines of descent from **O**. This section shows how the tools available to us might be used to clarify the relationship of these ten to each other, and other manuscripts of The General Prologue.

It must be emphasized that the manuscripts labelled as 'Q manuscripts' do not form a genetic group within the tradition, in the sense that manuscripts of the groups alpha a b cd e do belong to such genetic groupings. Each of these groupings represents a family of manuscripts which is descended from a hypearchetype below **O**. Each grouping is therefore a family within the tradition; they share an ancestor which other manuscripts do not have. However, the Q manuscripts share no ancestor below **O** itself. Thus, they share no ancestor not shared by every other manuscript in the tradition. It is more accurate to think of these as 'ungrouped manuscripts', rather than as manuscripts of any group. However, as long as we think of them as ' group Q', but with the reservation that group Q differs from other groups, this name could also be used.

3.4.1 Hg El

The most important relationship among all the manuscripts of The General Prologue, as indeed for the whole *Tales*, is the relationship between Hg and El. These two are among the earliest of all the manuscripts of the *Tales*, possibly written within a decade of Chaucer's death or even within his lifetime. Hg presents a text of uncommon excellence throughout the *Tales*; parts of the text of El appear drawn from an inferior exemplar (notably the first half of The Wife of Bath's Prologue) but the tale order in El is usually regarded as superior to that in Hg. Further, both manuscripts are (by common agreement of all who have examined the manuscripts themselves) written by the one scribe, known to Middle English scholars as 'scribe b' (Doyle and Parkes, 1978).

In The General Prologue Hg is not descended from El, nor El descended from Hg. Both have numerous variants not present in the other. For example: Hg has lines 253-254 in our numbering (And he yaf a certeyn ferme for the graunte..); El has lines 638-1 and 638-2 in our numbering (..Thanne wolde he speke no word but latyn). The first question is: do Hg and El share an exemplar below **O**, and so represent a single line of descent? If they do: are any other manuscripts descended from this same exemplar?

The VBase query GPElHg.cdb is designed to extract the variants shared by El Hg and not present in **O**, and so possibly introduced by a shared ancestor below **O**. This is the query:

in Hg El with !punct
AND in <5 of Bo^2 Ch Gg Ha^4 Ln Ps Ra^3 To^1 Cp En^3 Bo^1

The second line of the query is designed to exclude variants likely to have been in the archetype, **O**. If the variant were in **O**, it is probable that it will be in at least five of the eleven manuscripts in this line: En^3 (representing alpha), Cp (cd), Bo^1 (e) and the eight other O manuscripts. To be of any significance, a reading must not have been present in the archetype.

This query returns 32 readings. As only 35 and 48 readings were judged sufficient to warrant hypothesis of the alpha and cd hypearchetypes, this might appear enough to suggest the existence of a joint ancestor for El Hg below **O**. However, it is not. Scrutiny of the readings themselves suggests that by their nature almost all — indeed, probably all — of these are authorial in character and so likely to have been present in **O**. Accordingly, the existence of these readings in El Hg and rarely elsewhere is a welcome confirmation of what editors have long known: that El and Hg are excellent copies, and that they preserve many readings present in **O** and lost in other copies. In this case, the query only shows the inferiority of other branches of the tradition to the joint witness of El Hg. It is, of course, possible that El Hg did have a joint ancestor. However, this joint ancestor appears to have preserved the readings of **O** so well (as do El Hg in their turn) that its existence, if it did exist, is of no significance. Accordingly, we should treat El Hg as representing two distinct lines of descent from **O**. This, together with their early date and their closeness to **O**, mandates any readings which are present in both El

their closeness to **O**, mandates any readings which are present in both El Hg as virtually certain to have been present in **O**. Where El and Hg agree, whatever other manuscripts have is likely to be of historic interest only.

The focus, then, shifts to the points where El and Hg disagree. It is possible that other manuscripts may be descended from either El or Hg. Of even more importance would be instances where other manuscripts could be shown to descend from an ancestor between either El or Hg and **O**. The first step towards this is to identify where El and Hg differ from each other and from **O**.

The VBase queries GPElxHg.cdb and GPHgxEl.cdb aim to do this. They are designed to isolate all the variants not present in the archetype, introduced in one of the pair (or in an ancestor between it and **O**), and not present in the other. The query GPElxHg.cdb does this for those in El and not in Hg; GPHgxEl.cdb for those in Hg and not in El. Thus, the query GPElxHg.cdb:

in El with !punct
AND not in Hg
AND in <5 of Bo^{2} Ch Gg Ha^{4} Ln Ps Ra^{3} To^{1} Cp En^{3} Bo^{1} (i.e. likely not to be in **O**)

The query above, seeking variants introduced in El and not present in Hg, produced 54 readings. The reverse query, seeking variants introduced in Hg and not present in El (in GPHgxEl.cdb) produced 84 variants. Part of the reason there are so many more variants in Hg and not El is because this query retrieves the inclusion in Hg of lines 253-4 (found in Ch Cx^{2} Hg Ld^{2} Pn Py Tc^{1} Wy). This accounts for seventeen of the 84 variants. You can examine these variants yourself by running these queries.

The next step is to study the distribution of these two sets of variants across the other manuscripts, and especially across the other eight ungrouped manuscripts. We do this by the same process as earlier: making Variant Group profiles showing the distribution of these variants across the manuscripts.

The Variant Group file 'GPElxHg' contains the Variant Group for the 54 readings present in El and not Hg or **O**; 'GPHgxEl' that for the 67 readings present in Hg and not El or **O**, excluding the seventeen readings

in lines 253-4, which will be considered later. The file 'HgElSets.txt' contains the names of these two files, ready to make Variant Group profiles for all the manuscripts. The file 'HgElProf.txt' contains the Variant Group profiles for all the manuscripts, showing the distribution in the manuscripts. Here are some of the manuscript Variant Group profiles in this file:

Ad^{1}

7798 readings in this witness.
GPElxHg (54): 8
GPHgxEl (67): 10

Ad^{3}

4873 readings in this witness.
GPElxHg (54): 3
GPHgxEl (67): 16

Bo^{1}

7571 readings in this witness.
GPElxHg (54): 14
GPHgxEl (67): 16

Bo^{2}

7524 readings in this witness.
GPElxHg (54): 16
GPHgxEl (67): 3

Ch

7600 readings in this witness.
GPElxHg (54): 3
GPHgxEl (67): 38

Cx^{1}

7466 readings in this witness.
GPElxHg (54): 7

Cx[2]

7518 readings in this witness.
GPElxHg (54): 8
GPHgxEl (67): 17

En[3]

7834 readings in this witness.
GPElxHg (54): 8
GPHgxEl (67): 10

Ha[4]

7571 readings in this witness.
GPElxHg (54): 8
GPHgxEl (67): 25

Ht

6993 readings in this witness.
GPElxHg (54): 6
GPHgxEl (67): 16

Ld[2]

7549 readings in this witness.
GPElxHg (54): 4
GPHgxEl (67): 14

Ph[2]

7498 readings in this witness.
GPElxHg (54): 15
GPHgxEl (67): 17

Py

7506 readings in this witness.
GPElxHg (54): 7
GPHgxEl (67): 26

Tc[1]

7516 readings in this witness.

GPElxHg (54): 11

GPHgxEl (67): 19

For most manuscripts, the numbers of El variants and Hg variants are about equal. This is exactly as we would expect of manuscripts whose descent from **O** is quite independent of El Hg. Thus, the e pair Bo[1] Ph[2] have nearly equal numbers of these El Hg variants (16/14, 15/17 respectively). This suggests that the descent of e from O, at least, is independent of El Hg. The e ancestor introduced various readings into the text: by simple chance, many of these happened to coincide with readings also introduced quite separately in El or Hg (or their ancestors).

However, some representatives of the groups alpha a b cd, and some ungrouped manuscripts, do show a larger number (sometimes, significantly larger) number of El than Hg variants, or vice versa. Again (as in the discussion of variants shared by El Hg above) one has to be careful here of ancestral variants. Some of the variants in El and not in Hg, and vice versa, will be ancestral variants: that is readings present in **O** though lost in many other manuscripts close to **O** and — in individual cases — lost in either El or Hg. Study of these El or Hg variants indeed shows that very few of them are clear errors; almost all could have been, and some certainly were, present in **O**. It is likely that the higher incidence of Hg variants than El in several manuscripts related to alpha (e.g. Ad[3] Ht Py Tc[1]) is partly due to this. At least two other manuscripts close to alpha have approximately equal numbers of El and Hg variants (Ad[1] En[3]); further, the alpha manuscripts show little firm agreement as to which Hg variants they include, a sign that their occurrence in alpha manuscripts is either by inheritance from **O** or simply random. However, the evidence is equivocal. If in other parts of the *Tales* it became clear that alpha is descended from an exemplar near Hg then this would be consistent with matters in The General Prologue. The same is true of cd, which also shows a slight preference for Hg variants, but not sufficient to be decisive.

Exercises: now, try these for yourself..

1. Study the three groups of variants related to Hg and El discussed in this section: the agreements of Hg and El against others (GPElHg.cdb); Hg against El (GPHgxEl.cdb) and El against Hg (GPElxHg.cdb).

); Hg against El (GPHgxEl.cdb) and El against Hg (GPElxHg.cdb). Are you able to reach any conclusions about which of these are Chaucer's own readings, which scribal errors?

