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/

Nishinomiya-Yukawa International & Interdisciplinary Symposium
〈What is Life? - The Next 100 Years of Yukawa's Dream〉
15 - 20 October 2007, CO-OP Inn Kyoto, Kyoto

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）」

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

Nishinomiya-Yukawa International & Interdisciplinary Symposium
〈What is Life? - The Next 100 Years of Yukawa's Dream〉
15 - 20 October 2007, CO-OP Inn Kyoto, Kyoto

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

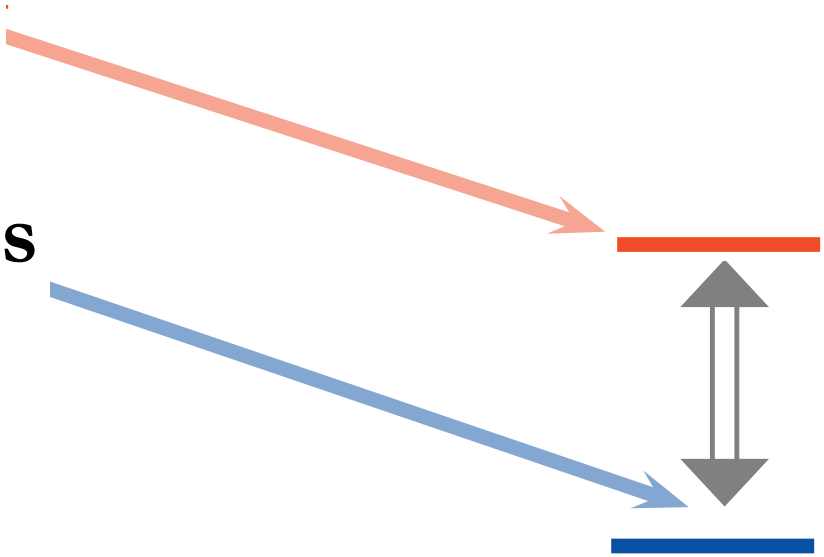
William Whewell (1840), *Philosophy of the Inductive Sciences*. (3 volumes)

