# Species Concepts and Speciation

3

Much thought has been given to the concept of a species. Darwin entitled his book *The Origin of Species*. By so naming his book, Darwin acknowledged that an entity called the species does exist. However, agreeing on exactly what a species is has been the subject of numerous and continued debate. Living organisms, including bacteria, show definable differences that can be observed and measured. Naturalists have used these differences to aid in the classification of life. The notion of classification is based on the assumption that observable differences are discreet in space and time. If we consider that evolution has been a continuous process since life originated, should we expect a discontinuous product such as a species to be formed, or should we expect continuous variation? Are the apparent discontinuities among living organisms due to random extinctions, or would we be able to identify unique groups even if there had been no extinctions and all life was before us? Variation both within and between species is recognized and either dealt with directly or tacitly acknowledged in discussion of species.

Early taxonomists placed new specimens into discreet hierarchical categories. The lowest and most restrictive category is the species, which is composed of a genus and species designation, the *binomial system of nomenclature* (Box 3.1). In the cataloging of species, a monumental undertaking that involved thousands of scientists, additional relationships among creatures were seen. Some organisms differing in hardly perceptible ways were classified as the same species but were never observed to mate and produce viable, fertile offspring. Other organisms differed in size, coloration, and shape and were mistakenly classified as separate species but later found to be the same species.

The species concept is ancient and originally linked to Providence, at least in Western philosophy. Groups or kinds of organisms were considered divinely created and immutable. A set of essential characteristics defined a particular organism and other observers could, based on those essential characters, identify the organism. Every living thing could, in theory, be placed within a specific group based on sets of observations. With the advent of evolutionary thought, the species concept became central to biology for a different set of reasons than the early naturalists were working under. Instead of Providence, natural selection and other evolutionary processes produced unique and wonderful products, which at the lowest discernible level were called *species*. However, the question remains: Are species real?

**Box 3.1 The Binomial System of Nomenclature**

Carolus Linnaeus developed a scheme to classify all known living organisms. His scheme was basically based on the form of the organism. Before the development of his system, the names of organisms were Latin descriptions that often were more than one word long. The simplicity and clarity of the linnaen system have been its strongest attributes. Each organism is given two names that correspond to a genus and a specific epithet. A specific organism is always called by both names because many organisms may have the specific epithet but are in entirely different genera (e.g., *Escherichia coli* and *Campylobacter coli*). Both of these bacteria have the species name of *coli*, but they are in different genera. The beauty of the system is that no matter the native language of the scientist, the organism being studied has the same name, decreasing confusion. In writing a binomial name, the genus is always written first and capitalized. A genus is a group of closely related species. The genus name can be used without the species modifier when speaking of this closely related group, such as all *Pseudomonas* bacteria. Taxonomy recognizes categories or levels of biological relatedness above that of the genus. Genera are grouped into families, families into orders, orders into classes, and classes into phyla or divisions. However, all classifications above the species level have no real biological meaning.

Is there a grouping of organisms such that no lower grouping exists? If so, what are the criteria on which the grouping is based? A species, by definition, is the fundamental unit of nature. The word *nature* comes from the Indo-European word *gene*, which is the same root word for genealogy. Identifying the fundamental unit of nature is literally finding the products of descent. Evolutionary descent is particularly difficult to determine. Based on numerous techniques, descent and origin of species can be inferred, but total reconstruction of lineage and relationships is not possible.

Within the scientific community and among nonscientists, species have a practical application and are the basic units of conservation and biodiversity. All attempts to preserve diversity or promote conservation are directed at species. For example, in the United States, legislation assumed that species were real and easily identified with the passing and enforcement of the Endangered Species Act. Much attention has been focused on the levels of biodiversity regionally, nationally, and globally and whether this diversity is decreasing. The concept of biodiversity has meaning only at the species level and presupposes that species exist. Higher-level diversity is confusing and misleading. Some higher-level taxonomic designations such as genus may exist as single groups within the next higher taxon but have numerous species within. Measurements of diversity based on the genus would be much lower than that at the species level.

