

# Basic Concepts and Recent Advances in Microbial Diversity, Taxonomy, Speciation and Evolution



# Basic Concepts and Recent Advances in Microbial Diversity, Taxonomy, Speciation and Evolution

By

Suchitra Godbole and Dhara P. Sachdev

**Cambridge  
Scholars  
Publishing**



Basic Concepts and Recent Advances in Microbial Diversity, Taxonomy,  
Speciation and Evolution

By Suchitra Godbole and Dhara P. Sachdev

This book first published 2024

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data

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

Copyright © 2024 by Suchitra Godbole and Dhara P. Sachdev

All rights for this book reserved. No part of this book may be reproduced,  
stored in a retrieval system, or transmitted, in any form or by any means,  
electronic, mechanical, photocopying, recording or otherwise, without  
the prior permission of the copyright owner.

ISBN (10): 1-5275-7597-7

ISBN (13): 978-1-5275-7597-4

# TABLE OF CONTENTS

	<b>List of Figures</b>	<b>x</b>
	<b>List of Tables</b>	<b>xiii</b>
	<b>Preface</b>	<b>xv</b>
<b>Chapter 1</b>	<b>HISTORY AND DEVELOPMENT OF THE CONCEPT OF SPECIES</b>	<b>1</b>
1.1	Introduction	1
1.2	History and development of the concept of species	2
1.3	Speciation	4
1.4	Biologists' working definition of species	5
1.5	Common names and species	7
1.6	Placement within genera	8
1.7	Abbreviated names	9
1.8	Identification codes	9
1.9	Number of prokaryotic species	10
1.10	Number of eukaryotic species	10
1.11	Conclusion	12
<b>Chapter 2</b>	<b>CONCEPTS OF SPECIES IN EUKARYOTES</b>	<b>13</b>
2.1	Introduction	13
2.2	History of defining the term "species"	14
2.3	Major species concepts	16
2.4	Other species concepts	22
2.5	Conclusion	29

<b>Chapter 3</b>	<b>THE CONCEPT OF MICROBIAL SPECIES</b>	<b>31</b>
3.1	Introduction	31
3.2	Species concepts for prokaryotes	32
3.3	Consensus in species concept	38
3.4	Perspectives and future trends	39
3.5	Conclusion	40
<b>Chapter 4</b>	<b>THEORETICAL MECHANISMS OF SPECIATION</b>	<b>43</b>
4.1	Introduction	43
4.2	Anagenesis	44
4.3	Cladogenesis	44
4.4	Macroevolution theories	45
4.5	Mechanisms of speciation	45
4.6	Four types of speciation	46
4.7	Adaptive radiation	50
4.8	Current status and evolving approaches	53
4.9	Conclusion	54
<b>Chapter 5</b>	<b>SPECIATION THEORIES IN PROKARYOTES</b>	<b>58</b>
5.1	Introduction	58
5.2	Anagenesis	59
5.3	Cladogenesis	59
5.4	Macroevolution theories	60
5.5	Phyletic speciation theory	60
5.6	Species fusion theory	61
5.7	The endosymbiont theory	61
5.8	Gradual speciation	62
5.9	Geographic speciation	63
5.10	Parapatric speciation	63
5.11	Driving forces behind microbial speciation	64
5.12	Mechanisms of microbial speciation – Conjugation, Transduction and Transformation	64

5.13	Domain cell theory of life	65
5.14	Conclusion	66
<b>Chapter 6</b>	<b>METHODS &amp; PARAMETERS FOR PROKARYOTIC SPECIES CIRCUMSCRIPTION</b>	<b>68</b>
6.1	Introduction	68
6.2	Gold standard for species circumscription	69
6.3	Paradigm of new standards for species circumscription	72
6.4	Conclusion	78
<b>Chapter 7</b>	<b>RECENT TRENDS IN MICROBIAL SPECIES CONCEPTS AND TAXONOMY</b>	<b>79</b>
7.1	Introduction	79
7.2	Microbial identification systems	79
7.3	Recent advances in Microbial identification systems	90
7.4	Pure cultures and culture collections centers	91
7.5	Guidelines to the recognition of a prokaryotic species	96
7.6	Future species circumscriptions	97
7.7	Microbial databases used in microbial taxonomy	99
7.8	Conclusion	106
<b>Chapter 8</b>	<b>EVOLUTION OF SPECIES</b>	<b>108</b>
8.1	Introduction	108
8.2	History and development of evolutionary theories	108
8.3	Neo-Darwinism and evolution – Modern Synthesis (MS)	110
8.4	Spontaneous Mutation controversy	116
8.5	Evolution of rates of mutations	117
8.6	Types of selection	121
8.7	Levels of selection	129
8.8	r and K selection	136
8.9	Neutral Evolution	138