2. There seemed slight evidence of a core group of variants present in Hg, in alpha, and not in El or **O**. If there is no such group, then alpha represents an independent line of descent from Hg. Are you able to produce firmer arguments that there was, or was not, such a group? (hint: start with the query El against Hg GPElxHg.cdb and search for combinations of the alpha and related manuscripts (e.g. Ad^1 Ad^3 En^3 Ht Py Tc^1) with it.
3. Can you find any evidence that the 'better manuscript' used by Caxton to correct his first edition was affiliated either to Hg or El? (hint: search for readings in Cx^2, not in Cx^1, in combination with the queries for Hg and El against others (GPElHg.cdb); Hg against El (GPHgxEl.cdb) and El against Hg (GPElxHg.cdb). Note that Cx^2 has the lines 253-4, present in Ch Hg but not El and rarely elsewhere. You might expect that the 'better manuscript' was (like Ch) affiliated to Hg throughout. Is it?

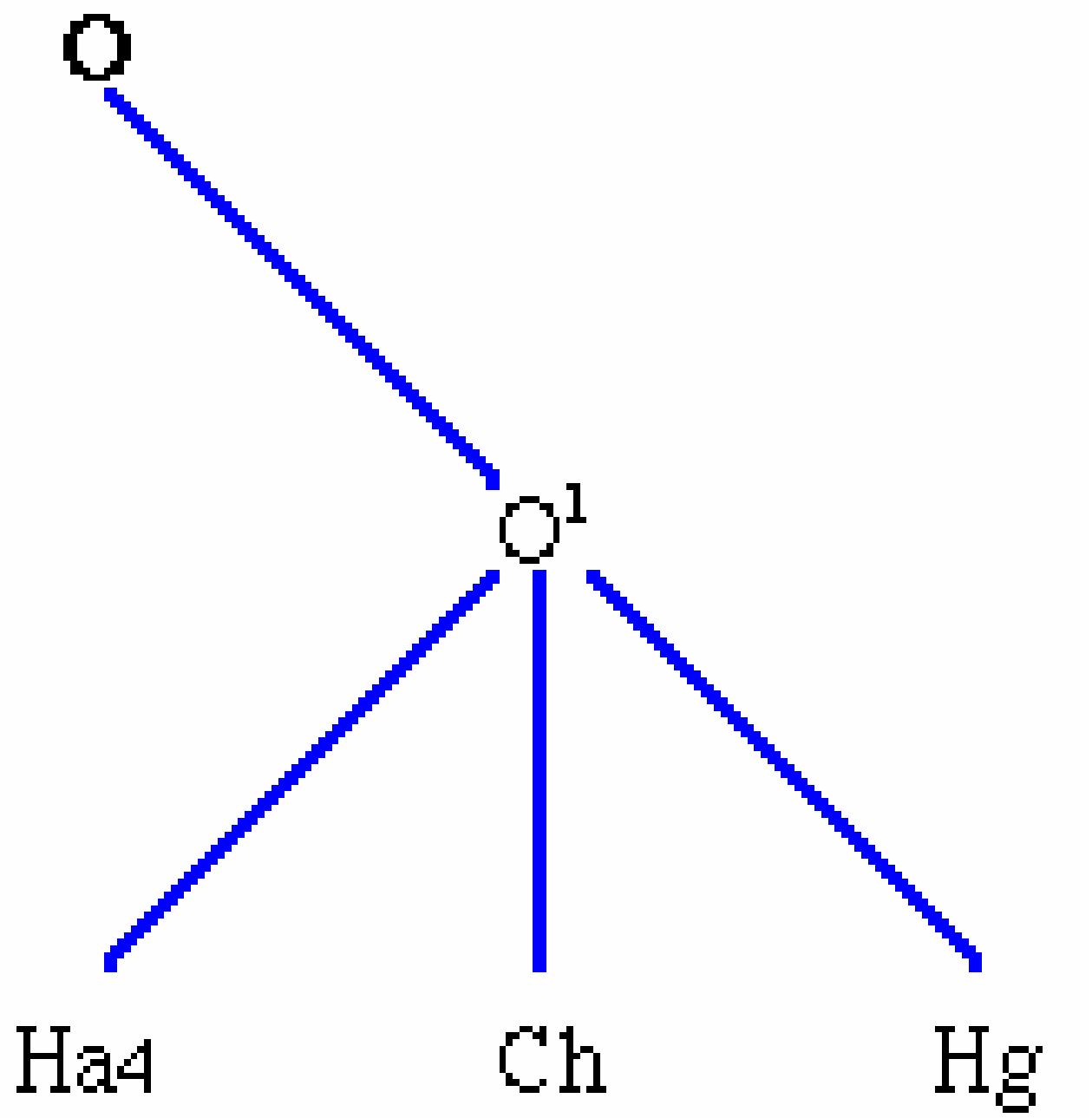
3.4.2 Hg Bo^2 Ch Ha^4

However, there are three manuscripts which seem to show a marked preference for either El or Hg variants. These are the three ungrouped manuscripts Bo^2 (16 El, 3 Hg) Ch (3 El, 38 Hg) Ha^4 (8 El, 25 Hg). The Bo^2 preference for El, while notable, is probably not sufficient in itself to be decisive (though would be important supporting evidence if a similar preference emerged in other parts of the *Tales*.) However, the incidence of agreement between Hg and Ch is so high, so evenly spread through the text, that it cannot possibly be either coincidence or the result of Ch and Hg both inheriting readings ancestral to the whole tradition but lost everywhere else. Either Ch is a descendant of Hg itself or it is a copy of an exemplar between Hg and **O**.

It seems very unlikely that Ch is a descendant of Hg itself. Ch includes the couplet 638-1 and 638-2, missing in Hg. Similarly, in The Wife of Bath's Prologue Ch included all five 'added passages' (again, missing in Hg), and I argued in the discussion of these that Ch had these from an exemplar close to **O**. Further, there are the three readings where Ch agrees with El against Hg. All these three readings are likely to have been in **O** (in the file GPChHgx.cdb). If Ch were descended from Hg one

would have to assume that Ch has independently restored these readings. However, if Ch is descended from an exemplar between **O** and Hg, then it might have these three readings and the couplet 638-1 and 638-2 all by descent from that exemplar.

The evidence of Ha^4 is also crucial. Of the 25 Hg agreements in Ha^4, 21 are shared with Ch (file GPHa4Chx.cdb.) The easiest explanation of this is that Ha^4 is descended from the same exemplar as Ch. It is even more difficult (indeed, really impossible) to argue that Ha^4 is descended from Hg itself. The likeliest explanation, then, for the distribution of these variants in Ch Ha^4 Hg is as follows:


O
O1
Ha4
Ch
Hg

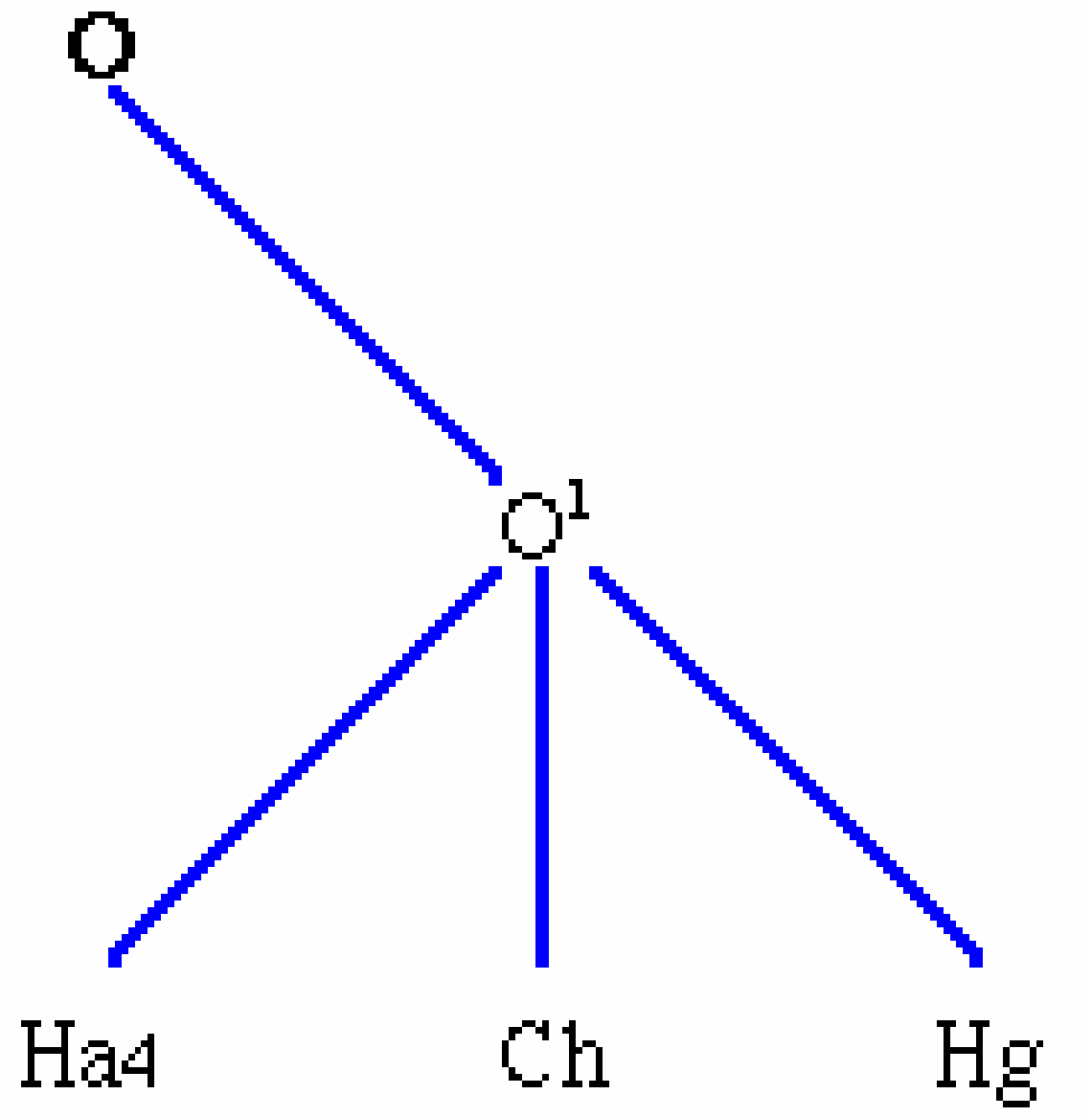


Figure 30: the descent of Ch Ha^4 Hg from a common exemplar

It has long been known that whatever else the defects of Ha^4, some readings at least show evidence of close relationship with **O**: the placement of Ha^4 here is consistent with that.