The criteria used to designate a species, even in the earliest attempts, were threefold:

1. The organism had to have been described from nature by a taxonomist.
2. The organism must be recognizable to others (i.e., characters are constant over generations).
3. There must be fertility when crossed with like organisms.

However, many specimens were dead and had been removed (collected) from their habitats with little ancillary information on the biology of the organism available. Most species designations were based on observed morphological differences, and the



third criterion wa
whereas taxonomi
was not until the
in the 1930s and
developed.

## Universal

Let us consider tl
In science, there a
(Hull, 1997). Thes
of the species con
Asexual organism
reduce their gener
is probably microl
less than universal

As observed by
difficult it is to ap
seeking to describ
morphology, there
phology is not suf
guidelines, such a
However, operatio
process.

The concept of
can be independer
Should it be based
Is a species a natu
or can be based o
term? What is the
lution has been ac
ucts of evolution?
fundamental to al
ability among and
However, much of
different response

Many species
Although there is
the concepts, each
Among the many
representative: th
evolutionary spec
describe each of t
plete comparison
in this chapter.

third criterion was never observed. Naturalists continued to use the earlier criteria, whereas taxonomists working in museums relied more and more on morphology. It was not until the great synthesis of genetics, systematics, and evolutionary biology in the 1930s and 1940s that a broadly defined concept of a species began to be developed.

## Universal Species Concept

Let us consider the problems associated with defining a universal species concept. In science, there are three common criteria a concept must meet in order to stand (Hull, 1997). These are universality, applicability, and theoretical significance. Most of the species concepts developed today have trouble with being general or universal. Asexual organisms present an intractable problem for these concepts and greatly reduce their generality. If we consider that most of the species diversity on this planet is probably microbial or parasitic, failure to include these groups makes the concept less than universal.

As observed by Hull (1997), the more theoretically significant a concept is, the more difficult it is to apply the concept. Ease of application is important to taxonomists seeking to describe species and species relationships. Operational guidelines, such as morphology, therefore can be formalized and used to describe species. When morphology is not sufficient to discriminate closely related organisms, other operational guidelines, such as genetic relatedness, can be used to define species boundaries. However, operational designations of species define the products of evolution, not the process.

The concept of the species has an evolutionary and a taxonomic meaning, which can be independent of each other. How is membership in a species to be determined? Should it be based on morphology, physiology, reproduction, or some other criterion? Is a species a natural kind or a set of organisms or individuals? If the term *species* is or can be based on morphology, physiology, or reproduction, is there meaning in the term? What is the importance of having a species concept? If we consider that evolution has been acting continuously for eons, why do we expect discontinuous products of evolution? Do species really exist in nature? The concept of the species is fundamental to all aspects of biology and especially to ecology and evolution. Variability among and between organisms is the starting material for evolutionary change. However, much of this variability is plastic– under different conditions organisms have different responses.

Many species concepts have been developed and championed over the years. Although there is overlap and after careful examination *synonymy* among some of the concepts, each has its own set of assumptions, theory (at times) and applicability. Among the many concepts currently being discussed we will choose four that are representative: the biological species concept, the phenetic species concept, the evolutionary species concept, and the phylogenetic species concept. We briefly describe each of these concepts, but the referenced articles provide for a more complete comparison and description of each concept and the many others not discussed in this chapter.

## Biological Species Concept

One of the oldest and most developed concepts is the *biological species concept* (BSC). This concept was championed by Earnst Mayr in 1942 and subsequently developed, discussed, and applied by many of the early, prominent evolutionary biologists. It is still being refined and applied today, and it is the concept of species most widely used in biology by botanists, zoologists, politicians, resource managers, and others concerned about biodiversity.

In the BSC, species exist as part of a reproductive community. Individuals within a reproductive community must be ecologically accessible to others within the same environment, and gene flow must be actually or potentially occurring. Mating maintains the gene pool that "regardless of the individuals that constitute it, interacts as a unit with other species with which it shares its environment." Because the definition is based on reproduction, selection must favor the acquisition of mechanisms that promote breeding with *conspecifics*. *Reproductive isolation* becomes a mechanism for the protection of genotypes. *Speciation* under this concept is the process of achieving reproductive isolation.