8.10	Molecular clocks	142
8.11	Phylogeny and molecular distances	144
8.12	Co-evolution	147
8.13	Molecular evolution	157
8.14	Speciation in sexual and asexual organisms	157
8.15	Origin and stability of diversity	162
8.16	Diversity of secondary metabolites	166
8.17	Conclusion	169
<b>Chapter 9</b>	<b>THE EXPLORATION OF UNCULTIVABLE BACTERIA</b>	<b>170</b>
9.1	Introduction	170
9.2	The great plate count anomaly	170
9.3	The VBNC state	172
9.4	Bacteria that can enter the VBNC state	173
9.5	Characteristics of the cells in the VBNC state	174
9.6	Inducing factors – conditions stimulating VBNC state	176
9.7	Resuscitation from the VBNC state	177
9.8	Reasons for cells to enter the VBNC state	178
9.9	Determination of viability of nonculturable cells	178
9.10	VBNC pathogens	179
9.11	Methods of detection of VBNC bacteria	184
9.12	Strategies for culture of ‘unculturable’ bacteria	189
9.13	Conclusion	196
<b>Chapter 10</b>	<b>MICROBIAL DIVERSITY</b>	<b>198</b>
10.1	Introduction	198
10.2	Overview of methods for estimation of microbial diversity	208
10.3	Measures and indices of microbial diversity	211
10.4	Conservation of biodiversity	232
10.5	Special note on soil biodiversity	239
10.6	Conclusion	242

<b>Chapter 11</b>	<b>METHODS FOR ESTIMATING MICROBIAL DIVERSITY</b>	<b>243</b>
11.1	Introduction	243
11.2	Culture and biochemical-based methods	244
11.3	Molecular-based (nucleic-acid based) methods	246
11.4	OMIC approach	276
11.5	Databases to assist study on microbial diversity	289
11.6	AI/ ML platforms	295
11.7	Note on biorthogonal chemistry	307
11.8	Conclusion	311
	Appendix A	312
	Appendix B	322
<b>Chapter 12</b>	<b>TAXONOMY OF BACTERIA</b>	<b>330</b>
12.1	Introduction	330
12.2	Classification	332
12.3	Nomenclature	339
12.4	Identification	342
12.5	Machine learning in bacterial classification	349
12.6	Case studies to highlight the impact of new technologies on bacterial taxonomy	354
12.7	Conclusion	358
	Appendix C	359
	Bibliography	<b>361</b>

## LIST OF FIGURES

Figure 4-1	Details and Types / Classification / Modes of Speciation
Figure 4-2	Types of Speciation
Figure 4-3	Allopatric Speciation (allo = other, patric = place): New species formed from geographically isolated populations
Figure 4-4	Sympatric (sym = same, patric = place): New species formed from within the range of the ancestral population
Figure 4-5	Peripatric (peri = near, patric = place): New species formed from a small population isolated at the edge of a larger population
Figure 4-6	Parapatric (para = beside, patric = place): New species formed from a continuously distributed population
Figure 4-7	Examples of adaptive radiation.
Figure 6-1	Overview of DNA – DNA hybridization (DDH) method for determination of relatedness between bacterial species
Figure 6-2	Plotted results of ANI <sub>b</sub> versus ANI <sub>m</sub>
Figure 6-3	Plotted values of DDH versus ANI <sub>m</sub>
Figure 6-4	Plotted values of TETRA versus ANI <sub>m</sub>
Figure 8-1	Darwinism, modern synthesis and integrated synthesis
Figure 8-2	Different types of mutations
Figure 8-3	Types of selection
Figure 8-4	Selection at individual level
Figure 8-5	Selection at cellular level
Figure 8-6 (a)	Selection at species level
Figure 8-6 (b)	Selection at species level
Figure 8-6 (c)	Selection at species level
Figure 8-7	Neutral evolution
Figure 8-8	Conceptual diagram showing how increasing diversity can stabilize ecosystem functioning
Figure 8-9	Conceptual model illustrating the insurance hypothesis
Figure 9-1	The Great Plate Count Anomaly
Figure 9-2	Model of mechanisms of co-culture helping unculturable bacteria
Figure 9-3	Diffusion chamber and Microbial trap method for cultivation of unculturable bacteria

Figure 10-1	Significance of biodiversity for sustainable human livelihoods
Figure 10-2	A schematic visualization of the potential of microbial bioprospecting for driving forward the development of sustainable solutions
Figure 10-3	General steps to understand microbial diversity using sequencing of 16S rRNA genes
Figure 10-4	Examples of calculation of beta diversity index (panels A and B)
Figure 10-5	Integrating molecular ecology and systematic conservation planning
Figure 10-6	Process for identifying and evaluating issues for Horizon Scan
Figure 10-7	Soil biodiversity and its global significance
Figure 11-1	A representation of 16S ribosomal RNA gene showing the nine hypervariable regions (in dark gray colour) and constant regions (in light gray)
Figure 11-2	Representation of pyrosequencing approach
Figure 11-3	The ligase-mediated sequencing approach of the Sequence by Oligonucleotide Ligation and Detection (SOLiD) sequencer (Life Technologies)
Figure 11-4	Cluster formation in Illumina NGS sequencing
Figure 11-5	Most common metabarcoding sequencing strategies for each sequencing technology generation- (a) First generation sequencing (Sanger)
Figure 11-6	Types of omics approaches for microbial community's analyses based on starting material
Figure 11-7	Workflow of metagenomics research
Figure 11-8	Schematic illustration representing the standard metagenomic workflow
Figure 11-9	Illustration for a workflow to generate a meta-pangenome for a species
Figure 11-10	Workflow for proteomic analysis of microbial community studies
Figure 11-11	Annotation of data using classical methods and AI
Figure 11-12	A workflow for functional annotation of proteins using AI
Figure 11-13	Overview of the metabolic labelling approach
Figure 11-14	Schematic of a standard BONCAT workflow
Figure 12-1	Tree of life (The three-domain system is a biological classification introduced by Carl Woese et al- in 1990)

