WARD C. WHEELER

SYSTEMATICS

A Course of Lectures

WILEY-BLACKWELL

Systematics

Systematics: A Course of Lectures

Ward C. Wheeler

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Registered office: John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial offices: 9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data has been applied for
9780470671702 (hardback)
9780470671696 (paperback)

A catalogue record for this book is available from the British Library.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Set in Computer Modern 10/12pt by Laserwords Private Limited, Chennai, India

1 2012

For

Kurt Milton Pickett
(1972–2011)
Ave atque vale

Contents

Preface

These notes are intended for use in an advanced undergraduate or introductory level graduate course in systematics. As such, the goal of the materials is to encourage knowledge of core systematic literature (*e.g.* works of Aristotle, Linné, Mayr, Hennig, Sokal, Farris, Kluge, Felsenstein) and concepts (*e.g.* Classification, Optimality, Optimization, Trees, Diagnosis, Medians, Computational Hardness). A component of this goal is specific understanding of methodologies and theory (*e.g.* Cluster Analysis, Parsimony, Likelihood, String Match, Tree Search). Exercises are provided to enhance familiarity with concepts and common analytical tools. These notes are focused on the study of pattern in biodiversity; notions of process receive limited attention and are better discussed elsewhere.

Each chapter covers a topic that could easily be the subject of an entire book-length treatment and many have. As a result, the coverage of large literatures is confined to what I think could be covered in a lecture or two, but may seem brief, idiosyncratic, but hopefully not too superficial. These notes are not meant to be the last word in systematics, but the first.

Students should have basic knowledge of biology and diversity including anatomy and molecular genetics. Some knowledge of computation, statistics, and linear algebra would be nice but not required. Relevant highlights of these fields are covered where necessary.

Using these notes

This is not a fugue. In most cases, sections can be rearranged, or separated entirely without loss of intelligibility. Several sections do build on others (*e.g.* sections on tree searching and support), while others can be deleted entirely if students have the background (*e.g.* sections on computational and statistical basics). The book was developed for a single semester course and, in general, each chapter is designed to be covered in a single 90 minute class period. The chapters on Parsimony, Likelihood, Posterior Probability, and Tree Searching are exceptions, spanning two such classes.

Exercises are of three types: those that can be worked by hand, those that require computational aids, and lastly those that are more suited to larger projects or group work. Hopefully, they are useful.

Acknowledgments

I would like to thank the support of the National Science Foundation (NSF) as well as the National Aeronautics and Space Administration (NASA) and the Defense Advanced Research Projects Agency (DARPA) for supporting the research that has gone into many sections of these notes.

John Wenzel and Adam Kashuba were supportive and persistent in urging me to publish these notes.

All errors, polemics, and disturbing asides are of course my own.

I thank the following people for offering expert advice in reviewing text, suggesting improvements, and identifying errors:

- Benjamin de Bivort, Ph.D.
  Rowland Institute at Harvard

- James M. Carpenter, Ph.D.
  Division of Invertebrate Zoology
  American Museum of Natural History

- Megan Cevasco, Ph.D.
  Coastal Carolina University

- Ronald M. Clouse, Ph.D.
  Division of Invertebrate Zoology
  American Museum of Natural History

- Louise M. Crowley, Ph.D.
  Division of Invertebrate Zoology
  American Museum of Natural History

- John Denton, M.A.
  Division of Vertebrate Zoology
  American Museum of Natural History

- James S. Farris, Ph.D.
  Molekylärsystematiska laboratoriet
  Naturhistoriska riksmuseet

- John V. Freudenstein, Ph.D.
  Director of the Herbarium and Museum of Zoology
  Herbarium, 1350 Museum of Biological Diversity

- Darrel Frost, Ph.D.
  Division of Vertebrate Zoology
  American Museum of Natural History

- Taran Grant, Ph.D.
  Departamento de Zoologia
  Instituto de Biociências
  Universidade de São Paulo

- Gonzalo Giribet, Ph.D.
  Department of Organismic and Evolutionary Biology
  Museum of Comparative Zoology
  Harvard University

- Pablo Goloboff, Ph.D.
  Instituto Superior de Entomologa
  Miguel Lillo

- Lin Hong, M.S.
  Division of Invertebrate Zoology
  American Museum of Natural History