The BSC excludes much of the life on this planet, including all *uniparental* species, as well as parthenogenic and self- or sib-mating species. Proponents of the concept sometimes call these other organisms *pseudospecies*, or as Gheslin (1987) posited, the organisms are not species. Hull (1997) said, "It should be kept in mind that very little in the way of gene exchange occurred during the first half of life on earth, and meiosis evolved even later. According to the biological species concept, no species existed for at least the first half of life on earth. Evolution occurred but in the absence of species." This is an important observation, and it relates directly to the application of the BSC concept to bacteria and other asexual organisms. Is it true that most of the world's biota are not species? This seems rather restrictive. It seems likely that if the products of evolution are species, the organisms that were living out their lives under the pressure of natural selection during the first half of the period of life on the planet were probably species.

The BSC has difficulty explaining hybrids because the concept is based on reproductive isolation. Hybrids, especially fertile hybrids, indicate that species designations for the parents are not restrictive. However, many species can hybridize and continue to maintain temporally ecological and genetic identities. Under the BSC, individual parasites and bacteria could be considered a species and each egg or fission a speciation event.

Although the BSC is widely used practically and theoretically, there are several problems associated with the concept. Mayden and Wood (1995) identified 10 elements of this concept that they considered counterproductive for understanding biological diversity:

1. Absence of a lineage perspective
2. Nondimensionality (With this concept, species exist in a brief segment of time with no linkage to the past. Spatially, the concept applies to other organisms that come in contact. Although the BSC has been extended to potentially interbreeding populations, the application is restricted to specific locations and times.)
3. Erroneous operational qualities used as definition
4. Exclusion of n
5. Indiscriminate
6. Confusion of i
7. Implicit relianc
8. Its relational n ductively isolat
9. Its teleological
10. Its employmen concepts

## Phenetic a

The phenetic specie the basis of overall ical score, which is unit in phenetics is mensional. Individ than between grou under this concept. phenetic. Multiple describe relationshi

## Evolutiona

Being dissatisfied w the evolutionary s Simpson defined ev lineages and posses modifications to th replaced with evolu argue that the ESC all known types of concept. As descrit species formed thr is considered a deri

## Phylogene

The phylogenetic s diagnosable cluster of ancestry and d

4. Exclusion of non–sexually reproducing organisms
5. Indiscriminate use of a reproductive isolation criterion
6. Confusion of isolating mechanisms with isolating effects
7. Implicit reliance on group selection
8. Its relational nature (i.e., A is a species relative to B and C because it is reproductively isolated from them)
9. Its teleological overtones
10. Its employment as a typological concept, no different from morphological species concepts

## Phenetic and Related Species Concepts

The phenetic species concept is based on *numerical taxonomy*. Species are defined on the basis of overall similarity or divergence in characters that can be given a numerical score, which is usually the presence or absence of the trait. The lowest taxonomic unit in phenetics is the *operational taxonomic unit* (OTU). The OTU is defined in terms of covarying characteristics using various statistical methods. This concept is nondimensional. Individuals for whom variance in characteristics is lower within a group than between groups are considered a distinct taxon. Species do not exist as lineages under this concept. Many molecular and morphological species concepts are basically phenetic. Multiple characters are measured or observed and the responses used to describe relationships. This concept is primarily operational.

## Evolutionary Species Concept

Being dissatisfied with the nondimensionality of the BSC, Simpson (1961) developed the evolutionary species concept (ESC), which extends the BSC through time. Simpson defined evolutionary species as groups that evolve separately from other such lineages and possess their own unitary roles and tendencies. Wiley (1981) made a few modifications to the concept and reworked it so that unitary roles and tendencies were replaced with evolutionary tendencies and historical fates. Wiley and Mayden (1997) argue that the ESC is the only available concept with the capacity to accommodate all known types of biologically equivalent diversity. The ESC is not an operational concept. As described by Mayden (1997), the concept "accommodates uniparentals, species formed through hybridization, and ancestral species." Reproductive isolation is considered a derived characteristic.