- Figure 12-2 Rhizome of life: Root of each gene are represented according to the current reclassification of living organism: eukarya (yellow), archaea (blue), nano archaea (light blue), bacteria (red), CPR (dark red), giant virus (orange) – in grey are genes without identified origin (ORFans)
- Figure 12-3 Proposed roadmap for nomenclature of uncultivated archaea and bacteria – Plan A and B provide two alternatives for inclusion of uncultivated archaea and bacteria into the classical Linnaean nomenclature system
- Figure 12-4 A simple representation of dendrogram with ‘K’ species taken as an outgroup; nodes (branch points) are categorized as internal nodes or HTU (Hypothetical Taxonomic Unit which include G, H, I, J) and external nodes or OTU (Operational Taxonomic Unit which include A, B, C, D, E, F); I, D and E together represent a clade (a clade includes an ancestral species and all its descendants)
- Figure 12-5 Diagrammatic representation of techniques and markers used in polyphasic approach for resolving the bacterial hierarchy
- Figure 12-6 Flow chart describing step by step process for characterization of newly isolated strain
- Figure 12-7 Validated number of bacterial names over the years with advancement in bacterial identification methodologies – the blue box indicates the major advances in molecular biology
- Figure 12-8 Flow chart with basic architecture of automatic bacterial image classification system
- Figure 12-9 Schematic representation of the gradual expansion of the phylogenetic tree of MTB in the bacteria domain

# LIST OF TABLES

Table 1-1	Estimated number of species on Earth
Table 4-1	The Theory of Speciation
Table 6-1	DNA based methodologies exploiting genomic features to understand the bacterial species for phylogeny and functional purpose
Table 7-1	Phenotypic, spectroscopic and genotypic characteristics used in microbial taxonomy
Table 7-2	List of some important culture collections and their holdings
Table 7-3	Some microbial culture collections in the world
Table 8-1	Comparison of types of selections
Table 8-2	Summary of characteristics of r and K selection species
Table 8-3	Few examples of methods classification by type of data
Table 8-4	Few examples of method classification by tree search strategy
Table 9-1	Bacteria described to enter the VBNC state
Table 9-2	Fluorescent dyes used for detection of VBNC bacteria
Table10-1	Microbial-based industries and the UN sustainable development goals
Table 10-2	Categories of diversity measurements
Table 10-3	Application of bacterial species: Examples of bioremediation of organic contaminants in soil with bacteria species
Table 11-1	Advantages and limitations of few molecular-based methods used to study microbial diversity
Table11- 2	Summary of sequencing methods
Table 11-3	Advantages of meta-barcoding approach for few sequencing methods
Table 11-4	Examples of public databases for microbial community analysis- Prevalent microbial sequence databases are listed below with indications of their omics integration and functional assignment integration where applicable
Table11-5	Summary of few ML/DP models and tools for studying microbial communities

Table 12-1	Description of taxon groups in rhizome of life and related terms
Table 12-2	Description of frequently used machine learning algorithms

## PREFACE

Welcome to '**Basics and Advances in Bacterial Speciation, Evolution, Diversity, and Taxonomy.**' This book delves into the fascinating world of microbiology, exploring the intricacies of concepts of bacterial species that can throw light on clear identification, classification and naming of new bacterial taxon with elaboration on evolution, significance of microbial diversity and its conservation to achieve sustainability goals.

Our aim in creating this book is to provide a comprehensive and accessible resource for students, researchers, and enthusiasts interested in understanding the fundamental principles of bacterial speciation, evolution, diversity, and taxonomy. From the basics to the cutting-edge advances, we strive to present a holistic view of the subject matter. Understanding the microbial world around and within us is yet an upcoming theme with several researchers all around the globe. The book covers the latest field of machine learning and artificial intelligence that is being employed in the area of microbial diversity and bacterial taxonomy; this shall provide a new perspective to researcher on significance of integration of computer, engineering and life science domains. With a blend of basic and multidisciplinary approach, this book is an excellent resource for students and research scholars that would like to excel in the area of microbiology, biotechnology, genetics, botany, and zoology. This Reference book focuses on the varied topics related to biodiversity such as speciation, evolution, estimation of microbial diversity, methods to explore microbial diversity and taxonomy of bacteria. The authors have tried their best to have a complete coverage of topics from the syllabi of varsities teaching subjects such as microbiology and biotechnology. As researchers and educators in the field of microbiology, the study of bacteria has always been an exhilarating pursuit for us. Through this book, we hope to inspire the same sense of wonder and curiosity that drives our passion for understanding the remarkable world of bacteria.

The book is divided into twelve major chapters with each chapter detailing one topic. Chapters 1 to 3 elaborate on species concept in eukaryotes and prokaryotes. Chapter 4 and 5 are focused on the different mechanisms of speciation for eukaryotes and prokaryotes, whereas Chapter 6 is on the methods and parameters for bacterial species circumscription. The later

Chapter 7 and 12 throw light on microbial taxonomy tools and techniques. Chapter 8 is dedicated on evolution of species, providing a holistic view on evolutionary theories. The microbes can either be in culturable or unculturable forms and thus the Chapter 9 of this book enlightens the reader on the uncultivable forms of bacteria. The significant part of the book lies on Chapter 10 and 11 that gives in details about microbial diversity measurements and recent advances in approaches to explore the microbial world around and within us.