- Jaakko Hyvönen, Ph.D.
  Plant Biology (Biocenter 3)
  University of Helsinki

- Daniel Janies, Ph.D.
  Biomedical Informatics
  The Ohio State University

- Isabella Kappner, Ph.D.
  Division of Invertebrate Zoology
  American Museum of Natural History

- Lavanya Kannan, Ph.D.
  Division of Invertebrate Zoology
  American Museum of Natural History

- Arnold G. Kluge, Ph.D.
  Museum of Zoology
  The University of Michigan

- Nicolas Lucaroni, B.S.
  Division of Invertebrate Zoology
  American Museum of Natural History

- Brent D. Mishler, Ph.D.
  Department of Integrative Biology
  and Jepson Herbaria
  University of California, Berkeley

- Jyrki Muona, Ph.D.
  Division of Entomology
  Zoological Museum
  University of Helsinki

- Paola Pedraza, Ph.D.
  Institute of Systematic Botany
  New York Botanical Garden

- Norman Platnick, Ph.D.
  Division of Invertebrate Zoology
  American Museum of Natural History
- Christopher P. Randle, Ph.D.
  Department of Biological Sciences
  Sam Houston State University
- Randall T. Schuh, Ph.D.
  Division of Invertebrate Zoology
  American Museum of Natural History
- William Leo Smith, Ph.D.
  Department of Zoology
  The Field Museum of Natural History
- Katherine St. John, Ph.D.
  Dept. of Mathematics & Computer Science
  Lehman College, City University of New York
- Alexandros Stamatakis, Ph.D.
  Heidelberg Institute for Theoretical Studies (HITS GmbH)
- Andrés Varón, Ph.D.
  Jane Street Capital
- Peter Whiteley, Ph.D.
  Division of Anthropology
  American Museum of Natural History

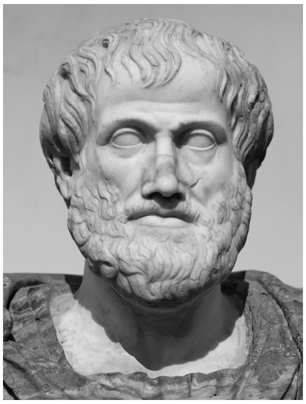
List of Algorithms

Part I

Fundamentals

Chapter 1

History



Roman bust of Aristotle
(384–322 BCE)

Ibn Rushd (Averroes)
(1126–1198)

Systematics has its origins in two threads of biological science: classification and evolution. The organization of natural variation into sets, groups, and hierarchies traces its roots to Aristotle and evolution to Darwin. Put simply, systematization of nature can and has progressed in absence of causative theories relying on ideas of "plan of nature," divine or otherwise. Evolutionists (Darwin, Wallace, and others) proposed a rationale for these patterns. This mixture is the foundation of modern systematics.

Originally, systematics was natural history. Today we think of systematics as being a more inclusive term, encompassing field collection, empirical comparative biology, and theory. To begin with, however, taxonomy, now known as the process of naming species and higher taxa in a coherent, hypothesis-based, and regular way, and systematics were equivalent.

1.1 Aristotle

Systematics as classification (or taxonomy) draws its Western origins from Aristotle[1]. A student of Plato at the Academy and reputed teacher of Alexander the Great, Aristotle founded the Lyceum in Athens, writing on a broad variety of topics including what we now call biology. To Aristotle, living things (*species*) came from nature as did other physical classes (*e.g.* gold or lead). Today, we refer to his classification of living things (Aristotle, 350 BCE) that show similarities with the sorts of classifications we create now. In short, there are three features of his methodology that we recognize immediately: it was functional, binary, and empirical.

Aristotle's classification divided animals (his work on plants is lost) using functional features as opposed to those of habitat or anatomical differences: "Of land animals some are furnished with wings, such as birds and bees." Although he recognized these features as different in aspect, they are identical in use.

[1]Largely through translation and commentary by Ibn Rushd (Averroes).

Systematics: A Course of Lectures, First Edition. Ward C. Wheeler.

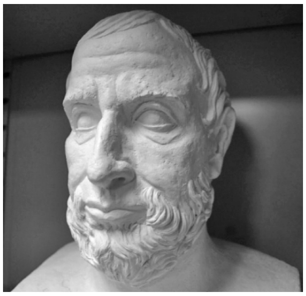
Features were also described in binary terms: "Some are nocturnal, as the owl and the bat; others live in the daylight." These included egg- or live-bearing, blooded or non-blooded, and wet or dry respiration.

