

The phylogeny of *The Canterbury Tales*

Geoffrey Chaucer's *The Canterbury Tales* survives in about 80 different manuscript versions¹. We have used the techniques of evolutionary biology to produce what is, in effect, a phylogenetic tree showing the relationships between 58 extant fifteenth-century manuscripts of "The Wife of Bath's Prologue" from *The Canterbury Tales*. We found that many of the manuscripts fall into separate groups sharing distinct ancestors.

Manuscripts such as these were created by copying, directly or indirectly, from the original material (written, in the case of *The Canterbury Tales*, in the late fourteenth century). In the process of copying, the scribes made (deliberately or otherwise) changes, which were themselves copied. Textual scholars have developed a system for reconstructing the relationships between textual traditions by analysing the distribution of these shared changes, and have constructed family trees (stemmata) on the basis of the results, with the ultimate aim of establishing precisely what the author actually wrote. This analysis is carried out manually and is feasible only for a few manuscripts of short texts. The sheer quantity of information in a tradition the size of *The Canterbury Tales* defeats any system of manual analysis.

However, the principle of historical reconstruction is similar to the computerized techniques used by evolutionary biologists to reconstruct phylogenetic trees of different organisms using sequence data. We therefore applied phylogenetic techniques to *The Canterbury Tales* using the 850 lines of 58 surviving fifteenth-century manuscripts of “The Wife of Bath’s Prologue”. We believe this to be the first full tradition of a major work to be analysed in this manner.

It may be inappropriate to impose a tree-like structure on such data sets, so we used the method of split decomposition implemented in the program **SplitsTree**², in addition to the cladistic analysis of PAUP³. Figure 1 shows a SplitsTree analysis of 44 of

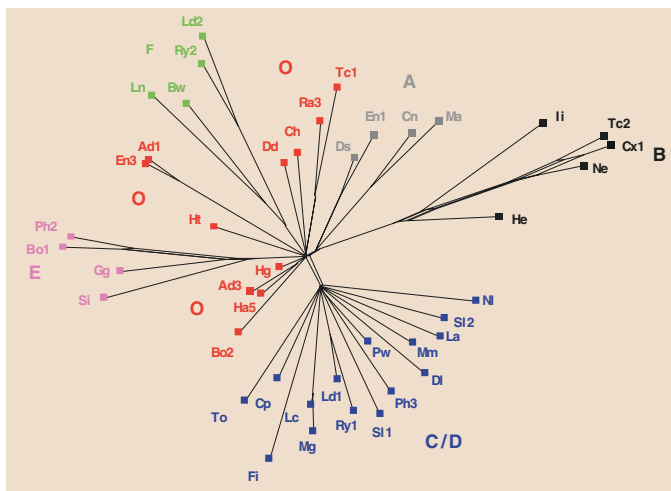


Figure 1 SplitsTree analysis of 44 manuscripts of "The Wife of Bath's Prologue" from Chaucer's *The Canterbury Tales*⁴. The two- or three-character codes indicate individual manuscripts, whereas the large capitals indicate groups of manuscripts, which are coloured the same.

the 58 manuscripts. Very similar results were given by PAUP (not shown). Several manuscripts form groups (A, B, C/D, E and F), each descended from a single and distinct common ancestor. The remaining 14 manuscripts were removed from the analysis shown in Fig. 1, as they were likely to have been copied from more than one exemplar, either by deliberate conflation of readings or by changing the exemplar during the course of copying. These manuscripts were identified by comparison of the trees generated with different regions of the text, which showed that their position in the analysis varied dramatically depending on which region was used. The central point is likely to represent the ancestor of the whole tradition. The manuscripts grouped as O are particularly crucial; their position near to the centre suggests that they all descend from Chaucer's original, and may therefore contain crucial evidence about this original. However, most of them have been ignored by scholars.

From this analysis and other evidence, we deduce that the ancestor of the whole tradition, Chaucer's own copy, was not a finished or fair copy, but a working draft containing (for example) Chaucer's own

notes of passages to be deleted or added, and alternative drafts of sections. In time, this may lead editors to produce a radically different text of *The Canterbury Tales*. These results also demonstrate the power of applying phylogenetic techniques, and particularly split decomposition, to the study of large numbers of different versions of sizeable texts.

Adrian C. Barbrook, Christopher J. Howe
Department of Biochemistry,
University of Cambridge,
Tennis Court Road, Cambridge CB2 1QW, UK
e-mail: c.j.howe@bioc.cam.ac.uk

Norman Blake
Humanities Research Institute,
Arts Tower, University of Sheffield,
Sheffield S10 2TN, UK

Peter Robinson
The Division of Learning Development,
Gateway House, De Montfort University,
Leicester LE1 5XY, UK

1. Blake, N. F. *The Textual Tradition of The Canterbury Tales* (Edward Arnold, London, 1985).
2. Huson, D. H. *Bioinformatics* 14, 68–73 (1998).
3. Swofford, D. L. *PAUP Version 3.1.1*. (Smithsonian Institute, Washington DC, 1993).
4. Robinson, P. M. W. in *The Canterbury Tales Project: Occasional Papers Vol. II* (eds Blake, N. F. & Robinson, P. M. W.), 69–132 (Office for Humanities Communication, London, 1997).

Green-wave phenology

Phenology, the traditional study of seasonal plant and animal activity driven by environmental factors, has found new relevance in research into global climate change^{1,4}. Global phenology research so far has concentrated on measurements obtained by satellites, downplaying connections of these measures to information obtained on the surface^{1,4}, perhaps because of a lack of conventional, surface-based phenological data. However, an integration of conventional

and satellite-derived measures is needed to understand better the mid-latitude spring onset of photosynthesis, known as the ‘green wave’ or ‘green-up’^{5–7}. Here I show how a surface-based green-wave model can extend the monitoring of climatic variability back to 1900, providing a longer-term context for the more limited, recent data obtained from satellites.

Simple events, such as flower blooms or insect hatchings, offer clues to the workings of complex processes⁸. The spring green wave signals the start of the growing season and profoundly influences subsequent

productivity⁷; plant activity is more responsive to spring weather than to other seasons. Thus, indices that measure the onset of the green wave are ideal biological measures of climatic variability. Despite these attributes, traditional phenological studies are typically local-scale agricultural projects, rarely involving ecosystem- or global-scale processes. In contrast, satellite remote sensing has been widely used for biospheric monitoring, including for phenology^{1,2,4,9}.

The daily global coverage by satellites makes them essential to investigations of green-up. However, remote-sensing data