We invite you to embark on this enlightening journey with us and discover the wonders that lie within the realm of bacterial speciation, evolution, diversity, and taxonomy. May this book inspire your curiosity, expand your knowledge, and encourage you to further explore the captivating world of microbiology.

# CHAPTER 1

## HISTORY AND DEVELOPMENT OF THE CONCEPT OF SPECIES

### 1.1: Introduction

Since the early times of the living world, man has been using classification in his normal life. He classified plants into edible, medicinal, fuel, and other uses after using them for a variety of purposes. Similar to other living kingdoms, the plant kingdom has a hierarchy that is mostly based on species rank. Over three centuries have passed with little exact definition of the word “species.” For various botanists, including both systematists and taxonomists, it has varied meanings. In practically all areas of biology, including anatomy, behavior, development, ecology, evolution, genetics, molecular biology, paleontology, physiology, and systematics, species are one of the fundamental comparison units. The significance of species in systematics, a field of biology on which all other areas of biology rely, accounts for a large portion of the value of species (de Queiroz, 2005).

Creating a theory to guide taxonomic practices in creating a lasting biological categorization is a major challenge in the study of biological systematics (Velasco, 2008). Additionally, concerns pertaining to species, such as speciation models, whether or not species even exist in nature according to taxonomy, and the delimitation of species, which typically involves species concept, are constantly mentioned in taxonomic and systematic literature. In addition to establishing what a species is, species notions also make speciation more clearly by specifying what a species is. As a result, research programs on speciation focus on the conditions and elements that depend on the species idea. Additionally, the definition of species is essential to many biological studies. For instance, ecological and behavioral research may look at how different species cooperate, including how species abundance interacts with biotic and abiotic factors. The majority of conservation biology research focuses on species. One of the most common biological units is the species, which is utilized in disputes in biology, agriculture, and medicine in addition to conservation and biology.

The term “species” therefore needs to be defined precisely and clearly for a wider audience than only academic biologists (Hausdorf, 2011; Aldhebiani, 2017). This chapter will cover the historical advancements of the notion of species and try to provide an outline of the history of the main species to help readers to better grasp how the notion of species has evolved.

## **1.2: History and development of the concept of species**

In the earliest scientific works, a species was just an individual organism that represented a collection of organisms that were similar to or nearly identical to one another. No relationships outside of that particular group were suggested. Genus and species are general and particular categories, respectively, according to Aristotle. Other pre-Darwinian scientists, like Aristotle, believed that each species had a unique, immutable “essence” like to the chemical elements. Early observers put previously isolated species into context as they started to create systems of organisation for living things. The term “species” was first defined biologically by English naturalist John Ray (Minelli, 2022).

In the 18th century, Swedish scientist Carl Linnaeus divided species into groups based on physical similarities rather than only differences. In order to reflect natural relationships, Carl Linnaeus also established the concept of a taxonomic hierarchy of classification that was mainly based on visible traits. However, despite how similar they might have appeared, it was still commonly accepted at the time that there was no organic relationship between species. This perspective was informed by the intellectual and theological education prevalent in Europe at the time, which thought that God has a hierarchical plan for the categories of life. Linnaeus worked to locate individual organisms that were exemplary of the species and thought that other non-exemplary organisms were deviant and abnormal, despite the fact that there are always distinctions (even if they are often minute) between individual organisms (Pokapu, 2007-2023).

By the nineteenth century, most naturalists realized that species might change shape over time and that the planet’s history allowed for significant alterations. In 1809, *Zoological Philosophy*, Jean-Baptiste Lamarck presented one of the first logical arguments against creationism. The new focus was on determining how a species might evolve over time. Lamarck proposed that an organism’s acquired property may be passed down to its progeny (for example, he attributed the giraffe’s long neck to generations of giraffes extending to reach the leaves of higher treetops). However, with the acceptance of Charles Darwin’s natural selection theory in the 1860s,

Lamarck's notion of goal-oriented evolution, sometimes known as a teleological process, was superseded. Scientists currently regard Charles Darwin and Alfred Wallace to have offered the most robust and compelling theory of evolution. Darwin maintained that groups, not individuals, evolved. His thesis was based on a significant shift of perspective from Linnaeus': instead of defining species in ideal terms (and hunting for an ideal representative and rejecting deviations), Darwin saw diversity among individuals as normal.