The nature of this joint ancestor of Ch Ha^4 Hg is crucial. It is clearly a

very good copy of **O**: hence the small number of variants (only 21 in file GPHa4Chx.cdb) shared by all three. By a rather remarkable chance, we may be able to deduce more about the nature of this **O1**. In an analysis of the spelling of scribe b (the scribe of both El and Hg) I observed that this scribe had highly distinctive spellings of some very common words (Robinson forthcoming). Almost alone among the manuscripts of the *Tales*, El and Hg use the forms dooth moore namoore oother (singular; regularly othre plural). The spelling

He slepte namoore than dooth a nyghtyngale

could be found nowhere else.

From study of the spelling databases for The Wife of Bath's Prologue I found that these spellings are found almost nowhere among the manuscripts of this part of the *Tales*, except in Hg and El where they are invariable. But one of the very few other manuscripts in which these spellings do occur is Ch, in The General Prologue (but not in The Wife of Bath's Prologue). A test of our transcript of Ch (still incomplete) showed the presence of these spellings in other parts of the *Tales* in Ch, but not in others. How do these spellings come to be present in Ch, in The General Prologue and elsewhere, as well as in El or Hg?

If Ch is not a copy of Hg (as I argue) then the spellings must have been present in **O1**. It is, of course, possible that these are Chaucer's own spellings, that they were in **O**, and that their presence in all of Ch Hg El is a sign once more only of their closeness to **O**. However, there is no reason to think that these spellings were characteristic of anyone except scribe b. These spellings occur not only in the scribe's copies of Chaucer but also in the scribe's stint in the Trinity College Gower, MS R 3.2.

This suggests that **O1**, the ancestor of all three of Ch Ha4 Hg, was a third copy of the *Tales* written by the same scribe who wrote El Hg, scribe b. In fact, we have a fragment of what might be this third copy, in Cambridge University MS Kk. 1. 3 (20), though Doyle and Parkes were reluctant to assign this certainly to scribe b.[5] Again, study of the relationship of Ch and Hg and the distribution of spellings in them in other parts of the *Tales* may illuminate this.

Exercises: now, try these for yourself..

1. On the CD-ROM, you have the spelling databases for all the manuscripts of The General Prologue. Examine the distribution of the spellings claimed as distinctive of El Hg (scribe b). Does their distribution in The General Prologue manuscripts support the hypothesis suggested above? Can you find other spellings which seem distinctive of scribe b, and are found also in Ch? (look, for example, at compaignye and note the variants on Auerill in line 1 of the text).
2. Examine the readings shared by Ch Hg (and also Ha^4). Are you able to reach any conclusions about which of these are Chaucer's own readings, which scribal errors or 'improvements'? You might care to compare these readings (presumably present in $\mathbf{O^1}$) with readings introduced in Ch or Ha^4 and not present in Hg.

3.4.3 Hg El; e; Manly and Rickert

In their discussion of The General Prologue, Manly and Rickert assert that the pair Bo^1 Ph^2 (e) is closely related to El up to line 288. Their wording appears to admit of no doubt: 'To 288 Bo^1 [i.e. the pair Bo&1; and Ph&2;] share all El variants (except 211, spelling)'. They nominate twelve points which support this suggestion. If true, this would be a relationship of the first importance.

However, I suggest above that e has no relationship with either El or Hg; hence the almost equal numbers of El and Hg variants in both Bo^1 and Ph^2. Further, VBase contradicts Manly and Rickert's assertion that Bo^1 /Ph^2 shares all El variants up to 288. The query GPElxE1.cdb finds eight variants in El, not in O or Hg, and not in Bo^1 or Ph^2, up to line 288. Several of these are highly significant: the phrasing of age he was in line 82; the replacement of any by is a in 207. Overall there are slightly fewer such counter-variants up to line 288 than in the remainder of the text (8 to 288; 29 in the remaining 570 lines). There are also rather more agreements between e and El up to line 288: 13 compared to 4 after line 289. But this variation could be the result of simple chance, again.

Manly and Rickert find 12 variants in support of a relationship between e and El up to line 288. The query GPElE289.cdb finds 13. Comparison of the two lists is instructive: only the nine variants in lines 23 148 188 196 199 215 240 251 289 are on both lists. In fact, examination of the variants returned by VBase suggests that the other four are either

variants returned by VBase suggests that the other four are either accidental or ancestral, so the number of variants is reduced to these nine: thin evidence indeed, especially as several of these nine may well be ancestral. The incidence of Hg variants throughout suggests that this apparent relationship with El is no more than chance.

Why, then, do Manly and Rickert think there is a relationship? The clue lies in the variants in lines 188, 234, 240 (query GPElEx.cdb). Here, the reading of the base text is first given with the evidence for it, followed by the reading in Bo^1 El Ph^2

188 Lat Austyn haue his swynk to hym reserued

40 mss: Ad^1 Ad^3 Bo^2 Cn Cp Cx^1 Cx^2 Dl Ds^1 En^1 En^3 Fi Gg Ha^2 Ha^3 Ha^4 Hg Ht Ii La Lc Ld^1 Ld^2 Ma Mg Mm Nl Pn Ps Pw Py Ra^3 Ry^1 Ry^2 Se Sl^2 Tc^1 Tc^2 To^1 Wy

his owene Bo^1 El Ph^2

234 And pynnes / for to yeuen faire wyues

40 mss: Ad^1 Bo^2 Ch Cn Cp Cx^1 Cx^2 Dl Ds^1 En^1 En^3 Fi Gg Ha^2 Ha^3 Ha^4 Hg Ht Ii La Lc Ld^1 Ld^2 Ma Mg Mm Nl Pn Ps Pw Py Ra^3 Ry^1 Ry^2 Se Sl^2 Tc^1 Tc^2 To^1 Wy

yonge Bo^1 El Ph^2

240 He knew the tauernes wel in euery town

38 mss Ad^1 Bo^2 Ch Cn Cp Cx^1 Cx^2 Dl Ds^1 En^1 En^3 Fi Gg Ha^2 Ha^3 Ha^4 Hg Ii La Lc Ld^1 Ma Mg Mm Nl Pn Ps Pw Py Ra^3 Ry^1 Ry^2 Se Sl^2 Tc^1 Tc^2 To^1 Wy

al the Bo^1 El Ph^2

This trio is certainly striking, and one can understand Manly and Rickert's excitement at their discovery of what appears (for once) to be firm evidence of a clear manuscript relationship among so much that is ambiguous. But one does not have to presume that Bo^1 El Ph^2 have a shared ancestor to produce this trio of readings. Recall that Bo^1 El Ph^2 represent two lines of descent from **O**, among perhaps some six lines of descent from **O** in all. It is quite possible for the two lines of descent to introduce the same error, just by chance, and for this error not to occur in the other lines of descent. This would produce exactly the distribution of variants here given. Indeed, one can find four points in the text where it appears that the single line of descent represented by Bo^1/Ph^2 introduces, just by chance, the same error as appears to have been introduced by the single line of descent represented by Hg. The query GPEHgx.cdb

identifies these four readings. There is another possibility: some of these readings might be ancestral to the whole tradition, though lost in every other line of descent. This can only be judged by assessing the readings themselves.

It is also possible that some of the readings found in only one or two of the lines of descent might be the result of Chaucer's own revision, expressed as annotations or substitutions in the text of **O** itself. The reading yonge in line 234 is arguably sharper than the conventional faire. The uneven distribution of these possible revisions across the different lines of descent would be the result of the different first copies of **O** making different ad hoc decisions about what to do with the competing authorial readings they found in **O** itself. Thus, I argued that for The Wife of Bath's Prologue Chaucer decided to remove the 'added passages' from the text (possibly, when he reassigned what was originally her tale to The Shipman, Robinson 1997). Chaucer indicated that the text should be deleted in **O**, and all the first copyists of The Wife of Bath's Prologue respected these marks except for the copyist of alpha, who restored the passage. Similarly, Chaucer may have either written lines 253-4 of The General Prologue in the margin, or marked it for deletion: all the first copyists except for the copyist of the ancestor of Ch Hg (termed $\mathbf{O^1}$; also perhaps the ancestor of Py and the 'better manuscript' behind Cx^2) chose to exclude the lines.