An additional feature of Aristotle's work was its empirical content. Aspects of creatures were based on observation rather than ideal forms. In this, he recognized that some creatures did not fit into his binary classification scheme: "The above-mentioned organs, then, are the most indispensable parts of animals; and with some of them all animals without exception, and with others animals for the most part, must needs be provided." Sober (1980) argued that these departures from Aristotle's expectations (Natural State Model) were brought about (in Aristotle's mind) by errors due to some perturbations (hybridization, developmental trauma) resulting in "terata" or monsters. These forms could be novel and helped to explain natural variation within his scheme.

- Blooded Animals
  - Live-bearing animals
    - humans
    - other mammals
  - Egg-laying animals
    - birds
    - fish
- Non-Blooded Animals
  - Hard-shelled sea animals: Testacea
  - Soft-shelled sea animals: Crustacea
  - Non-shelled sea animals: Cephalopods
  - Insects
  - Bees
- Dualizing species (potential "terata," errors in nature)
  - Whales, seals and porpoises—in water, but bear live young
  - Bats—have wings and can walk
  - Sponges—like plants and like animals.

Aristotle clearly had notions of biological progression (*scala naturae*) from lower (plant) to higher (animals through humans) forms that others later seized upon as being evolutionary and we reject today. Aristotle's classification of animals was neither comprehensive nor entirely consistent, but was hierarchical, predictive (in some sense), and formed the beginning of modern classification.

1.2 Theophrastus



Theophrastus (*c*.371–*c*.287 BCE)

Theophrastus succeeded Aristotle and is best known in biology for his *Enquiry into Plants* and *On the Causes of Plants*. As a study of classification, his work

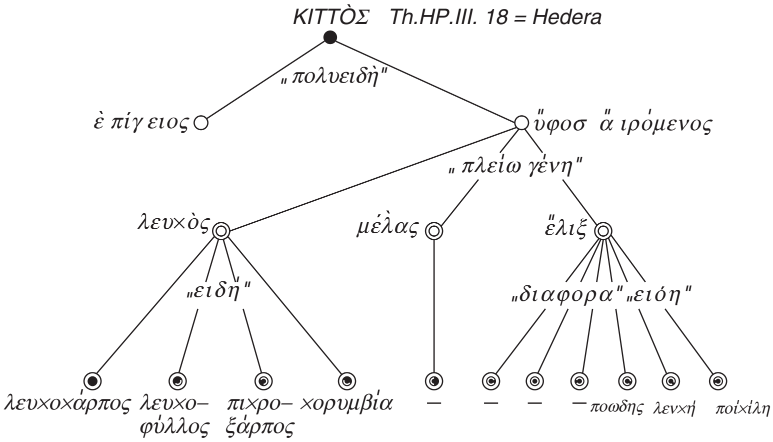
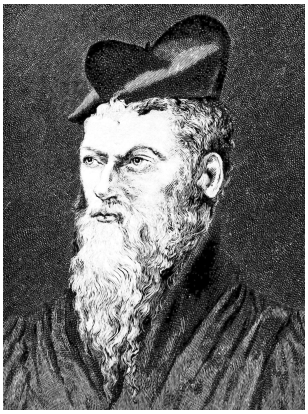
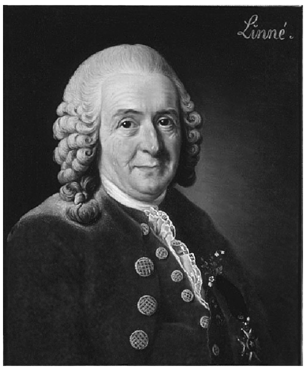


Figure 1.1: Branching diagram after Theophrastus (Vácsy, 1971).



Pierre Belon
(1517–1564)



Carl von Linné
(1707–1778)

on ivy (*κιττός*) discussed extensively by Nelson and Platnick (1981), has been held to be a foundational work in taxonomy based (in part at least) on dichotomous distinctions (*e.g.* growing on ground versus upright) of a few essential features.

Theophrastus distinguished ivies based on growth form and color of leaves and fruit. Although he never presented a branching diagram, later workers (including Nelson and Platnick) have summarized these observations in a variety of branching diagrams (Vácsy, 1971) (Fig. 1.1).