Darwin's work was inspired by Thomas Malthus' observation that the rate of increase of a biological population will always surpass the rate of growth of the environment's resources, such as food supply. As a result, not all individuals of a population will be able to survive and reproduce, Darwin argued. Those who did will, on average, have small changes that make them marginally better acclimated to the environment. If these changeable features are heritable, the survivors' progeny will have them as well. Thus, adaptive variations will accumulate in the population over many generations, whereas counter-adaptive features will tend to be removed (Darwin and Wallace, 2023, University of California, Berkeley, n.d.). The environment determines whether a variation is adaptive or non-adaptive: different settings favor different qualities. Because the environment essentially chooses which creatures live to reproduce, it is the environment (the "fight for existence") that chooses which features are passed down. This is the natural selection hypothesis of evolution. The length of a giraffe's neck would be explained in this hypothesis by supposing that proto-giraffes with longer necks had a large reproductive advantage over those with shorter necks. Over many generations, the entire population would be a long-necked animal species. The process behind the inheritance of individual features remained unknown in 1859, when Darwin published his theory of natural selection. Although Darwin speculated on how qualities are inherited (pangenesis), his theory is based only on the fact that inheritable features exist and are varied (making his achievement even more impressive). Despite the fact that Gregor Mendel's work on genetics was published in 1866, its importance was not recognized. Hugo de Vries, Carl Correns, and Erich von Tschermak rediscovered Darwin's work in 1900, realizing that the "inheritable traits" in Darwin's hypothesis are genes (Renáta and Kutáš, 2013).

### 1.3: Speciation

The emergence of a new species from a parental line is referred to as speciation. There is no visible dividing line between the original and descendent species. Despite the fact that current scientific understanding of species shows that there is no rigorous and complete technique to differentiate between various species in all circumstances, biologists continue to explore tangible ways to operationalize the concept. Richard Dawkins describes two creatures as conspecific if and only if they have the same number of chromosomes and the same number of nucleotides for each chromosome. Most taxonomists, however, would disagree. For example, the genome of many amphibians, most notably the *Leiopelma* frogs of New Zealand, consists of “core” chromosomes that are mainly invariable and accessory chromosomes, of which there are a number of possible combinations. Despite the fact that chromosomal counts vary greatly amongst populations, they may effectively interbreed and create a single evolutionary unit. Polyploidy is extremely common in plants, with few restrictions on interbreeding; because individuals with an odd number of chromosome sets are usually sterile, depending on the actual number of chromosomes sets present, this results in an unusual situation in which some individuals of the same evolutionary unit can interbreed with certain others and others cannot, with all populations eventually linked to form a common gene pool (Chung-I Wu, 2022).

The classification of species has been profoundly influenced by technological advances that have enabled researchers to determine relatedness based on molecular markers, beginning with the relatively crude blood plasma precipitation assays in the mid-twentieth century and progressing to Charles Sibley’s DNA-DNA hybridization studies in the 1970s, which led to DNA sequencing techniques. The outcomes of these methodologies resulted in dramatic modifications in higher taxonomic groups (such as phyla and classes), rearranging numerous branches of the phylogenetic tree (molecular phylogeny). So far, the results for taxonomic groups below genera have been mixed; the speed of evolutionary change on the molecular level has been modest, generating apparent differences only after long years of reproductive isolation. The pomarine skua–great skua phenomena is a classic example of DNA-DNA hybridization findings leading to incorrect conclusions. Turtles have been found to evolve at one-eighth the pace of other reptiles on the molecular level, while the rate of molecular evolution in albatrosses is half that of the very closely related storm-petrels, both of which are members of the Procellariiformes. Hybridization is no longer used and has been superseded by more reliable computational algorithms

for sequence comparison. Molecular taxonomy is based on the total change caused by evolutionary processes rather than on the evolutionary processes themselves. Mutation, crossover, and selection are not uniform processes that contribute to the development and preservation of variation. Changes in the DNA sequence enduring over generations are a result of natural selection rather than being a direct target of the latter; for example, silent transition-transversion combinations would alter the melting point of the DNA sequence but not the sequence of the encoded proteins and thus are a possible example where, for example, in microorganisms, a mutation confers a change in fitness all by itself (Brenner, Staley and Krieg. 2015; Suárez-Díaz, 2013).

#### **1.4: Biologists' working definition of species**

Traditionally, before a suggested species can be considered a species, numerous specimens of it must be evaluated for unifying characteristics. In contrast to the traditional notion of a species as a class of creatures, some biologists may see species as statistical phenomenon. In such situation, a species is defined as a lineage that evolves independently and generates a single gene pool. Although DNA sequences and morphology are utilized to assist distinguish closely related lineages, the borders of this characterization are hazy. However, the precise meaning of the term “species” remains debatable, particularly among prokaryotes, and this is why it is called the species problem. In practice, scientists define species as communities of creatures with strong genetic resemblance. This might be due to adaptation to the same niche and the transmission of genetic information from one individual to another via a number of methods. The precise amount of resemblance employed in such a definition is arbitrary, but it is the most commonly used definition for organisms that reproduce asexually (asexual reproduction), such as some plants and microbes (Gibson, 2012).

Because there is no defined species idea in microbiology, several scholars argue that the term “species” is useless for examining bacterial evolution. Instead, they detect genes freely travelling across even distantly related species, with the entire bacterial domain functioning as a single gene pool. Nonetheless, a form of rule of thumb has been created, stating that bacteria or archaea species with 16S rRNA gene sequences that are more than 97% identical to each other should be examined by DNA-DNA hybridization to check if they belong to the similar species. This design was recently modified, stating that the boundary of 97% was too low and should be raised to 98.7%.