It appears that Manly and Rickert wanted to believe in this relationship between e (Bo^1 Ph^2) and El, and then sought evidence for it. However, several of the variants they saw as evidence are likely to have been ancestral to the whole tradition, although not present in Hg (for example: line 215 where And Ad^1 Bo^1 El En^3 Ld^1 is likely to have been present in alpha and therefore **O** beside e and El). Kane criticizes Manly and Rickert severely for their consistent refusal to identify readings likely to be ancestral and to exclude these from their analysis (Kane 1984). This criticism is well merited. Because ancestral readings can occur absolutely anywhere within the tradition, one can use them to prove any relationship one likes. This is exactly what Manly and Rickert do. Thus, their evidence for the existence of the 'composite group' (see above 3.2.1) melts to nothing if one excludes variants likely to be ancestral (query MRComp.cdb.)

The failure of Manly and Rickert to recognize significant groupings like

that of ab and alpha and their willingness (on the other hand) to postulate groupings which do not exist on the basis of an unsafe use of ancestral variants, render every part of their analysis suspect. At no point do they use the technique of counter-groups, here used to validate (or invalidate) group relationships. Many excuses can be made for them: the lack of time, support and money; the inadequate tools that made their effort heroic almost beyond our imagining. Manly and Rickert knew the tradition as, quite possibly, no-one ever will again. Their suggestion that manuscript groups are to be identified by 'persistent and consistent' agreement (II 20) is very close to the technique of 'significant number' advocated here. There is example after example in their pages of analysis of inspired intuition: one such is their suggestion that El in the second half of The Wife of Bath's Prologue derives its text from the a ancestor, the manuscript we call alpha. But there is example after example on their pages of hypotheses, often expressed in the most positive language, which are simply wrong. Their method is so defective that there is no way of telling what is right from what is wrong. Any analysis built on anything but their most broad conclusions, which trusts their assertions alone without further testing and validation, is in danger of being valueless.

Exercises: now, try these for yourself..

1. The pages of Manly and Rickert's analysis of The General Prologue are strewn with hypotheses which run counter to the analysis suggested in this workshop. For example, they argue for the following relationships:

   Bo^{1} Ph^{2} (e) as a genetic group with Py (II 80-81)

   Ps as descended from a b manuscript (II 81)

   Bo^{1} Ph^{2} (e) Py as descended from a b manuscript (II 81-2)

   Ha^{4} a b as sharing an ancestor (II 82)

   Ad^{3} Tc^{1} Ld^{2} as sharing an ancestor up to line 531 (II 84)

   Can you find what Manly and Rickert have seen that made them see relationships? Are they right, or wrong? Why?

2. ' One can use ancestral variants to prove any relationship one likes'. Using VBase, pick some random groupings of manuscripts and see whether you can find evidence using ancestral variants to 'prove' a

family relationship. A nice example would be to 'prove' that b and cd had a common exemplar (as Manly and Rickert actually suggest, II 90).

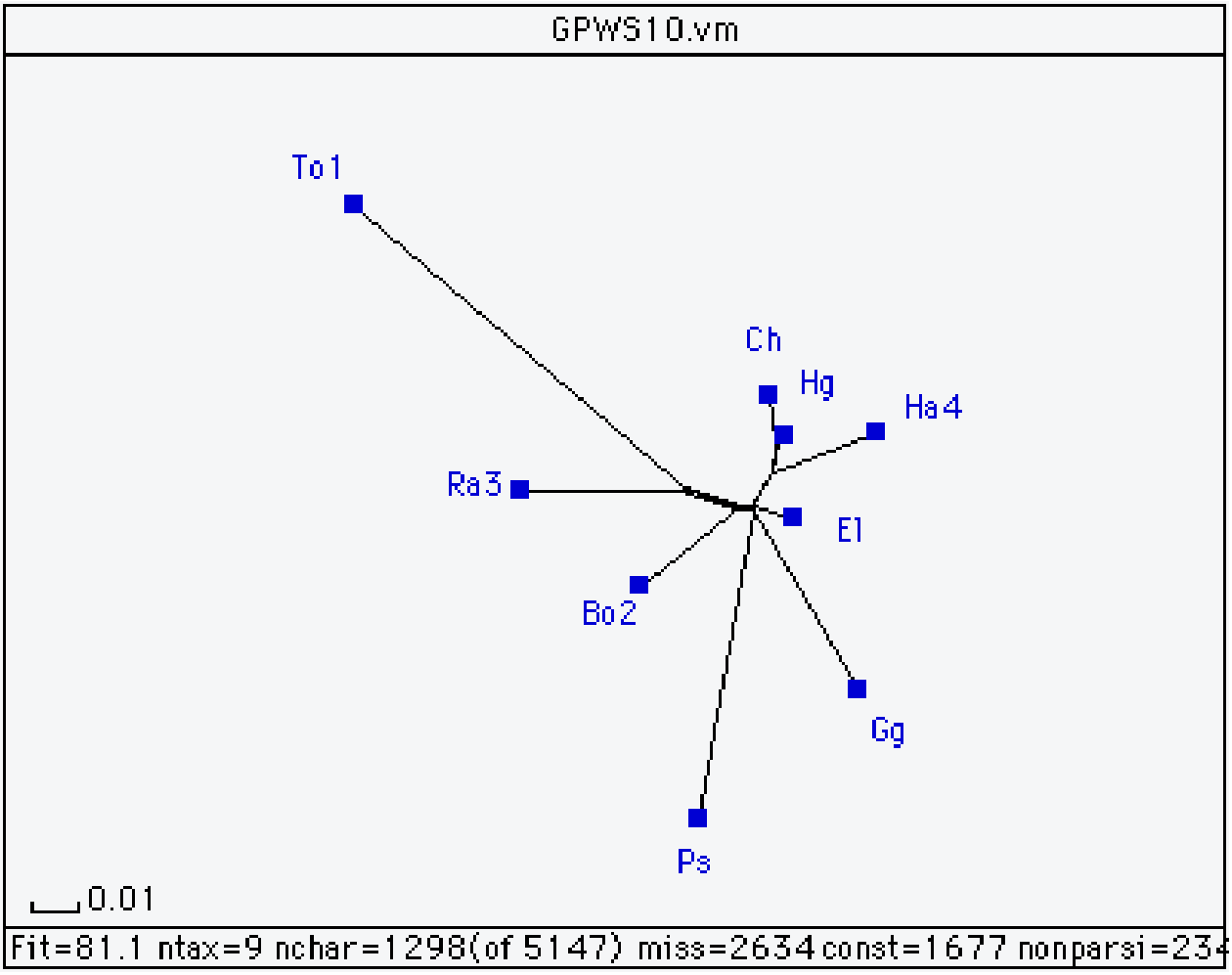
3.4.4 Our biggest problem: scribes who do not make mistakes

Our analysis of these ten ungrouped manuscripts Bo^2 Ch El Gg Ha^4 Hg Ln Ps Ra^3 To^1 (excluding the fragmentary Do) in the last sections has been able to establish only the following:

- El and Hg represent distinct lines of descent from **O** itself
- Ch is certainly affiliated with Hg, and may be a descendant of the same hypearchetype as Hg. Ha^4 may also be descended from this same hypearchetype, and Bo^2 from the same hypearchetype as El.

This then establishes two further lines of descent (El Hg) and three other manuscripts (Ch certainly; perhaps Ha^4 and Bo^2) may be affiliated with these two. What of the other five manuscripts, Gg Ln Ps Ra^3 To^1?

We could try and use SplitsTree and VBase to isolate groups among just these five and the other ungrouped texts. First, we divide the General Prologue into the same three segments (1-250; 251-500; 501-end) and look for evidence of tree-like relationships among just these ungrouped manuscripts. The file GPWS10.vm does this for lines 1-250 (excluding Ln, OUT up to line 383):


GPWS10.vm
To1
Ch
Hg
Ha4
Ra3
El
Bo2
Gg
Ps
0.01
Fit=81.1 ntax=9 nchar=1298(of 5147) miss=2634 const=1677 nonparsi=234