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

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）」

Taxonomic map

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）」

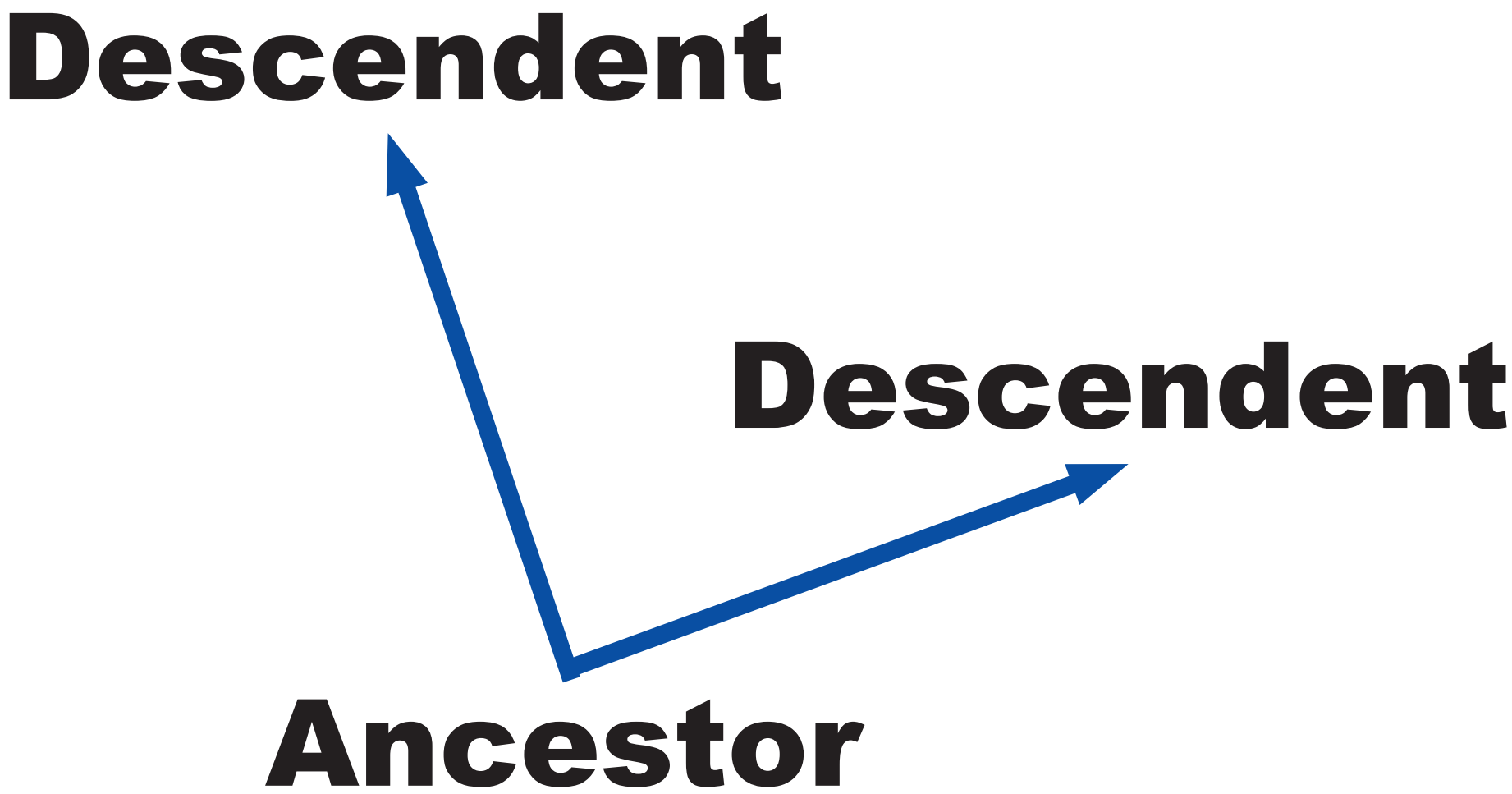
日本生物地理学会第 62 回年次大会シンポジウム
〈進化と系譜：ツリー，ネットワーク，視覚言語リテラシー〉
（8 April 2007，立教大学）

日本生物地理学会第 62 回年次大会シンポジウム
〈進化と系譜：ツリー，ネットワーク，視覚言語リテラシー〉
（8 April 2007，立教大学）

日本生物地理学会第 62 回年次大会シンポジウム
〈進化と系譜：ツリー，ネットワーク，視覚言語リテラシー〉
（8 April 2007，立教大学）

Ernst Haeckel (1866), *Generelle Morphologie der Organismen*

日本生物地理学会第 62 回年次大会シンポジウム
〈進化と系譜：ツリー，ネットワーク，視覚言語リテラシー〉
（8 April 2007，立教大学）


Descendent
Descendent
Ancestor

Human genealogical trees

Christiane Klapisch-Zuber (2003)

日本生物地理学会第 62 回年次大会シンポジウム
〈進化と系譜：ツリー，ネットワーク，視覚言語リテラシー〉
（8 April 2007，立教大学）

Le tableau généalogique (9th century)

日本生物地理学会第 62 回年次大会シンポジウム
〈進化と系譜：ツリー，ネットワーク，視覚言語リテラシー〉
（8 April 2007，立教大学）

Arbor consanguinitatis (9th century)

日本生物地理学会第 62 回年次大会シンポジウム
〈進化と系譜：ツリー，ネットワーク，視覚言語リテラシー〉
（8 April 2007，立教大学）

Tree of Knowledge

Tree of knowledge (18th century)

arbor scientiae (13th century)

日本生物地理学会第 62 回年次大会シンポジウム
〈進化と系譜：ツリー，ネットワーク，視覚言語リテラシー〉
（8 April 2007，立教大学）

Tree of Art

Astrit Schmidt-Burkhardt (2005)