1.3 Pierre Belon

Trained as a physician, Pierre Belon, studied botany and traveled widely in southern Europe and the Middle East. He published a number of works based on these travels and is best known for his comparative anatomical representation of the skeletons of humans and birds (Belon, 1555) (Fig. 1.2).

1.4 Carolus Linnaeus

Carolus Linnaeus (Carl von Linné) built on Aristotle and created a classification system that has been the basis for biological nomenclature and communication for over 250 years. Through its descendants, the current codes of zoological, botanical, and other nomenclature, his influence is still felt today. Linnaeus was interested in both classification and identification (animal, plant, and mineral species), hence his system included descriptions and diagnoses for the creatures he included. He formalized the custom of binomial nomenclature, genus and species we use today.

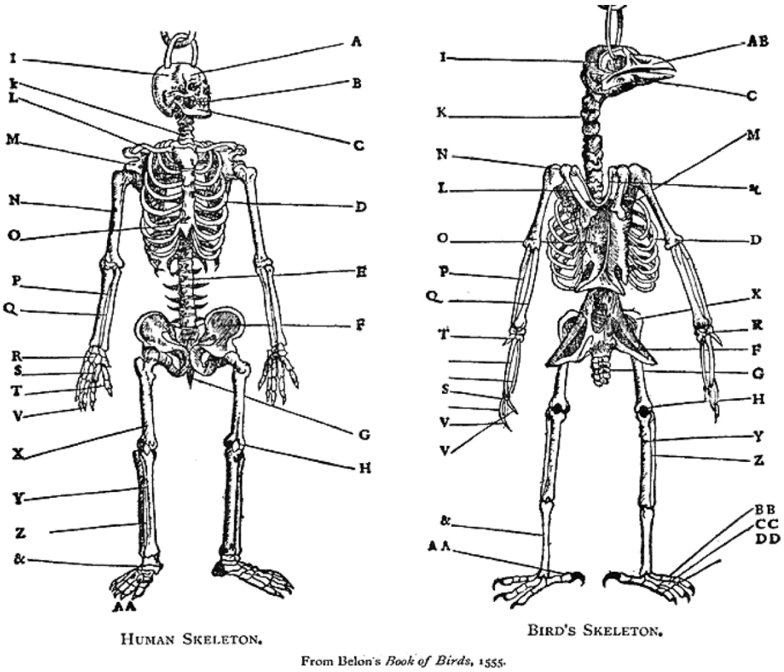


Figure 1.2: Belon's funky chicken (Belon, 1555).

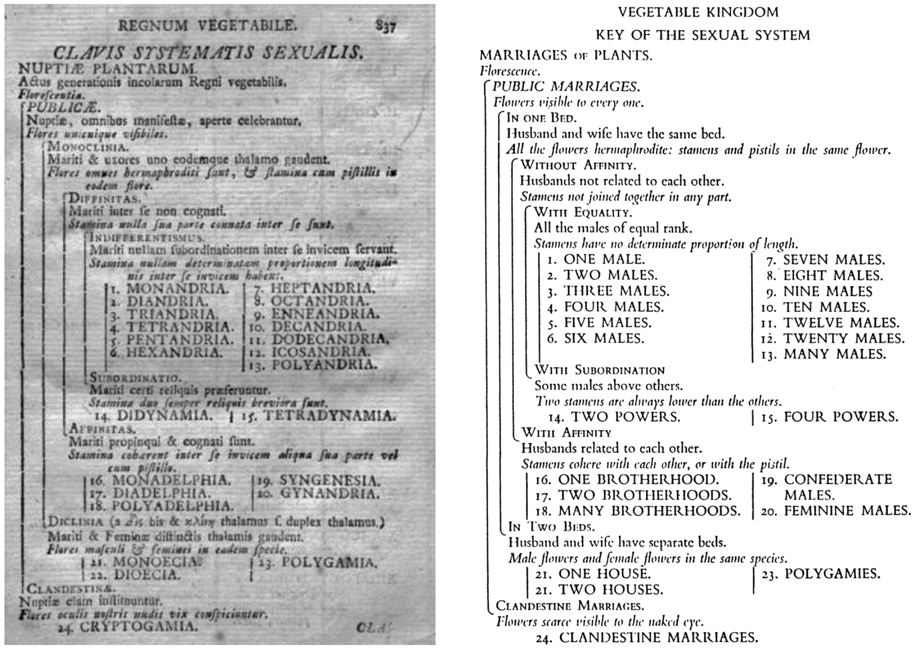
Linnaeus was known, somewhat scandalously in his day, for his sexual system of classification (Fig. 1.3). This was most extensively applied to plants, but was also employed in the classification of minerals and fossils. Flowers were described using such terms as visible (public marriage) or clandestine, and single or multiple husbands or wives (stamens and pistils). Floral parts were even analogized to the foreskin and labia.