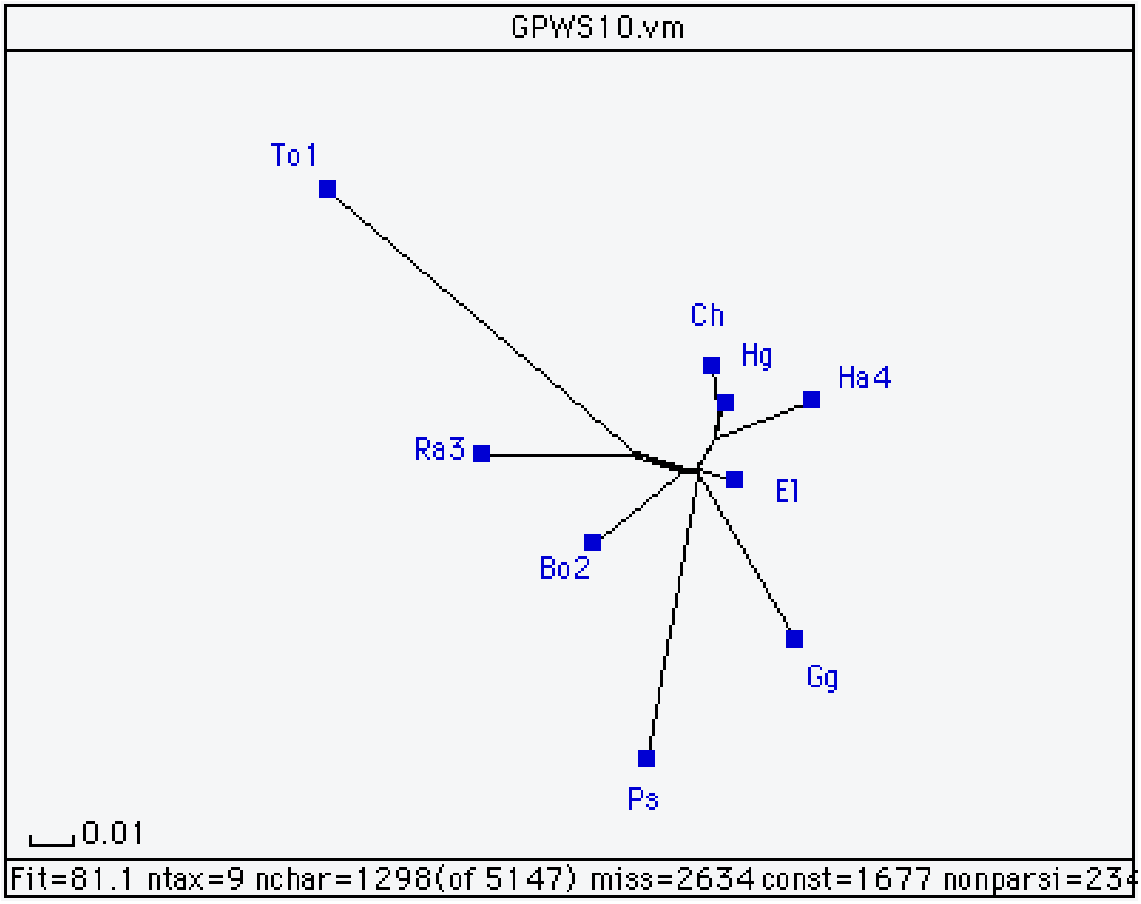
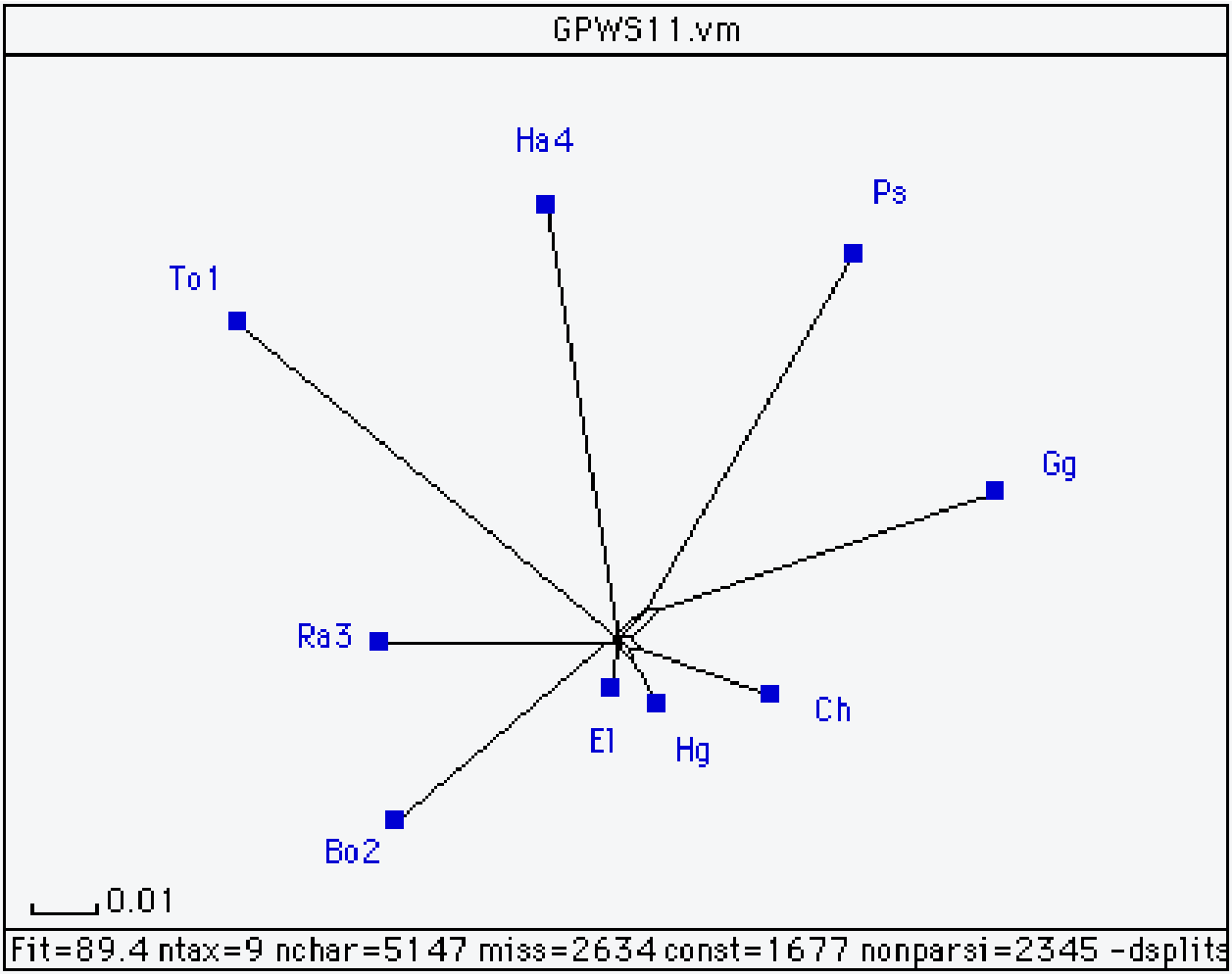


Figure 31: Ungrouped manuscripts, lines 1-250

Rather pleasingly, this seems to confirm quite clearly that Ch Ha^4 Hg all share a hypearchetype, as proposed above. Other relationships are less clear, though To^1 and Ra^3 may share an ancestor with Bo^2.

Here is the SplitsTree graph for lines 251-500, still excluding Ln (GPWS11.vm.)


GPWS11.vm
Ha4
Ps
To1
Gg
Ra3
Ch
El
Hg
Bo2
0.01
Fit=89.4 ntax=9 nchar=5147 miss=2634 const=1677 nonparsi=2345 -dsplits