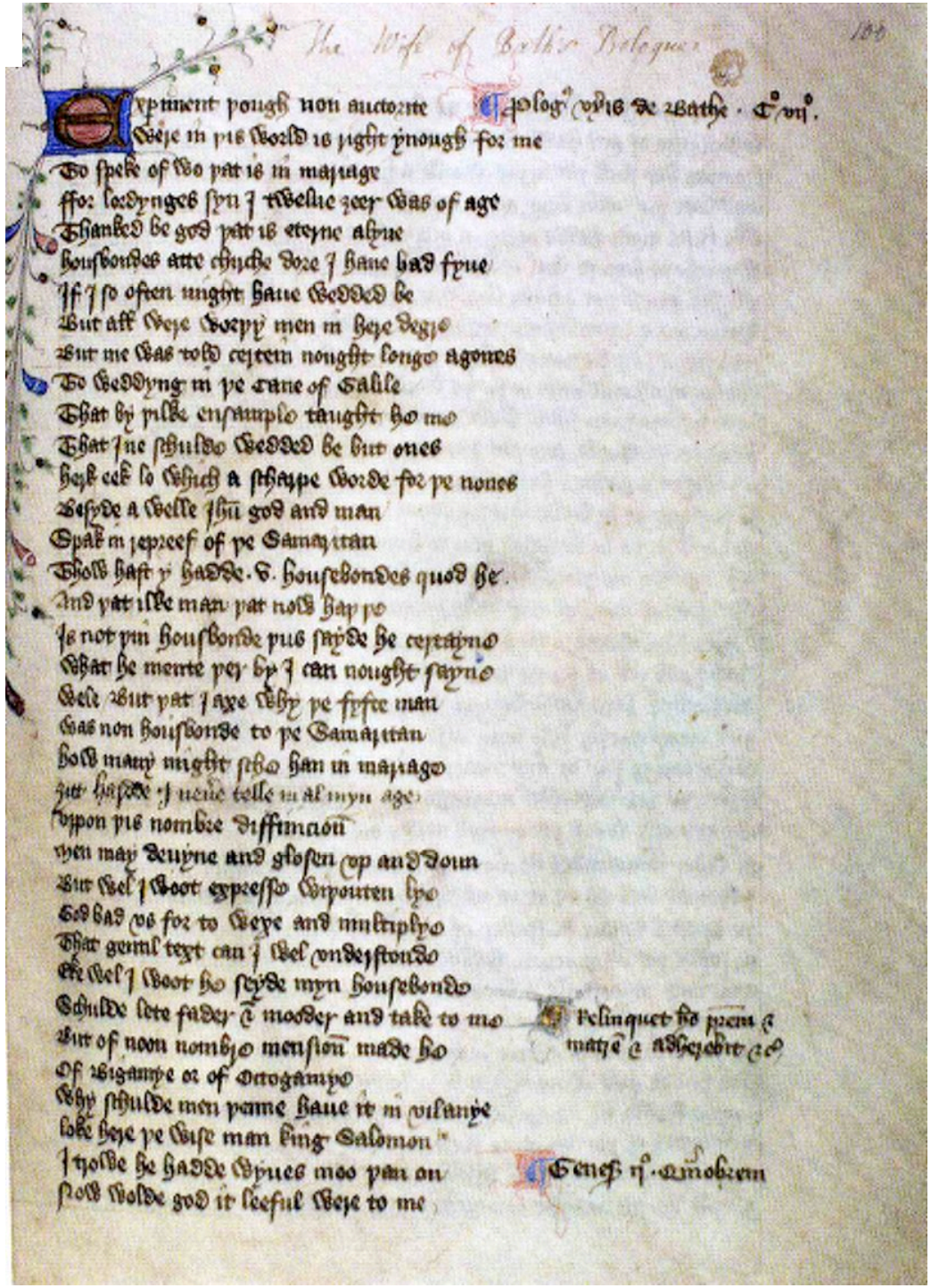
日本生物地理学会第 62 回年次大会シンポジウム
〈進化と系譜：ツリー，ネットワーク，視覚言語リテラシー〉
（8 April 2007，立教大学）

Tree of Technology

Curt Brandis (2005)

