Historical reasoning and abductive inference in phylogenetic reconstruction

MINAKA Nobuhiro

National Institute for Agro-Environmental Sciences
and
Graduate School of Agricultural and Life Sciences, The University of Tokyo

minaka@affrc.go.jp
http://cse.niaes.affrc.go.jp/minaka/
http://d.hatena.ne.jp/leeswijzer/
Understanding Biodiversity
朝日カルチャーセンター（October ~ December 2005、新宿）
三中信宏：公開講座「進化する生物（全5回）」
Organizing the knowledge of diversity
Organizing the knowledge of diversity

Systematics [in general]

= the science of systematization
Organizing the knowledge of diversity

**Systematics** [in general]

= the science of systematization

1) **Taxonomy**

= categorization by similarity
Organizing the knowledge of diversity

**Systematics** [in general]

= the science of systematization

1) **Taxonomy**

= categorization by similarity

2) **Phylogenetics**

= inference of history
William Whewell (1840),

*Philosophy of the Inductive Sciences.* (3 volumes)
William Whewell (1840),
*Philosophy of the Inductive Sciences.* (3 volumes)

**Classificatory Sciences**

**Palaetiological Sciences**
Taxonomy *versus* Phylogenetics

Working on the same objects with different purposes and methods
Phylogenetic tree
Taxonomic map
Ernst Haeckel (1866), *Generelle Morphologie der Organismen*
Human genealogical trees

Christiane Klapisch-Zuber (2003)
Le tableau généalogique (9th century)
Arbor consanguinitatis (9th century)
Tree of Knowledge

arbor scientiae (13th century)
Tree of Art

Astrit Schmidt-Burkhardt (2005)
Tree of Technology

Curt Brandis (2005)
The Canterbury Tales
The phylogeny of The Canterbury Tales

Geoffrey Chaucer's The Canterbury Tales survives in about 80 different manuscript versions. 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 defies 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, in addition to the cladistic analysis of PAUP.

Figure 1 shows a SplitsTree analysis of 44 of 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 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 sizable 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 5SV, UK

Aviezer Tucker (2004)

Our Knowledge of the Past:

A Philosophy of Historiography

(Cambridge University Press)

Historiographic sciences
Historiographic sciences

1) biblical criticism
2) classical philology
3) comparative linguistics
4) evolutionary biology
The common nature of historiographic sciences

- similar effects that preserved information about their common cause or causes

- the probable reconstruction of intervening stages between the common cause or causes and the similar effects
common cause

reconstruction

abductive inference
abductive inference

Charles Sanders Peirce
- deduction
- induction
- abduction
Abductive inference

1) D is a collection of data.
2) H explains D.
3) H is best among others.
4) Choose H.
第3回最先端育種セミナー（17～19 November 2005，神戸）
〈家畜・水産動物における遺伝的多様性研究—その目的と新しい方法論—〉
三中信宏：生物進化の歴史を推定する—分子系統樹を構築する理論と方法
Global optimum

Optimality criterion

Local optimum

initial tree

initial tree

Tree space
phylogenetics
The Tree of Life
Theodosius Dobzhansky's famous epigram (1973)

*Nothing in biology makes sense except in the light of evolution*
Johann Sebastian Bach's newly discovered cantata (2005)

Alles mit Gott und nichts ohn' ihn
(BWV 1127)

Everything with God,
and nothing without Him
Alles mit Gott und nichts ohn' ihn
Alles mit Gott und nichts ohn' ihn

Alles mit Stammbaum

und nichts ohn' ihn
Everything with *The Tree of Life*, and nothing without Him.
Thank you very much!

ご清聴ありがとうございました