Nomenclature for many fungal, plant, and other eukaryote groups[2] is founded on the *Species Plantarum* (Linnaeus, 1753), and that for animals the 10th Edition of *Systema Naturae* (Linnaeus, 1758). The system is hierarchical with seven levels reflecting order in nature (as opposed to the views of Georges Louis Leclerc, 1778 [Buffon], who believed the construct arbitrary and natural variation a result of the combinatorics of components).

- Imperium (Empire)—everything
- Regnum (Kingdom)—animal, vegetable, or mineral
- Classis (Class)—in the animal kingdom there were six (mammals, birds, amphibians, fish, insects, and worms)
- Ordo (Order)—subdivisions of Class
- Genus—subdivisions of Order

[2] For the current code of botanical nomenclature see `http://ibot.sav.sk/icbn/main.htm`.

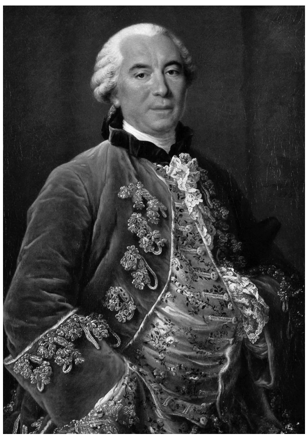
- Species—subdivisions of Genus
- Varietas (Variety)—species varieties or "sub-species."


REGNUM VEGETABILE. 837
CLAVIS SYSTEMATIS SEXUALIS.
NUPTIÆ PLANTARUM.
Actus generationis incolarum Regni vegetabilis.
Florescentia.
PUBLICÆ.
Nuptiæ, omnibus manifestæ, aperte celebrantur.
Flores unicuique visibiles.
MONOCLINIA.
Mariti & uxores uno eodemque thalamo gaudent.
Flores omnes hermaphroditi sunt, & stamina cum pistillis in eodem flore.
DIFFINITAS.
Mariti inter se non cognati.
Stamina nulla sua parte connata inter se sunt.
INDIFFERENTISMUS.
Mariti nullam subordinationem inter se invicem servant.
Stamina nullam determinatam proportionem longitudinis inter se invicem habent.
1. MONANDRIA. | 7. HEPTANDRIA.
2. DIANDRIA. | 8. OCTANDRIA.
3. TRIANDRIA. | 9. ENNEANDRIA.
4. TETRANDRIA. | 10. DECANDRIA.
5. PENTANDRIA. | 11. DODECANDRIA.
6. HEXANDRIA. | 12. ICOSANDRIA.
13. POLYANDRIA.
SUBORDINATIO.
Mariti certi reliquis præferuntur.
Stamina duo semper reliquis breviora sunt.
14. DIDYNAMIA. | 15. TETRADYNAMIA.
AFFINITAS.
Mariti propinqui & cognati sunt.
Stamina cohærent inter se invicem aliqua sua parte vel cum pistillo.
16. MONADELPHIA. | 19. SYNGENESIA.
17. DIADELPHIA. | 20. GYNANDRIA.
18. POLYADELPHIA.
DICLINIA (a δὶς bis & κλίνη thalamus f. duplex thalamus.)
Mariti & Feminæ distinctis thalamis gaudent.
Flores masculi & feminei in eadem specie.
21. MONOECIA. | 23. POLYGAMIA.
22. DIOECIA.
CLANDESTINÆ.
Nuptiæ clam instituuntur.
Flores oculis nostris nudis vix conspiciuntur.
24. CRYPTOGAMIA.