日本生物地理学会第 62 回年次大会シンポジウム
〈進化と系譜：ツリー，ネットワーク，視覚言語リテラシー〉
（8 April 2007，立教大学）

The Canterbury Tales



The Wife of Bath's Prologue

Experiment pough non auctorite
Were in pis world is right ynough for me
To speke of wo pat is in marriage
For lordynges syn I twelue yeer was of age
Thanked be god pat is eterne alyue
Housbondes atte chirche dore I have had fyue
If I so often might have wedded be
And yet I wyll yeppy men in heye degre
But me was told certeyn nought longe agones
To weddyng in pe cane of Galile
That bi pilke ensample taught ho me
That I ne shulde wedded be but ones
Herk eek lo which a sharpe worde for pe nones
Besyde a welle jhu god and man
Spake in reprefe of pe Samaritan
Thow hast y hadde .v. housebondes quod he
And pat ilke man pat now hap pe
Is not pin housbonde pus sayde he certayne
What he meente pey by I can nought sayne
Wele but pat I axe why pe fyfte man
Was non housbonde to pe Samaritan
How many might sho han in marriage
yet hadde I neue telle in al myn age
Upon pis nombre diffinacion
Men may deuyne and glosen up and down
But wel I woot expresso withouten lye
God bad vs for to wexe and multiplye
That gentil text can I wel vndeystonde
Eke wel I woot he sayde myn housebonde
Shulde lete fadey & modey and take to me
But of noon nombre mencion made he
Of bigamye or of Octogamye
Why shulde men panne hauen it in vilanye
Loke heye pe wise man king Salomon
I trowe he hadde wyues moo pan on
Trow wolde god it leefful weye to me

Prologus vxis de Bathe · Tunc

Relinquet ho patrem & matrem & adherebit eo

Genesis ij · Quomodo

The phylogeny of *The Canterbury Tales*

Geoffrey Chaucer's *The Canterbury Tales* survives in about 80 different manuscript versions[1]. 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 PAUP[3]. 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 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.

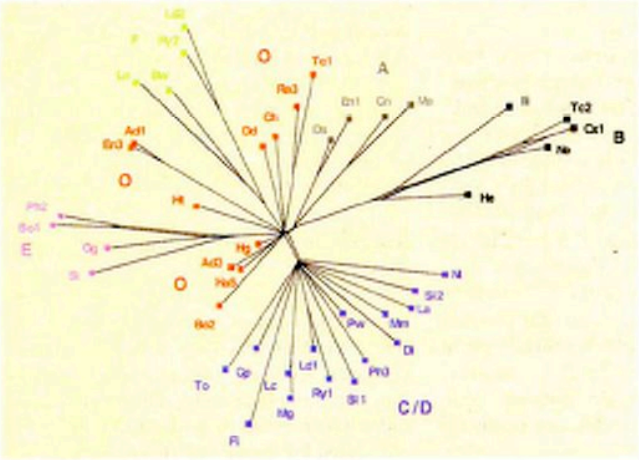


Figure 1 SplitsTree analysis of 44 manuscripts of "The Wife of Bath's Prologue" from Chaucer's *The Canterbury Tales*[4]. The two- or three-character codes indicate individual manuscripts, whereas the large capitals indicate groups of manuscripts, which are coloured the same.

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).

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.