## Phylogenetic Species Concept

The phylogenetic species concept is based on the idea that a species is "the smallest diagnosable cluster of individual organisms within which there is parental pattern of ancestry and descendent" (Cracraft, 1983). A species is easily observed as the

terminal organism in a lineage (Moreno, 1996) that has a common ancestor to other terminal organisms. The phylogenetic species concept does not rely on higher taxonomic and probably meaningless ranks above the genus level. Avise (1994) pointed out that the problem with this approach is that of determining the difference between gene phylogenies and pedigrees and how to recognize monophyly.

The approach is based on two very different types of data. The first uses comparisons of characters observed between fossil and living or extant species. The second is based on comparisons of molecular sequences obtained from different organisms.

## Bacterial Taxonomy

Many scientists hold to a five-kingdom system that includes plants, animals, fungi, protists, and monerans (i.e., bacteria). With the rise of modern molecular techniques, techniques that can determine differences in the building blocks of genes, an interesting system has been put forward. This system (Figure 3.1), based on differences in the patterns of individual building blocks of a certain type of DNA, suggests that there are only three major groups of living things. Two of the three major groups are microbial. All other living things fall into a single category. In other words, differences among puny, nondescript bacteria are greater than differences between cypress trees and humans.

A major problem with comprehending this immense genetic diversity in microbes is the inability to remember that they have been around for a very long time. Processes that select for certain traits have had plenty of time to fine tune and modify bacterial genes—far longer than all other organisms added together!

We have been considering the genetic diversity that exists among bacteria; let us return to the species problem. All taxonomic categories above that of species exist only in the imagination of man. In theory, the species designation should be the most "real" category. Microbes have been assigned to various species groups. In the past, this was done based on the ability of the microbes to perform various enzymatic and

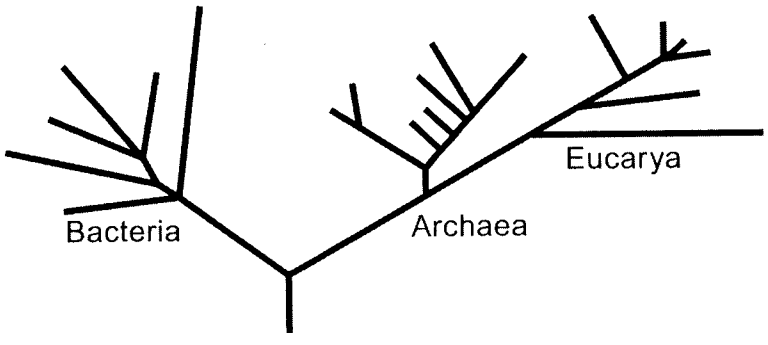


**Figure 3.1** The tree of life as determined by 16S rRNA sequences. This is a simplified version of the tree. Considerable detail has been added to each of the branches. Notice that most of the diversity of life forms is microbial. (Adapted from Woese CR, Kandler O, Wheelis ML. 1990. Toward a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 87:4576–4579, 1990.)

metabolic activities. Using modern molecular biological techniques, scientists can show the degree of relatedness by examining the similarity between DNA from one organism and another. This method of species designation has its own set of ambiguities. For example, at what level of similarity are two organisms the same species? There are no clear boundaries. Let us muddy the waters a little more.

Some bacteria are *promiscuous*, not in the moral sense of the word, but in the sense of the strict scientific meaning. Promiscuity is not restricted to one set or class. The term is most frequently used to describe a person who has multiple sexual relationships. In this sense, a promiscuous person is not restricted to a single individual. In a larger sense, promiscuous could be used to describe businesses that broaden their base and are not restricted to one customer. However, the term does fit bacteria. They are promiscuous. Bacteria can share their genes with other bacteria through several unique evolutionary mechanisms. Bacteria can even share their genes with totally unrelated "species" of bacteria, and there is evidence that they can share genes with eukaryotes.

What is a bacterial species? Many bacterial genes are mobile and can be passed between unrelated organisms. The oldest organisms on the planet are the most difficult to assign to a group. Over 3.8 billion years of evolutionary experimentation, they have broken down barriers that seem to exist among other organisms. They have developed a more cosmopolitan approach to life.