The ability of two organisms to interbreed and produce fertile offspring of both sexes is generally accepted as a simple indicator that the organisms share enough genes to be considered members of the same species in the study of sexually reproducing organisms, where genetic material is shared through the process of reproduction. As a result, a “species” is a group of creatures that interbreed. This definition may be expanded to state that a species is a collection of creatures that can possibly interbreed—fish can be considered the same species even though they reside in separate lakes, as long as they can interbreed if they come into touch with each other. On the other hand, there are several examples of series of three or more separate populations in which people from the center population may interbreed with persons from either side, but individuals from either side cannot interbreed. As a result, one may argue that these populations are either one species or two separate species. This is not a contradiction; rather, it demonstrates that species are defined by gene frequencies and so have ambiguous borders. As a result, any single, universal definition of “species” is inherently arbitrary. Instead, biologists have offered a variety of definitions; whatever term a biologist employs is a pragmatic decision based on the specifics of that biologist’s study.

In practice, these definitions frequently overlap, and the discrepancies are more a question of emphasis than blatant disagreement. Nonetheless, no species idea has yet been offered that is completely objective or that can be used in all instances without resorting to judgement. This is the biological species concept (BSC) for most vertebrates, and the Phylogenetic Species Concept (PSC) to a lesser extent (or for other purposes). Many BSC subspecies are recognized species under the PSC; the distinction between the two is that the BSC defines a species as a result of manifest evolutionary history, whereas the PSC defines a species as a result of manifest evolutionary potential. As a result, a PSC species is “made” as soon as an evolutionary lineage begins to split, but a BSC species begins to exist only after the lineage separation is complete. As a result, there can be significant conflict between alternative classifications based on the PSC versus BSC, because they treat taxa that would be considered subspecies under the latter model completely differently (e.g., the numerous subspecies of honey bees (Shapiro and Polz, 2015)).

It is surprisingly difficult to define the word “species” in a way that applies to all naturally occurring organisms, and the debate among biologists about how to define “species” and how to identify actual species is called the species problem. Over two dozen distinct definitions of “species” are in use amongst biologists. This problem dates as early as to the writings of Charles

Darwin. While Darwin wrote a definition in *Species*, no one definition has satisfied all naturalists. Most modern textbooks follow Ernst Mayr's definition, known as the Biological Species Concept (BSC) of a species as "groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups". It has been argued that this definition of species is not only a useful formulation, but is also a natural consequence of the effect of sexual reproduction on the dynamics of natural selection.

Various parts of this definition serve to exclude some unusual or artificial mating:

- Those that as a result of deliberate human action, or occur only in captivity (when the animal's normal mating partners may not be available)
- Those that involve animals that may be physically and physiologically capable of mating but, for various reasons, do not normally do so in the wild. The typical textbook definition above works well for most multi-celled organisms, but there are several types of situations in which it breaks down:
  - By definition it applies only to organisms that reproduce sexually. So it does not work for asexually reproducing single-celled organisms and for the relatively few parthenogenetic or apomictic multi-celled organisms. The term "phylotype" is often applied to such organisms.
  - Biologists frequently do not know whether two morphologically similar groups of organisms are "potentially" capable of interbreeding.
  - There is considerable variation in the degree to which hybridization may succeed under natural conditions, or even in the degree to which some organisms use sexual reproduction between individuals to breed.
  - In ring species, members of adjacent populations interbreed successfully but members of some non-adjacent populations do not.

A useable definition of the term "species" as well as accurate techniques of identifying specific species are required for asserting and testing biological ideas as well as quantifying biodiversity.

## 1.5: Common names and species

Common names for plant and animal taxa may correlate to species: for example, "lion," "walrus," and "camphor tree" are all species names. In

some circumstances, popular names do not: for example, “deer” refers to a family of 34 species, including Eld’s deer, red deer, and elk (as used in American English to indicate wapiti, not moose). The final two species were originally considered a single species, demonstrating how species boundaries may shift as scientific understanding expands (Gibson, 2012).

## 1.6: Placement within genera

A species should ideally be given a formal, scientific name, although in practice, there are many unidentified species (which have merely been described, not named). When a species is named, it is assigned to a genus. This might be interpreted scientifically as a belief that the species is more closely related to other species within its genus (if any) than to species from other genera. Species and genera are often defined within the context of a wider taxonomic hierarchy. Life, domain, kingdom, phylum, class, order, family, genus, and species are the most well-known taxonomic levels. This genus attribution is not permanent; a different (or the same) taxonomist may subsequently place it to a different genus, in which case the name will also change (Pyle et al., 2021). In biological nomenclature, the name of a species is a two-part name (a binomial name), which is considered as Latin, however roots from any language, as well as names of places or persons, can be used. The general name appears first (with the first letter capitalized), followed by another phrase. The second word is defined differently in zoological and botanical nomenclature.

The second half of the name is known as the particular name or the specific epithet in zoological nomenclature. Grey wolves, for example, belong to the species *Canis lupus*, coyotes to *Canis latrans*, golden jackals to *Canis aureus*, and so on, all of which are members of the genus *Canis* (which also includes many other species). The grey wolf’s genus name is *Canis*, its specific name or specific epithet is *lupus*, and its binomen, or species name, is *Canis lupus*. The second element of the name is known as the specific epithet in botanical nomenclature. In botany, the ‘specific name’ is usually a combination of the genus name and the particular epithet. For example, *Pinus palustris* is the popular name for the longleaf pine; the genus name is *Pinus*, the specific epithet is *palustris*, and the specific name is *Pinus palustris*. This binomial naming system was initially employed by Leonhart Fuchs and then formalized in biological rules of nomenclature by Carolus Linnaeus in his 1753 *Species Plantarum* (followed by his 1758 *Systema Naturae*, 10th edition) (Gibson, 2012).