VEGETABLE KINGDOM
KEY OF THE SEXUAL SYSTEM
MARRIAGES OF PLANTS.
Florescence.
PUBLIC MARRIAGES.
Flowers visible to every one.
IN ONE BED.
Husband and wife have the same bed.
All the flowers hermaphrodite: stamens and pistils in the same flower.
WITHOUT AFFINITY.
Husbands not related to each other.
Stamens not joined together in any part.
WITH EQUALITY.
All the males of equal rank.
Stamens have no determinate proportion of length.
1. ONE MALE. | 7. SEVEN MALES.
2. TWO MALES. | 8. EIGHT MALES.
3. THREE MALES. | 9. NINE MALES
4. FOUR MALES. | 10. TEN MALES.
5. FIVE MALES. | 11. TWELVE MALES.
6. SIX MALES. | 12. TWENTY MALES.
13. MANY MALES.
WITH SUBORDINATION
Some males above others.
Two stamens are always lower than the others.
14. TWO POWERS. | 15. FOUR POWERS.
WITH AFFINITY
Husbands related to each other.
Stamens cohere with each other, or with the pistil.
16. ONE BROTHERHOOD. | 19. CONFEDERATE MALES.
17. TWO BROTHERHOODS. | 20. FEMININE MALES.
18. MANY BROTHERHOODS.
IN TWO BEDS.
Husband and wife have separate beds.
Male flowers and female flowers in the same species.
21. ONE HOUSE. | 23. POLYGAMIES.
21. TWO HOUSES.
CLANDESTINE MARRIAGES.
Flowers scarce visible to the naked eye.
24. CLANDESTINE MARRIAGES.

(a) Sexual system for plants (Linnaeus, 1758). (b) English translation.

Figure 1.3: Linnaeus' sexual system for classification (a) with English translation (b) (Linnaeus, 1758).



Georges Louis Leclerc, Comte de Buffon (1707–1788)

The contemporary standard hierarchy includes seven levels: Kingdom, Phylum, Class, Order, Family, Genus, and Species, although other levels are often created as needed to describe diversity conveniently (*e.g.* McKenna and Bell, 1997).

1.5 Georges Louis Leclerc, Comte de Buffon

Georges Louis Leclerc, Comte de Buffon, began his scientific career in mathematics and probability theory[3]. He was appointed director of the *Jardin du Roi* (later *Jardin des Plantes*), making it into a research center.

Buffon is best known for the encyclopedic and massive *Histoire naturelle, générale et particulière* (1749–1788). He was an ardent anti-Linnean, believing taxa arbitrary, hence there could be no preferred classification. He later thought, however, that species were real (due to the *moule intérieur*—a concept at the

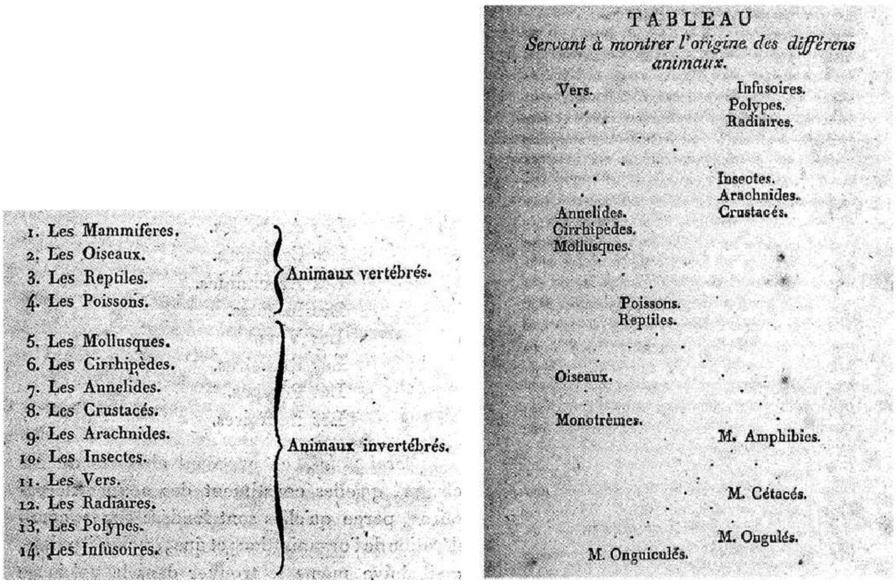
[3] Buffon's Needle: Given a needle of length l dropped on a plane with a series of parallel lines d apart, what is the probability that the needle will cross a line? The solution, $\frac{2l}{d\pi}$ can be used to estimate π.

foundation of comparative biology). Furthermore, Buffon believed that species could "improve" or "degenerate" into others, (*e.g.* humans to apes) changing in response to their environment. Some (*e.g.* Mayr, 1982) have argued that Buffon was among the first evolutionary thinkers with mutable species. His observation that the mammalian species of tropical old and new world, though living in similar environments, share not one taxon, went completely against then-current thought and is seen as the foundation of biogeography as a discipline (Nelson and Platnick, 1981).