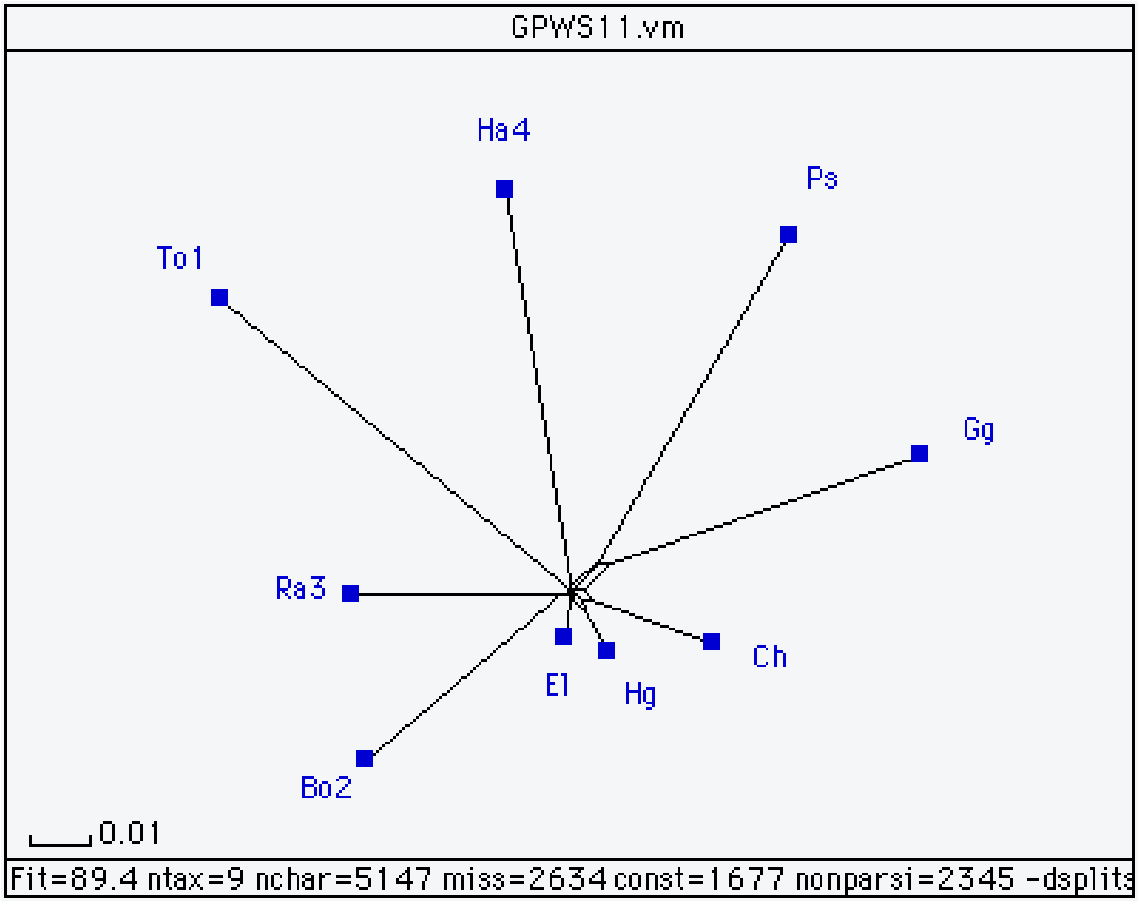
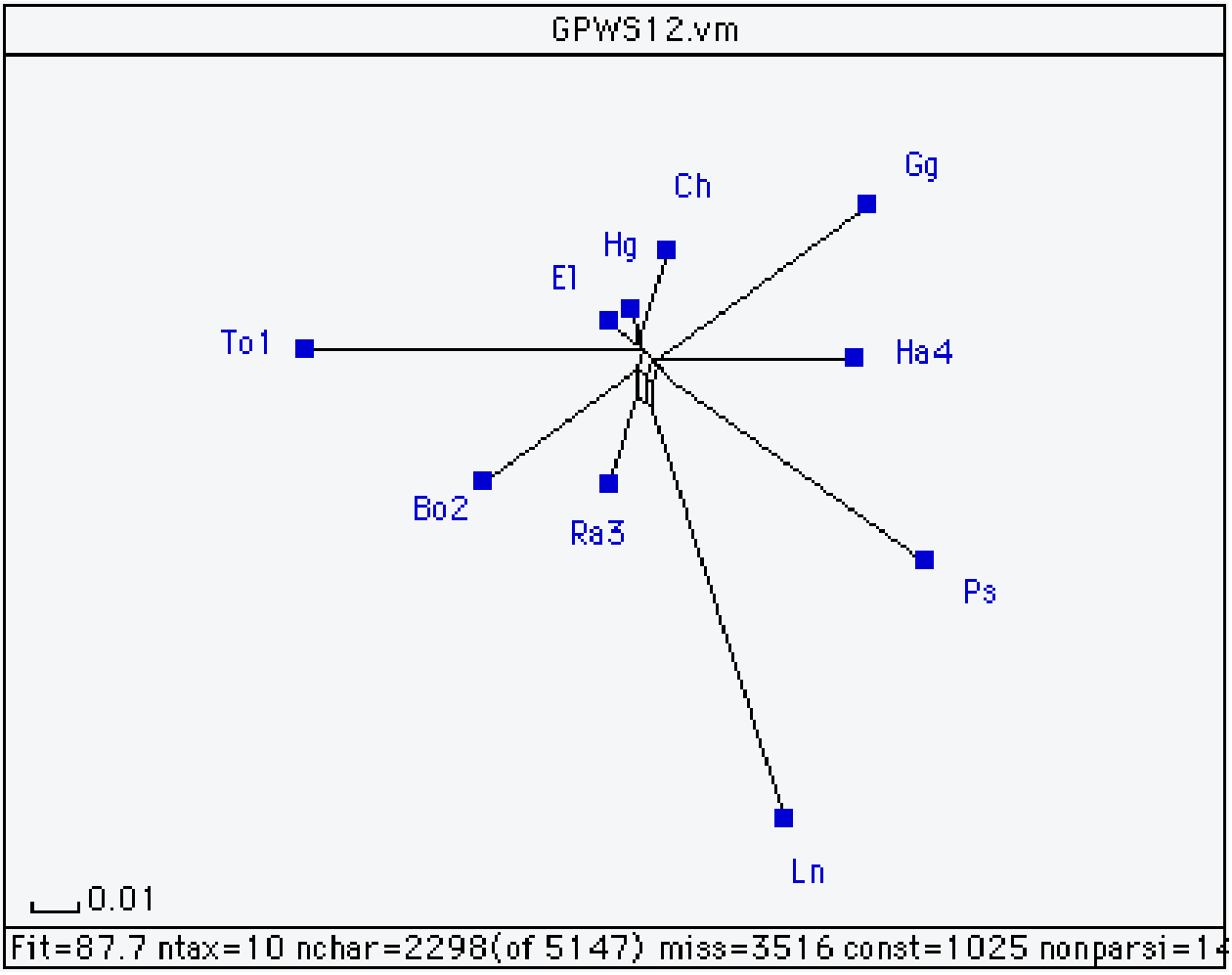


Figure 32: Ungrouped manuscripts, lines 251-500

The relationship between Ch Hg is still rather clearly marked here, but that between Bo^{2} Ra^{3} To^{1} appears less clear, while Ha^{4} has moved away from Ch Hg. Finally, for the last section of text, from lines 501-end SplitsTree produces this graph (GPWS12.vm):


GPWS12.vm
Gg
Ch
Hg
El
To1
Ha4
Bo2
Ra3
Ps
Ln
0.01
Fit=87.7 ntax=10 nchar=2298(of 5147) miss=3516 const=1025 nonparsi=14

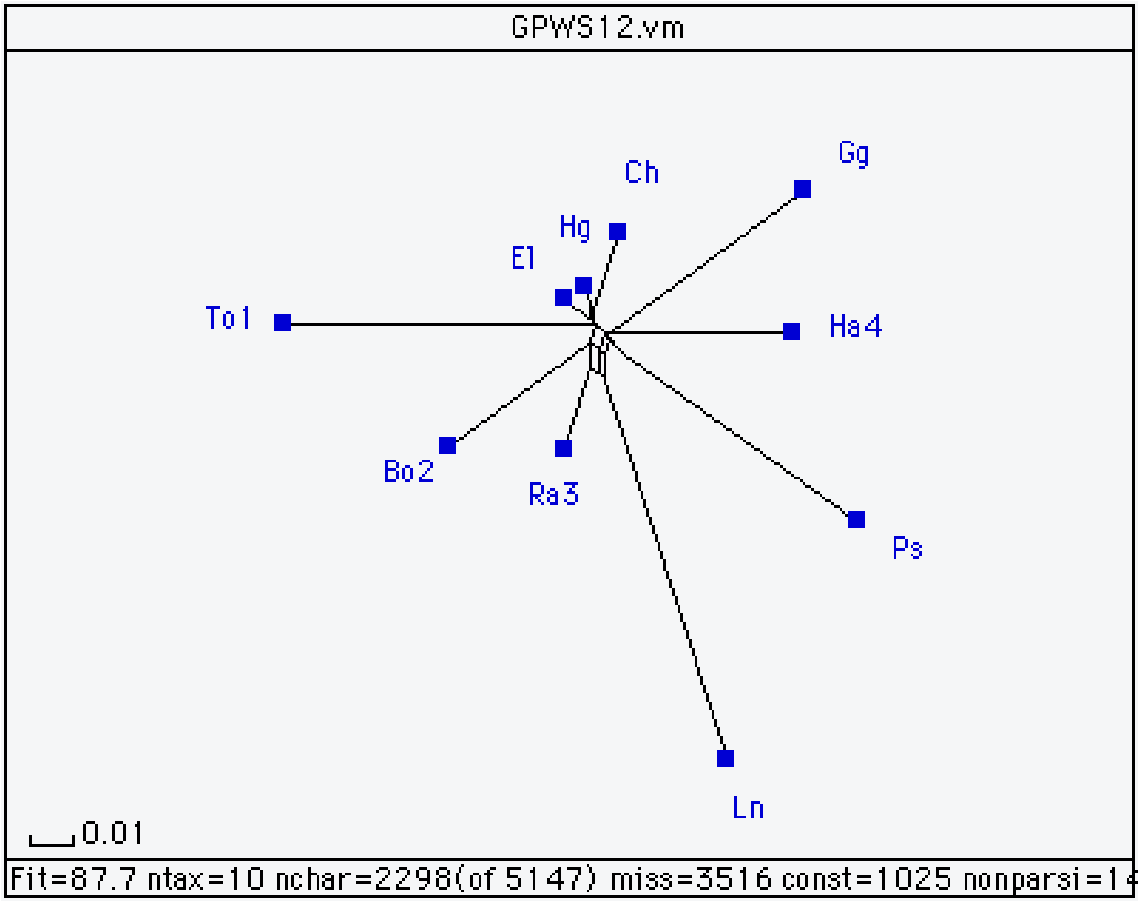


Figure 33: Ungrouped manuscripts, lines 501-end

Now, Ra^{3} appears to have moved right away from To^{1}, perhaps to join Ln; Ha^{4} remains apart from Hg/Ch; and other manuscripts appear to have no close relation with any other.

This picture of unstable relations between the manuscripts is exactly what we would expect if none of these ungrouped manuscripts shared any hypearchetype below **O**. In that case, the few agreements between any pair of sub-group against other manuscripts would be simply random; in another section of text, different random agreements might produce different pairings.

We can use VBase to test this hypothesis, that apart from Hg/Ch (and

possibly Ha^4 with them; and El/Bo^2), none of the ungrouped manuscripts share an ancestor below **O**. The file GPGgBo2.cdb searches for all variants which might have been introduced by a shared ancestor of Gg and Bo^2, below **O**:

in Gg Bo^2 with !punct
AND in <5 of Ch Ha^4 Ln Ps Cp En^3 Bo^1 Ra^3 To^1
AND not in El Hg

The first line of the query seeks variants in both Gg and Bo^2. The second line seeks to exclude all variants which might have been in **O**. If a variant were in **O**, then it would very likely be in at least five of the ancestors of alpha cd e (En^3 Cp Bo^1, as representative of these), and in the ungrouped manuscripts Ch Gg Ha^4 Ln Ps Ra^3 To^1. The third line excludes variants in both El and Hg: their unique proximity to and separate descent from **O** means that any reading in both manuscripts is highly likely to have been in **O** also.