## 1.7: Abbreviated names

Books and articles may purposefully fail to properly identify species, instead using the abbreviation “sp.” in the singular or “spp.” (Species pluralis, Latin abbreviation for several species) in the plural in place of the particular epithet (e.g. *Canis* sp.). The term “spp.” is used as a shorthand for anything that pertains to several species within a genus but not all species within that genus. The genus term is used by scientists to indicate that something applies to all species within a genus name without the specific epithet. The aforementioned plural is sometimes transcribed as “sps.”, which may be confused with “ssp.”, which stands for subspecies instead. Genus and species names are often printed in italics in books and articles. Italicized abbreviations such as “sp.”, “spp.”, “sps.”, “ssp.”, “subsp.”, and so on should not be used (Winston, 1999).

## 1.8: Identification codes

Several codes have been devised for identifying particular species; for example:

- National Center for Biotechnology Information (NCBI) employs a numeric ‘taxid’ or *Taxonomy identifier*, a “stable unique identifier”, e.g. the taxid of *H. sapiens* is 9606.
- (KEGG) employs a three- or four-letter code for a limited number of organisms; in this code, for example, *H. sapiens* is simply *hsa*.
- UniProt employs an “organism mnemonic” of not more than five alphanumeric characters, e.g. *HUMAN* for *H. sapiens*.
- Integrated Taxonomic Information System (ITIS) provides a unique number for each species.

Among microorganisms, in particular, the problem of species identification is made difficult by discordance between molecular and morphological investigations; these can be categorized as two types:

- (i) One morphology, multiple lineages (e.g. morphological convergence, cryptic species) and
- (ii) One lineage, multiple morphologies (e.g. phenotypic plasticity, multiple life-cycle stages).

In addition, in these and other organisms, horizontal gene transfer (HGT) makes it difficult to define the term species. All species definitions assume

that an organism acquires its genes from one or two parents very like the “daughter” organism, but HGT makes that assumption false. There is strong evidence of HGT between very dissimilar groups of prokaryotes, and at least occasionally between dissimilar groups of eukaryotes.

## 1.9: Number of prokaryotic species

Prokaryotes are by far the most numerous species on the earth. They are also by far the most varied, both metabolically and phylogenetically; they include bacteria and archaea, two of the three major classifications of living beings. The current taxonomy of prokaryote species is based on a mix of genomic and phenotypic features. When compared to higher eukaryotes, the proposed cut-off value of 70% DNA-DNA similarity to distinguish species represents an exceptionally broad species definition for prokaryotes. There are now somewhat more than 6200 legitimately named prokaryotic species (Oren, 2004). The International Code of Nomenclature of Prokaryotes (ICNP), which is based on the Linnaean binomial system, oversees prokaryote naming. The current Code limits only apply to cultivated minority nomenclature. The International Committee on Prokaryotic Systematics has rejected suggestions to incorporate the vast majority of uncultivated bacteria and archaea inside the Code’s restrictions. The provisional rank of *Candidatus* can be used to designate uncultivated prokaryotes whose names cannot be properly published under ICNP standards, but whose names can currently be validated under the SeqCode, which was recently formed to address uncultivated majority nomenclature. Metagenomics, single-cell genomics, and high-throughput culture technologies have resulted in a flood of new species that need to be identified. GAN and Protologger are examples of automated program that can assist researchers in naming and describing newly discovered prokaryotes, both cultivated and uncultivated. However, knowledge of Latin and Greek is still essential for proper quality control of names that must meet the standards set by nomenclature regulations (Oren, 2023).

## 1.10: Number of eukaryotic species

One of the most fundamental yet perplexing topics in science is the number of species on Earth. Unfortunately, getting an actual estimate is difficult due to the fact that most species have yet to be identified and indirect attempts to address this issue have been very contentious. We show that the taxonomic categorization of species into higher taxonomic groups (from genera to phyla) follows a consistent pattern that may be used to estimate

the total number of species in any taxonomic group. Based on this trend, there will be 8.7 million (1.3 million SE) species on Earth, with 2.2 million (0.18 million SE) of these being marine. Findings indicate that around 86% of the species on Earth, and 91% in the ocean Still waiting for a description (Table 1-1). Closing this knowledge gap will need a revived interest in exploration and taxonomy, as well as an ongoing effort to classify existing biodiversity data in publicly accessible databases (Mora et al., 2011).

Table 1-1: Estimated number of species on Earth

<b>Kingdom</b>	<b>Number of species (<i>Ocean</i>)</b>	<b>Number of species (<i>Terrestrial</i>)</b>	<b>Number of species (<i>Total</i>)</b>
Animals	2,150,000	5,620,000	7,770,000
Chromists	7400	20,100	27,500
Fungi	5320	605,680	611,000
Plants	16,600	281,400	298,000
Protozoa	36,400	0	36,400
Archaea	1	454	455
Bacteria	1320	8360	9680
<b>Total species</b>	<b>2,210,000</b>	<b>6,540,000</b>	<b>8,750,000</b>

Source: Mora, Camilo et al, 2011.