## Bacterial Species Concepts

We know remarkably little about the taxonomy of bacterial species. Even today, the characterization of bacterial species is swayed by the perceived need. Bacteria that are industrially important are better characterized than are medically important species and many times more so than ecologically important species. More than 3,000 species of *Streptomyces* have been identified and patented by the pharmaceutical industry.

Based on various newly devised molecular biological techniques, most known bacterial diversity falls within distinct *phenetic* clusters. These technique-driven methods provide some basis to classify bacteria, but they, as Goodfellow et al. (1997) state, "overlook the fact that species are the product of biological processes." The cookbook nature of these techniques makes them readily accessible to many, and the procedures have been used to describe bacterial species across many habitats. Bacterial species designations are usually assigned to groups of strains that show high levels of similarity in biochemistry, genetics, morphology, nutrition, and structure (Goodfellow et al., 1997). This approach to describing bacterial species is operational but not evolutionary sound, and it relies on the product and not the process to define a species. Even studies that seek to describe evolutionary relationships among bacteria and other taxa (e.g., Woese) use an operational method. Comparisons of 16S rRNA sequences can and do show relationships among groups of organisms, but they do not provide a clear definition of what a species is or how it came to be (discussed later). Numerous studies describe the microbial diversity found in soil, marine, estuarine, freshwater, acid mine drainage, thermal springs, and many other habitats using an operational definition of a species. These studies are important and underscore our need to develop an evolutionary definition of bacterial species.

Most bacterial species names in use today are *taxospecies*. The species designation comes from some application of numerical taxonomy (Sneath and Sokal, 1973). This procedure assigns a numerical code to a series of phenotypic observations of a particular isolate. Numerical taxonomy is the foundation of diagnostic test strips or plates that contain various organic compounds or that indicate the presence of particular degradative or metabolic capabilities. These methods are quick. However, classification to a specific species requires that a complete numerical description of the species be in a database. Because relatively few bacteria can be cultivated and most environmental isolates have not been carefully and thoroughly examined, the databases are restrictive. A number can be generated for an isolate, but classification of that isolate to a species may or may not be valid. Numerical scores are often unique and can be used to designate an unknown species. Levels of microbial diversity can be estimated using the frequency of the numerical scores. Numerical scores that are identical are considered to identify the same organism. Subtle differences in scores can be resolved using multivariate statistical analyses.

A second bacterial species concept is that of the *genomic species*. Genomic species are strains that show DNA:DNA relatedness values greater than some specified value and thermal denaturation values less than some specific rating (Box 3.2). From previous work with enteric bacteria, genomic species were recognized when individuals had 70% or more DNA:DNA relatedness with a difference of 5°C or less in thermal stability.

Neilsen et al. (1995) showed that there was good agreement between DNA:DNA hybridizations and data obtained through numerical taxonomy or chemotaxonomy. This correspondence gives some support to the taxospecies designations and does support the idea that metabolic and functional differences are maintained at the genetic level. The results from numerical taxonomy seemed to suggest that microbial responses are continuous. However, what appeared to be a continuum of responses under numerical taxonomy resolved into defined groups of bacteria when DNA:DNA pairings were made. Exceptions to these clearly defined groups have been found and draw into question the species designations or the universal application of this technique to answer bacterial species questions.

For example, within the genus *Xanthomonas*, DNA:DNA pairings range from 0% to 100% between various pathovars (Hildebrandt et al., 1990), although these pathovars appear indistinguishable when compared biochemically. Such incongruities need further investigation but suggest that the genus designation may be too inclusive. Bac-

**Box 3.2 DNA:DNA Hybridization and Bacterial Species**

Ward et al. (2002) pointed out that the concept of a bacterial species based on DNA:DNA hybridizations is fraught with difficulties. Among these difficulties is the arbitrary nature of the selection criterion: more than 70% homology. The danger of setting a number for species designations is that researchers may feel these numbers are truly thresholds that, once crossed, identify a species. However, true bacterial species may have hybridization percentages that are higher than 70% or that show greater than 97.5% sequence similarity. On the other hand, these types of data describe a continuum of possible relatedness among bacteria and suggest that species designations may be meaningless for some or all groups of bacteria. For many higher organisms, homologies and similarities at these levels would make many recognized species disappear. For example, most of the angiosperms would end up as a single group, and most primates would be designated as a single species. Given enough data, what appears now as a continuum of similarities may resolve into disjunctive unique groupings.