Inference to the best explanation

Global optimum

Local optimum

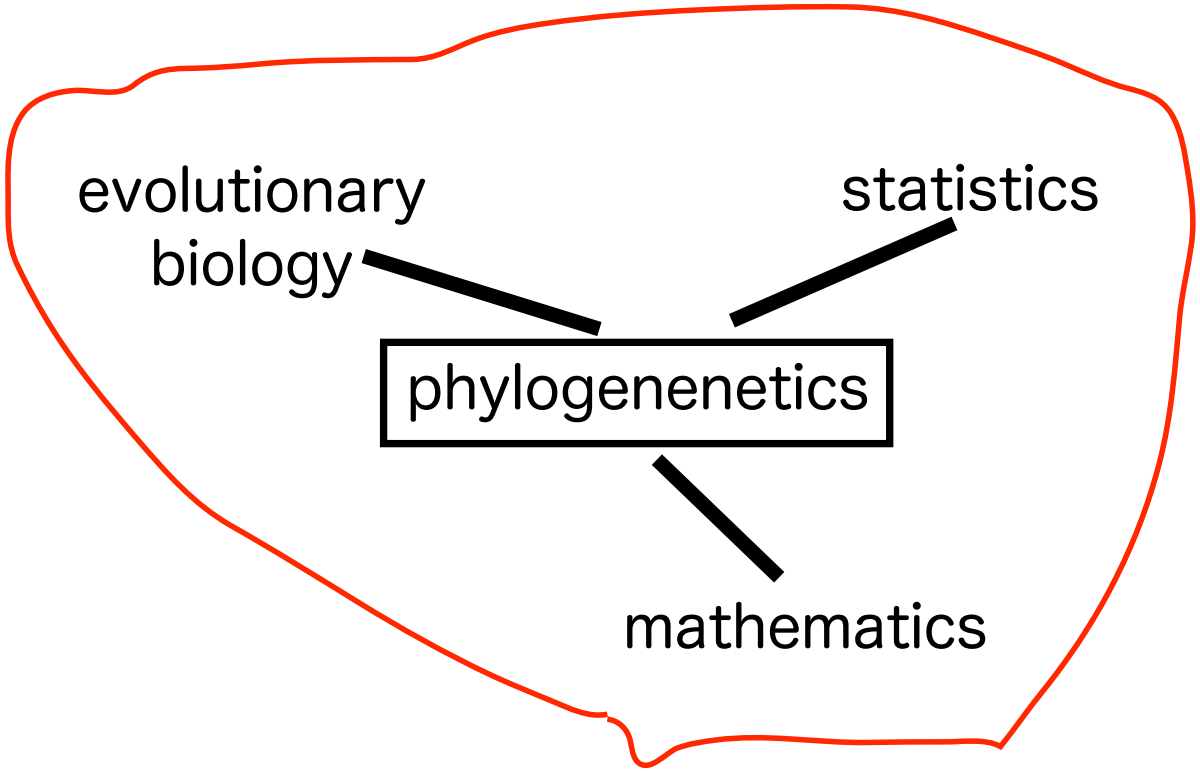
Optimality criterion

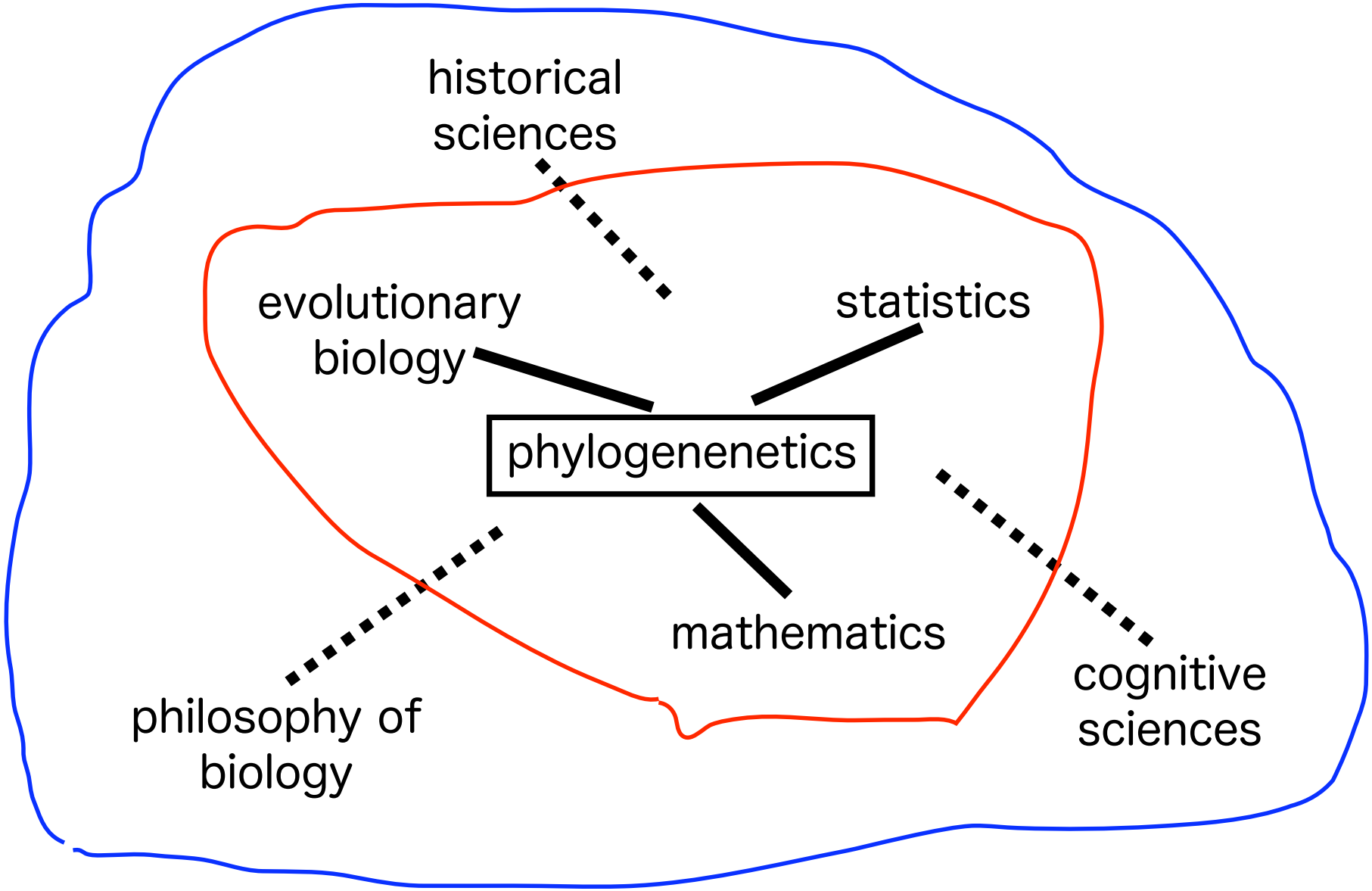
initial tree

initial tree

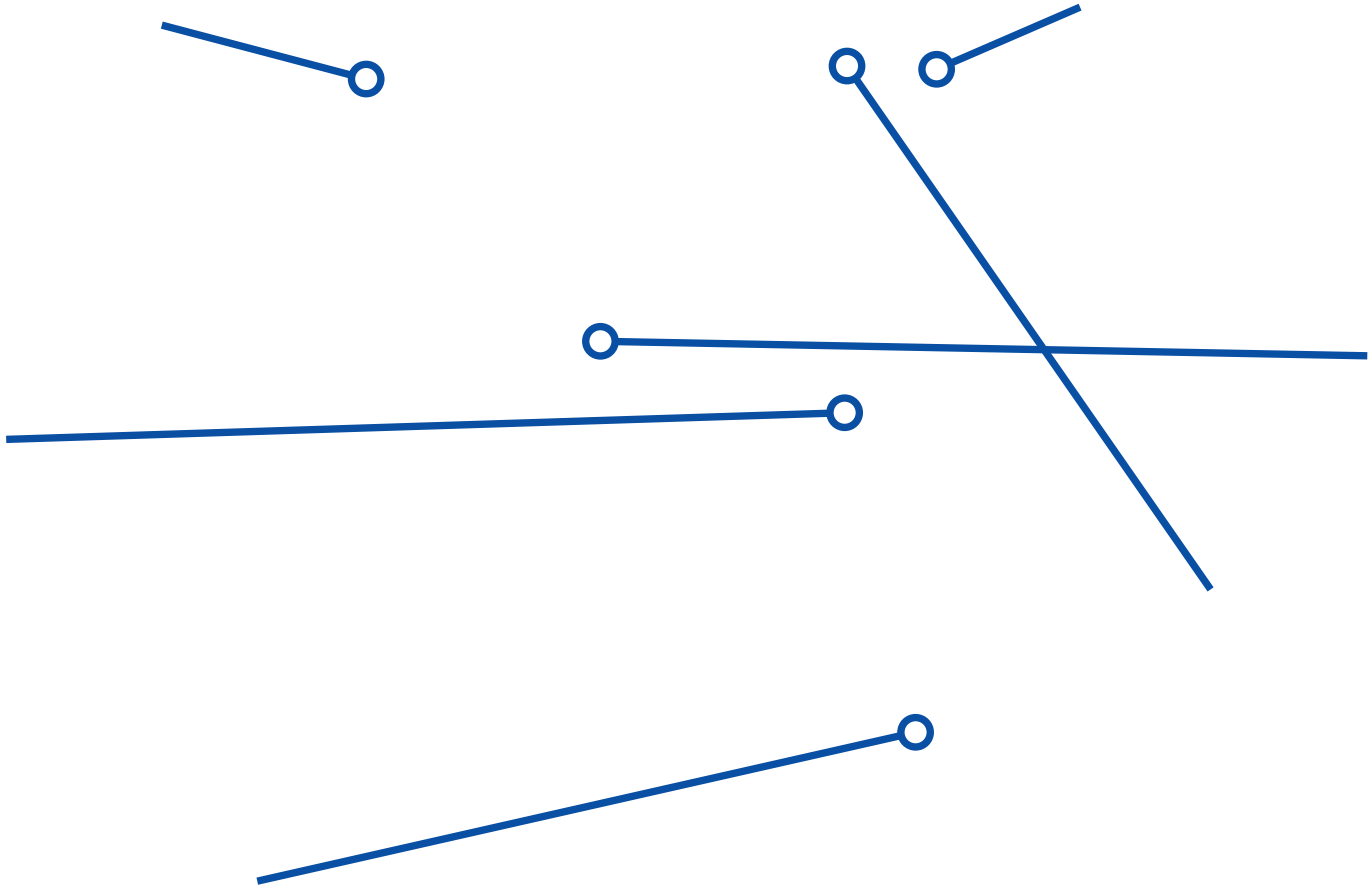
Tree space

phylogenenetics


evolutionary
biology
statistics
phylogenenetics
mathematics


historical sciences
evolutionary biology
statistics
phylogenenetics