The IUCN Red List tracks the number of described species and updates this figure annually based on the latest work of taxonomists. In 2021 it listed 2.13 million species on the planet. In the chart we see the breakdown across a range of taxonomic groups – 1.05 million insects; over 11,000 birds; over 11,000 reptiles; and over 6,000 mammals (Ritchie, 2022). Organizations such as the Global Taxonomy Initiative, the European Distributed Institute of Taxonomy, and the Census of Marine Life (the latter solely for marine animals) are now working to enhance taxonomy and add previously unknown species to the system. Current knowledge only covers a subset of the creatures in the biosphere and so does not allow for a thorough understanding of how the ecosystem works. Humanity is also eradicating previously unknown species at an unprecedented rate, which means that even before a new species has had the opportunity to be investigated and categorized, it may already be gone. iBOL (<https://www.cbd.int/gti/>) is the pioneer in the development of DNA-based technologies for biodiversity monitoring and conservation. iBOL assists CBD Parties in improving their capacity to identify and monitor biodiversity via capacity building and technology transfer (Mora et al., 2011; Sweetlove, 2011).

### **1.11: Conclusion**

For a long time, defining and recognizing a species has been a contentious topic. Many concepts have been offered to determine the variance and limitation between species. When a taxonomist studies a specific taxon, he or she must accept a species concept and offer a species limitation in order to define that taxon. Various attempts have been made to define a species and the species concept kept on changing from time to time. Also knowing the total number of species has been a question of great interest motivated in part by our collective curiosity about the diversity of life on Earth and in part by the need to provide a reference point for current and future losses of biodiversity. Unfortunately, incomplete sampling of the world's biodiversity combined with a lack of robust extrapolation approaches has yielded highly uncertain and controversial estimates of how many species there are on Earth. Several definitions of species and the species concepts developed over the time for eukaryotes and prokaryotes have been discussed in Chapter 2 and 3, respectively.

## CHAPTER 2

# CONCEPTS OF SPECIES IN EUKARYOTES

### 2.1: Introduction

Historically, any distinguishing features that could be viewed, counted, and measured sufficed to define new species. Individual organisms are typically recognizable, while larger units of life, such as populations, subspecies, species, or genera, are not. Taxonomists divide species into genera, families, orders, and kingdoms, whereas ecologists divide species into higher-level structures such as communities and ecosystems. The foundation for these group terms is usefulness rather than intrinsic naturalness; nonetheless, we aim to delimit groups of species as much as possible along natural fault lines so that independent observers may recover essentially the same groupings.

Given that life has evolved in a more or less hierarchical fashion, there will be an almost infinite number of alternate, yet nested, methods of classifying the same animals. Although Darwin (1859) considered that species were similar in type to lower and higher taxonomic levels, many current academics think that species are more objectively recognizable, and so more real, than populations or genera. Today, the species appears to be the most important taxon in much of ecology, evolution, and biodiversity. It has been proposed that if genera or subspecies were accorded the same logical position as species, these disciplines would be jeopardized.

The International Commission of Zoological Nomenclature states species concepts from taxonomy and defines species as “the basic unit of classification”. Taxonomy and allied fields generally employ species to organize and retrieve information about individual specimens in collections or biodiversity databases. It may also be important in evolutionary biology to differentiate a sort of divergent evolution known as “speciation,” which produces qualitatively different effects than within-species evolution while employing the same processes.

Taxonomists devised species concepts at the beginning of the twentieth century to give categorization an acceptable logic or operational

mechanism. Each species notion remains a theoretical construct incapable of depicting all aspects of a complicated world. A species notion is plausible if it pertains to real-world occurrences that are important from an evolutionary standpoint and if its theoretical concept can be turned into a good taxonomic working routine.

It would be great if a single definition of species could meet all of these applications, but such a definition has yet to be discovered, and some feel it is impossible. However, if species are more real, objectively definable, and basic than, say, genera or subspecies, as many people appear to believe, a unified definition should be attainable. In contrast, even if species have no higher objectivity than other taxa, unitary nominalist principles for species delimitation may be implemented, maybe after much diplomacy via an international agreement among biologists. However, knowledge of the whole range of today's competing answers to the species idea issue will very certainly be required in order to find a universal species definition (Mallet, Seixas and Thawornwattana, 2022).

This chapter discusses the traditional concepts of species and major species concepts in eukaryotes with their advantages and disadvantages.

## **2.2: History of defining the term “species”**

The English scholar John Ray (1627-1705) was the first to provide a species definition. “No matter what variations occur in individuals or species if they spring from the seed of one and the same plant, they are accidental variations and not such as distinguish a species permanently; one species never springs from the seed of another, nor vice versa,” he wrote in his major work *Historia Plantarum* (Ray, 1686). As a result, Ray strove to define species as groups of plants that actually reproduce within their genetic limitations.

Over 50 years later, Carolous Linnaeus (1707-1778), whose work was the most prominent and consequential in the area of taxonomy, embraced a larger notion and established a new definition of species. Linnaeus documented quickly and methodically around 5900 species of plants known to man at the time in his work *Species Plantarum* (Linnaeus, 1753), focusing mostly on floral structure and sexual traits. As a result, he employed a “natural system” sexual system to define species. Linnaeus' notion was straightforward, relevant, and universally accepted. Carolus Linnaeus is credited with developing the first widely recognized hierarchical structure, which now includes seven categories: kingdom, phylum, class, order,