teria that have the same functional capacity as evidenced by the biochemical test but that differ genetically is suggestive of *ecological equivalents*. Ecological equivalents are species found in similar but isolated habitats that look and behave the same although they are unrelated phylogenetically. Examples from higher organisms include the mix of mammals found in North America and those found in Australia. Although no placental mammals were originally found in Australia, equivalents of dogs, mice, and deer can be recognized. The complete overlap in biochemical capacity of bacteria that differ genetically may echo the similarity of habitats colonized by these bacteria.

DNA:DNA pairings do not reflect the actual degree of sequence homology among groups of bacteria. Stackebrandt and Goebel (1994) estimated that bacteria with 70% DNA similarity have 96% to 98% DNA sequence identity. Embley and Stackebrandt (1997) observed that strains that show less than 98% sequence similarity rarely have DNA:DNA homology above 60%. From this observation, they suggest that any environmental sequence that shares 98% 16S rRNA sequence or less similarity to known sequences potentially represent new species. These levels of divergence are not trivial and probably reflect deep evolutionary differences at DNA:DNA pairing values less than 70%.

For bacteria that cannot be isolated or cultured from environmental samples, the amplification by PCR of 16S rRNA has been effective in describing some of the hidden diversity of microbial communities. The use of this technique is based on at least two major assumptions (Goodfellow et al., 1997): that there has been no lateral transfer of 16S rRNA genes and that the amount of evolution or dissimilarity between sequences is representative of the entire genomes of bacteria. If bacteria share rRNA genes between taxa, the use of the method to characterize groups would be compromised. Differences would not be due to evolution acting on these genes but to random mixing of the genes through lateral transfer and recombination. Given that these genes code for the maintenance and tertiary structure of ribosomes, it seems unlikely that they are not highly conserved. Fox et al. (1992) showed that these genes may be too conserved to serve as a means of resolving species differences. Bacteria that take up and incorporate these genes would run the risk of having nonfunctional or functionally reduced ribosomes. This would place these bacteria at a distinct evolutionary disadvantage. Although the possibility exists for transfer of these genes, it has low probability of occurring.

If divergence of bacterial rRNA genes is not tightly correlated with divergence within the genome, the technique cannot be used to differentiate species. Although the sequence of the ribosomal genes is highly conserved, differences do exist among species. These differences are presumably from point mutations and the degree of divergence related to the time since divergence of the species. If other molecules found in organisms do not show similar patterns of divergence, this marker would be considered suspect. Various other molecular markers such as 23S rRNA, ATPase subunits, elongation factors, and RNA polymerase genes, when used to discriminate among bacteria, show a high degree of correspondence with the 16S rRNA results.

Although this and other similar techniques can differentiate bacteria, the resulting designations are operational. We cannot directly infer evolutionary process in the divergence.

Species-specific differences in 16S rRNA genes are primarily found in certain hypervariable regions of the gene. If only part of the 16S rRNA gene is sequenced

similarity cannot be determined. Only by sequencing the entire gene can comparisons between studies be effectively made. Relationships constructed from partial sequences are not stable and change with additional sequence data. To compare environmental samples from various sites and times, it is necessary to sequences the entire gene.

## Application of the Phenetic Species Concept to Bacteria

Most, if not all, bacterial species designations are based on phenetics. The compendium of bacterial nomenclature, Bergey's *Manual*, is based almost entirely on the notion of a phenetic species. Morphological distinctions separate bacteria into basic units such as rods, cocci, and spirochetes, but these classifications are subject to observer error and may change with the physiologic state of the bacteria being observed. Other coarse phenotypic divisions being used include the Gram stain, terminal electron acceptor, aerobe/anaerobe tests, pathogenicity, and various chemical and metabolic properties. To be recognized as a unique species, an isolate must be cultured and the following morphological and biochemical descriptions made: general morphology as observed by light and electron microscopy; various physiological and biochemical tests, including growth on various organic compounds; nitrogen source use; ability to fix nitrogen; pH range; various enzymatic tests; analysis of fatty acids; determination of G + C content; DNA:DNA hybridizations with closely related organisms; and phylogenetic analysis. This list is not comprehensive, but it does give an indication of the work necessary to describe a new species of bacteria.