mathematics
philosophy of biology
cognitive sciences

The Tree of Life

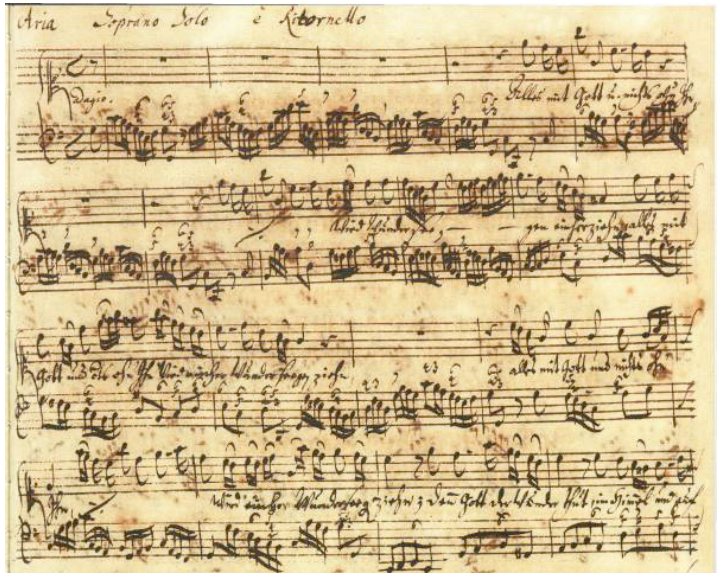


第 3 回最先端育種セミナー（17 ～ 19 November 2005，神戸）
〈家畜・水産動物における遺伝的多様性研究―その目的と新しい方法論―〉
三中信宏：生物進化の歴史を推定する―分子系統樹を構築する理論と方法

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

Nishinomiya-Yukawa International & Interdisciplinary Symposium
〈What is Life? - The Next 100 Years of Yukawa's Dream〉
15 - 20 October 2007, CO-OP Inn Kyoto, Kyoto

Alles mit Gott und nichts ohn' ihn

Alles mit <u>Gott</u> und nichts ohn' ihn

Alles mit <u>Stammbaum</u>

und nichts ohn' ihn

Alles mit <u>Gott</u> und nichts ohn' ihn

Alles mit <u>Stammbaum</u>

und nichts ohn' ihn

Everything with <u>The Tree of Life</u>, and nothing without Him.


Thank you very much!
ご清聴ありがとうございました
http://www.365pic.com