Jean-Baptiste Lamarck (1744–1829)

1.6 Jean-Baptiste Lamarck

Jean-Baptiste Lamarck (who coined the word "Biologie" in 1802) believed that classifications were entirely artificial, but still useful (especially if dichotomous). His notion of classification is closer to our modern keys (Nelson and Platnick, 1981). An example of this comes from his *Philosophie zoologique* (Lamarck, 1809), with the division of animal life into vertebrates and invertebrates on the presence or absence of "blood" (Fig. 1.4(a)).



(a) Lamarck's classification of animals. (b) Lamarck's transmutational tree.

Figure 1.4: Lamarck's division of animal life (a) and transmutational tree (b) (Lamarck, 1809).

Lamarck is best known for his theory of Transmutation (Fig. 1.4(b))—where species are immutable, but creatures may move through one species to another based on a motivating force to perfection and complexity, as well as the familiar "use and dis-use." Not only are new species created in this manner, but species can "re-evolve" in different places or times as environment and innate drive allow.

Georges Cuvier
(1769–1832)

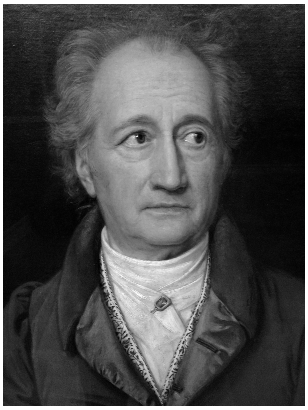
1.7 Georges Cuvier

The hugely influential Léopold Chrétien Frédéric Dagobert "Georges" Cuvier divided animal life not into the *Scala Naturae* of Aristotle, or two-class Vertebrate/ Invertebrate divide of Lamarck, but into four "embranchements": Vertebrata, Articulata, Mollusca, and Radiata (Cuvier, 1812). These branches were representative of basic body plans or "archetypes" derived (in Cuvier's view) from functional requirements as opposed to common genealogical origin of structure. Based on his comparative anatomical work with living and fossil taxa, Cuvier believed that species were immutable but could go extinct, ("catastrophism") leaving an unfillable hole. New species, then, only appeared to be new, and were really migrants not seen before. Cuvier established the process of extinction as fact, a revolutionary idea in its day.

Étienne Geoffroy Saint-Hilaire
(1772–1844)

1.8 Étienne Geoffroy Saint-Hilaire

Although (like Lamarck), the comparative anatomist Étienne Geoffroy Saint-Hilaire is remembered for his later evolutionary views[4], Geoffroy believed that there were ideal types in nature and that species might transform among these immutable forms. Unlike Lamarck, who believed that the actions of creatures motivated transmutation, Geoffroy believed environmental conditions motivated change. This environmental effect was mediated during the development of the organism. He also believed in a fundamental unity of form for all animals (both living and extinct), with homologous structures performing similar tasks. In this, he disagreed sharply with Cuvier and his four archetypes (embranchements), not with the existence of archetypes, but with their number.



Johann Wolfgang von Goethe
(1749–1832)

1.9 Johann Wolfgang von Goethe

With Oken and Owen, Goethe was one of the foremost "ideal morphologists" of the 19th century in that he saw universal patterns underlying the forms of organisms. He coined the term "Morphology" to signify the entirety of an organism's form through development to adult as opposed to "gestalt" (or type—which was inadequate in his view). This is similar to Hennig's concept of the "semaphoront" to represent the totality of characters expressed by an organism over its entire life cycle.

Goethe applied these ideas to the comparative morphology and development of plants (von Goethe, 1790)[5] as Geoffroy did to animals, creating morphological ideals to which all plants ascribed. He claimed, based on observation, that

[4] "The external world is all-powerful in alteration of the form of organized bodies... these are inherited, and they influence all the rest of the organization of the animal, because if these modifications lead to injurious effects, the animals which exhibit them perish and are replaced by others of a somewhat different form, a form changed so as to be adapted to the new environment" (Saint-Hilaire, 1833).

[5] In his spare time, he wrote a book called *Faust*.