Metabolic diversity can be estimated and scored using diagnostic strips or multiplates on which the response of an isolate to various organic compounds can be visualized. The metabolic phenotype of an isolate is used to characterize the organism. Theoretically, organisms with increasing similarity in metabolic response are considered the same species. Divergence is an indicator of species separation. Unfortunately, there really are no effective ways to determine degree of similarity or dissimilarity. Potential responses end up along a continuum of responses and no clear divisions are present. Researchers have performed these tests on many medical and industrial bacteria. Levels of discrimination can be quite high in the cases where numerous bacteria have been screened and some estimate of within "species" variation is known. However, environmental isolates present considerable difficulty to these methods because little if any information is available on the metabolic variance of free-living bacteria.

Falkwell and others have shown that bacteria isolated from the deep subsurface are metabolically indistinguishable with various aerobic pseudomonads. However, based on DNA:DNA pairings and 16S rRNA analysis of these isolates, they were found to be very different organisms from the pseudomonads. Another example involves the infamous "Jack-in-the-box" *Escherichia coli* strain. This bacterium has been implicated in the deaths of several people. Based on metabolic tests the organism is very similar to *E. coli* K12. Genetic comparisons between the two strains show little overlap. Is the Jack-in-the-box strain the same or a different species from the well-known *E. coli* strains?

On the other extreme, some bacteria have been shown to have incredible variance in their metabolic potential with low genetic differences. Without knowledge of this

variance, specific isolates collected from the extremes of the metabolic distribution would be considered different species.

Phenetic species are functional in that distinct classifications can be made, but they fail to show evolutionary relationships and give little information on modes of speciation and change.

### Application of the Phylogenetic Species Concept

Moreno (1996) made several observations about the application of the phylogenetic species concept to bacteria. Remember that the phylogenetic species concept is based on two types of data: fossil and molecular sequences. Microbial fossils are rare and provide little discriminatory power. The evolutionary history of microbes has been inferred from molecular data that are expressed as trees or often bushes. Do molecular trees or gene trees represent species trees? This topic is still being debated, but there appear to be some genes that give support to the notion of a nearly one-to-one correspondence between genes and species. These genes often code for essential cell functions, and they are highly conserved sequences. Genes for ribosomal RNA, cytochrome *c*, elongation factor, chaperone proteins, ATPsynthase, and others are often used. The properties of these genes make them good for phylogenetic analysis at levels above the species, but they are usually not very good at discriminating between closely related sister species. Sometimes, the genes of interest may be duplicated in the same chromosome or be found on different chromosomes in the same organism and potentially increase the diversity within the same group. Sequence differences can be used to discern between closely related bacteria and determine if they are species but there are no clear guidelines on what the limits should be. At some levels of resolution, every individual can be classified as a species, and the concept becomes devoid of any meaning and usefulness in the study of ecology or evolution.

Figure 3.2 is an example of a tree constructed based on the 16S rRNA gene sequences. These trees are ubiquitous in the literature. Unknown sequences are compared with and against known sequences and degree of relatedness determined. In many such examples from the literature, major groups of bacteria are identified (e.g., α-Proteobacteria, β-Proteobacteria, γ-Proteobacteria). From these studies, it is fairly clear that most bacteria fall into the known major groups of bacteria but that specific designations are not as easily determined. Because all bacterial DNA in a sample is seldom sequenced, the true taxonomic diversity is unknown.

## Speciation

We have been considering the products of evolution and trying to determine which species concepts are most applicable to bacteria. We have seen that the evolutionary species concept is the most applicable and theoretically sound concept whereas the phenetic species concept is the most frequently applied to bacteria. The process of forming species (i.e., speciation) is an exciting aspect of bacterial evolutionary biology. Bacteria do things differently from other organisms. To understand how the unique evolutionary mechanisms of bacteria affect their speciation we must first discuss speciation as applied to higher organisms.