This query returns just nine hits: so few as to be likely the result of simple chance. Other queries seeking manuscript pairings also produced so few hits that it appeared that chance rather than shared ancestry was responsible:

GPRa3To^1.cdb: variants in Ra^3 To^1, suggesting a common ancestor for this pair, 14 hits
GPHa4Gg.cdb: variants in Ha^4 Gg, suggesting a common ancestor for this pair, 18 hits
GPTo1Bo^2.cdb: variants in To^1 Bo^2, suggesting a common ancestor for this pair, 19 hits

A few queries did, however, suggest relationships within these manuscripts:

GPChHg.cdb: variants in Ch Hg, suggesting a common ancestor for this pair: 64 hits
GPChHaHg.cdb: variants in Ch Ha^4 Hg, suggesting a common ancestor for these three: thirty hits

The first of these confirms the relationship between Ch and Hg, discussed above, while the second of these suggests that Ha^4 joins this pair.

The relationship between Bo^2 and El remains uncertain. The query GPBo2El.cdb, seeking evidence for their shared ancestry, returns just 22

hits, and this may be no more than chance.

In assessing these results, the behaviour of Ps must give us pause. Ps is a famously 'corrected' manuscript. It was written for Jean d'Angoulême, probably while he was a prisoner at Maxey Castle, Northamptonshire from 1422-1436, by a scribe 'Duxworth' and contains extensive corrections by the scribe and by Jean himself. In The Wife of Bath's Prologue, I suggested that the exemplar of Ps was a cd manuscript, but with corrections from an ab or O manuscript (or, alpha). There appears to be a connection still with alpha from the figures given in the Variant Group profile for Ps (see 3.3.3), but no link with cd in The General Prologue. The connections of Ps with other ungrouped manuscripts appear baffling:

> GPPsHa4.cdb: variants in Ps Ha^4, suggesting a common ancestor for these two, 39 hits
>
> GPPsGg.cdb: variants in Ps Gg, suggesting a common ancestor for these two, 37 hits
>
> GPPsBo2.cdb: variants in Ps Bo^2, suggesting a common ancestor for these two, 23 hits

The first two of these are specially difficult to explain. If there is a common ancestor for each of the pairs Ps/Ha^4 and Ps/Gg then there should be a common ancestor for Gg/Ha^4. But the VBase query GPHa4Gg.cdb produces only 18 hits. This is likely to be chance, and (further) the supposition of a common ancestor for Gg/Ha^4/Ps appears to contradict the evidence for a relationship between Ch/Ha^4/Hg. The explanation for these contradictory relationships is likely to lie in the nature of Ps itself. It is heavily 'corrected'; it may combine readings from at least two different exemplars; the many more than usual places where its scribe and corrector departed from their copy text make it likely that there will be, by simple chance, many points than usual where Ps happens to agree with another manuscript. Ha^4 is similarly a highly corrected manuscript, and a high level of agreement between Ha^4 and other manuscripts (and particularly with Ps) may not then be of any significance.

For at least five of these ungrouped manuscripts Gg Ln Ps Ra^3 To^1, and perhaps also Bo^2 Ha^4, we are unable to show any affiliation with each other or with any other of the manuscripts. However, it is unlikely that each one of these ungrouped manuscripts represents an independent line

of descent from **O**. They may be severally or together linked to other manuscripts or groups. However, if the ancestor from which a grouping descends had itself very few variants, then there would be very few variants which these manuscripts might inherit and which might then show their shared descent. I suggest earlier that alpha might have had only some forty variants from its exemplar, **O**. Indeed, our analysis of cd suggests that the cd exemplar also had only some fifty variants. Such small numbers lie on the edge of our ability to discriminate truly significant groupings from mere chance.

It is a paradox, that our greatest difficulty in this stemmatic analysis is not contamination, or shift of exemplar, or scribes who introduce wanton error, but manuscripts which have very few variants. A case in point is Gg. In The Wife of Bath's Prologue, Gg was clearly affiliated with e, represented there by Bo^1 Ph^2 Si. However, there appears no evidence of affiliation, in the form of shared introduced readings, between Gg and e in The General Prologue. Yet, it is quite possible that Gg could in fact share an ancestor below **O** with e in The General Prologue, as it does in The Wife of Bath's Prologue. This ancestor might have had very few variants, and so there is no evidence of the relationship. The same may be true of alpha. The very first copy of **O** made in the alpha line of descent (that is, the copy from which all the ab manuscripts ultimately descend) may have had many fewer variants from **O** than even the forty or so identified by us as shared by ab and the manuscripts apparently descending from alpha (Ad^1 Ad^3 En^3 etc.) The scattering of alpha variants throughout the ungrouped manuscripts, and also in e, may suggest that alpha lies behind many more than the 21 manuscripts identified above as descended from it.

In these cases, the evidence we have of agreements and disagreements within this text alone does not of itself permit certain statement of relationships. However, it does indicate a range of possibilities. Analysis of other areas (e.g. tale order; the evidence of glosses, manuscript layout and decoration; relationships in other parts of the *Tales*) might then determine this further.

Exercises: now, try these for yourself..

1. Test out further the argument presented above, that one cannot distinguish subgroupings among the ungrouped manuscripts, and

particularly the five Gg Ln Ps Ra^3 To^1. Can you, using VBase, find evidence of shared ancestors below **O** for any of these?

2. Test the hypothesis, that a very early exemplar in the alpha line of descent may underlie many other manuscripts not allocated to alpha in 3.3.3 above: particularly, the e group and ungrouped manuscripts. Can you find evidence for this?
3. Can you find any evidence of relationships between individual cd manuscripts and individual ungrouped manuscripts? (if you do: how would you explain this? hint: look at Fi. Where do the readings in this manuscript come from?)

---

4. Using what we have learnt: the 'New Stemmatics'

From the remarks in the last section about the impossibility of achieving a certain classification for many of the manuscripts in the tradition, our analysis will clearly not be able to produce the kind of iron classification described by Maas and Greg. We are not able to specify exactly how many lines of descent there are; not able to determine exactly for every manuscript what line of descent it is in; and not able to fix just what readings were in the ancestor of each line of descent. Therefore, editors will not be able to use our work to produce, in any mechanical fashion, an edition by which readings are chosen by simple 'majority verdict': reading 'a' in five lines of descent, reading 'b' in two lines, choose reading 'a'. Our present knowledge permits no such certainty. The nature of the case suggests that no such certainty, no such edition, will ever be possible.

When we began this project, we thought that our aim was to accumulate materials which would help editors determine the early history of the text, and so help them edit the text for others to read. As the project has proceeded, we have come to see that our aim is not to help editors edit, but to help readers read. Through a perspicuous presentation of the material, readers might discover, with reasonable effort, both what choices are available at each point and also what lies behind these choices, and so judge how the text might best be read. This is not just presentation of the evidence. If that were all, one might just publish the transcripts and manuscript images. Rather, we aim to provide the tools for evaluation of the evidence, and instances of our conclusions drawn from use of these tools to instruct and challenge the reader. I have called this the 'new

stemmatics'.[1] Like the stemmatics of the last century, its aim is to illuminate the history of the text. Unlike the stemmatics of the old century, its aim is not a well-made edition, but a well-informed reader.

In the next section and in a separate part of this CD-ROM, I give examples of how these tools and what we have learnt from their use might help us read the *Tales*, both as a whole and line-by-line. The next section concentrates on what comparison of the analyses of The Wife of Bath's Prologue and The General Prologue might tell us of the whole *Tales*. The separate 'Stemmatic Commentary' on this CD-ROM looks at some 120 readings from The General Prologue, assessing for each the weight of manuscript support for this reading or that.

4.1 The General Prologue and The Wife of Bath's Prologue: one *Tales* or many tales?

Scholars have long argued about the state of the *Tales* as Chaucer left it at his death. Is the *Tales* as we have it a collection of separate publications, which Chaucer 'published' as a series of discrete sections, and had only partially integrated into a single work? Or did he leave a single collection of papers, in various states of readiness and incomplete as a whole, but nevertheless representing a single and conceptually coherent work in progress?

The textual archaeology we have performed on the two sections of text, for The General Prologue and The Wife of Bath's Prologue, might be expected to cast light on this. If we discover that the two sections have distinct manuscript histories, then the case for the first alternative is considerably strengthened. However, if we discover that the one early textual history underlies both sections, the case for the *Tales* as a single collection is strengthened.

The most complete earlier attempt to answer this question, by Manly and Rickert, left no clear answer. Indeed, one can find words in their discussions to support both views. On II, 37 they assert that Chaucer permitted 'single copies of some of his tales' to circulate, and that 'these single copies...were made use of by the scribes who after Chaucer's death attempted to assemble the parts of the unshaped CT'. Four pages later, on II 41, they argue that while there was never 'a completed MS of Chaucer's'

yet all existing texts go back 'to a body of incomplete material, in different stages of composition and only in part put in order and corrected'. However, the import of their classification is that each part of the *Tales* has its own unique textual history, and this has been taken by later scholars to support the notion of separate publication of separate parts of the *Tales*, presumably in Chaucer's lifetime. The mention of The Wife of Bath as ye han rede in the 'Lenvoy de Chaucer a Bukton' is used to support this view.

On the face of it, the dissimilarities between the textual histories of The Wife of Bath's Prologue and The General Prologue, as I have presented them here and in Robinson 1997, appear to support the argument for separate textual history, and therefore separate early publication. El and Hg are very close throughout The General Prologue; but in the first half of The Wife of Bath's Prologue El is with the e group, as is Gg, and then moves very close to Hg in the second half. Gg moves similarly: with e throughout The Wife of Bath's Prologue then close to Hg through The General Prologue. Also, the group of manuscripts we have identified as close to alpha in The General Prologue appear close to **O** rather than to alpha in The Wife of Bath's Prologue (thus, Ad^1 Ad^3 En^3 Tc^1).

However, one could explain such differences as being the result of shifts of exemplar, or of later developments in the textual tradition. These are irrelevant. It is only the early history of the tradition, and specifically the exemplar and its first copies, which are relevant. One can divide the enquiry into two parts:

1. What was the nature of the exemplar underlying all early copies?
2. From study of the existing manuscripts, how many early copies were made and how do the manuscripts trace back to these early copies?

4.1.1 The nature of the exemplar underlying the two sections

Concerning the exemplar from which all existing manuscripts descend, we concluded the following for The Wife of Bath's Prologue:

1. There was one Chaucerian exemplar only. That is, the differences between the manuscripts and their distribution over the tradition do not show evidence of two (or more) Chaucerian versions, with one representing a word-by-word revision of the other.
2. However, this exemplar was not a fair copy, but Chaucer's working copy of the *Tales*. It contained marks for passages to be deleted, as

in the so-called 'added passages'. The different treatment of such authorial markings by the first copyists then accounts for uneven distribution of the passages across the tradition. One might expect that the working copy might contain alternative authorial readings at particular words, and these too might then be distributed unevenly across the tradition (however, there is little evidence for this in The Wife of Bath's Prologue; cf. Solopova 1997).

The analysis of The General Prologue presented here points to exactly the same conclusions. There is no evidence of distinct Chaucerian exemplars, or of one branch of the tradition representing a word-by-word revision by Chaucer of the text. I have pointed out a few readings where Hg and El disagree and where the alternatives may represent Chaucer's own changes-of-mind: see 3.4.3 above. But, as in The Wife of Bath's Prologue, the infrequency of these suggests that what Chaucer left was a clean working copy, so far as the word-by-word text is concerned, with very few instances of authorial reworking of individual words.

The treatment of lines 253-4 in The General Prologue, and their distribution across the manuscripts, can be explained as due to the same factors responsible for the distribution of the 'added passages' in The Wife of Bath's Prologue. That is: Chaucer may have written the lines and then later marked them for deletion. In The Wife of Bath's Prologue, it seems that the first scribes of all except the alpha copy respected these marks for deletion; in the General Prologue, the first scribes except that of the Hg ancestor respected these marks. In both traditions, the 'extra lines ' may have found their way to other manuscripts by contamination. Thus, lines 253-4 were imported into Cx^2 (they are absent from all ab manuscripts preceding Cx^2) presumably from the 'better manuscript' Caxton used in preparation of this edition, a manuscript we know to have been very close to **O**.

Therefore, it can be argued that the one exemplar underlies both The General Prologue and The Wife of Bath's Prologue. This exemplar was Chaucer's own working copy of the *Tales*, the lost archetype we call **O**.

4.1.2 How many early copies were there of O?

The analysis of The Wife of Bath's Prologue suggested that the majority of the manuscripts represent just three lines of descent for the

manuscripts, and so derive from three distinct copies of **O**:

through alpha: El (second half); the exemplar of Cx^2; hence, through ab all the manuscripts of the a and b groups

through cd: Cp and all the manuscripts of the cd groups

through ef: e (Bo^1 Ph^2 and others); f (Ld^2 Ry^2 Bw Ln)

In addition, analysis showed a group (labelled O) of some seventeen manuscripts which could not be allocated to these groups, and so represented an uncertain number of additional lines of descent. These manuscripts were: Ad^1/En^3 Ad^3/Ha^5 Bo^2 Ch El (also e) Gl (also c) Ha^4 Hg Hk Ht Py Ps (also c) Ra^2 Ra^3/Tc^1.

The analysis in this workshop of The General Prologue tradition suggests, likewise, that the majority of manuscripts represent just three lines of descent, and so derive from three distinct copies of **O**:

through alpha: Ad^1 Ad^3 En^3 Tc^1; hence, through ab Ht Py Ra^2 Ry^1; and thence all the manuscripts of the a and b groups

through cd: Cp and all the manuscripts of the cd groups

through e: e (Bo^1 Ph^2)

In addition, analysis showed a group of some ten manuscripts (leaving aside the fragmentary Do) which could not be allocated to these groups, and so represented an uncertain number of additional lines of descent. These manuscripts were: Bo^2 Ch El Gg Ha^4 Hg Ln Ps Ra^3 To^1. Seven of these ten (all but Gg Ln To^1) are among the seventeen classified as O group in The Wife of Bath's Prologue. Within this ten, it appears that El and the trio Ch/Ha^4/Hg may represent two independent lines of descent.

Thus presented, there are striking similarities between the accounts of the early copies of the two sections. However, a judgement based on just these two sections of the *Tales*, amounting to less than ten per cent of the whole and only two of some forty distinct sections, would be unsafe. There are, indeed, differences between the two textual histories, and analysis of further sections of the *Tales* is likely to cast further light on relationships among the ungrouped (or O group) manuscripts. Such further analysis will also clarify whether the inference drawn above, that the differences between the different parts are due to late or local factors, is valid. If it is valid, these differences will not disturb the argument that

there is one textual history, and one exemplar only, underlying the whole *Tales*.

This conclusion would not prove that separate publication of some parts of the *Tales* did not occur in Chaucer's lifetime; just that any such publication has left no trace in the manuscript tradition.

Exercises: now, try these for yourself..

1. 'There is no evidence of distinct Chaucerian exemplars, or of one branch of the tradition representing a word-by-word revision by Chaucer of the text.' If there were any evidence of such a revision, or of distinct Chaucerian exemplars, how would you expect to see it in the tradition? Where would you look for it? (then, try and find it...)
2. Analysis of both The Wife of Bath's Prologue and The General Prologue suggests that revision by Chaucer within **O** shows itself most clearly (and perhaps exclusively) in deletions and additions of whole lines and passages. Examine the other 'whole line' variants (that is, whole lines or sequences of lines) in The General Prologue (hint: use the 'with blk' parameter in VBase searches). Are there any other cases, besides the couplets at 253 and perhaps 638-1, where the absence or presence might be the result of Chaucerian revision?
3. It is suggested above that in rare cases, Chaucer might have altered single words or short phrases in **O** and that these revisions might be reflected in some of the first copies. Can you find any other alternative readings beside those suggested in 3.4.3 above which might result from such a revision? (hint: look at readings in El and not in Hg, and vice versa).
4. Manly and Rickert, on II 41, assert that it is possible to reduce the number of manuscripts needed to establish the text to just 'the ancestors of ... a b c d and the single MSS El, Gg, and Hg, and in connection with them Ha^4 and Ad^3'. In accordance with this, the editors of *The Riverside Chaucer* and of *The Variorum Chaucer* typically collate only Ch Ad^3 El Gg Hg Ha^4 Cp Pw En^1 Ne, with slight variants. In view of what is said above about the likely lines of descent, assess how adequate these choices might prove as the basis for an edition. (hint: what lines of descent appear over-represented by these choices; what under-represented; what not represented at all?)

Notes

1. See further section 4 of this workshop. I first used the description 'New Stemmatics' in a talk to the Dutch Society of Stemmatologists in Wassenaar, The Netherlands, in May 1998. However, like many 'new' things it is not new at all. The pragmatic use of stemmatics proposed in this workshop is very similar to the practice of some modern Italian textual critics, notably Barbi and Contini. According to Paolo Cherchi's description, these scholars used the Lachmann method as an historical tool, to provide a working hypothesis towards comprehension of the transmission of the text, and not as a mechanical technique for determination of readings (Cherchi 1995, 440-44.) The same method is employed by Federico Sanguineti, in the preparation of his new edition of Dante's *Commedia*, the first to be based on study of all 600 extant manuscripts of the poem complete at least one canticle. The Stemmatic Commentary on this CD-ROM attempts to use the findings of the Analysis Workshop in the same spirit.
2. Variant drift seems particularly likely where a text is especially well known, so that copyists who know a 'better' reading (very likely, the original) will introduce that better reading in place of the introduced variant. This leads to the phenomenon (worth a study in itself) of later texts actually being 'better'.
3. On the parallel history of stemmatics and evolutionary biology, see the outline in Robinson and O'Hara, 1993.
4. This distinctive system of punctuation is discussed by Elizabeth Solopova (forthcoming), who argues that it is scribal and contrasts it with the different system found in El and Hg.
5. See Doyle 1995, 60-65. It is notable that in just the two pages of the Kk fragment (reproduced by Doyle) there are several spellings which are characteristic of scribe b and rare elsewhere: e.g. weere (19 times in Hg in The General Prologue, and only once in any other manuscript) and compaignye, the usual spelling of El Hg (and Ch!) and rare elsewhere. For these spellings see the spelling databases.