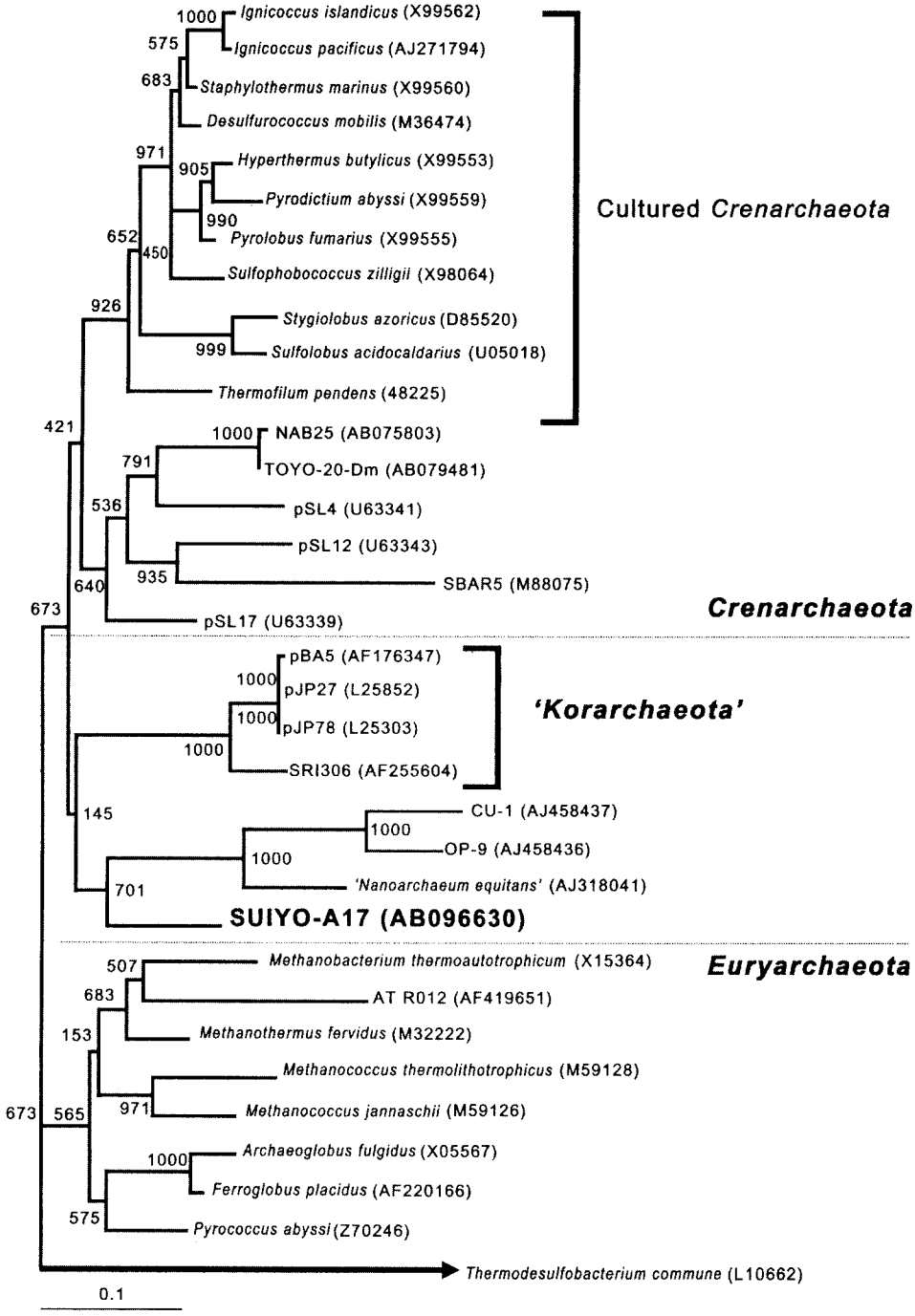


**Figure 3.2** Phylogenetic tree based on 16S rRNA sequences. Sequences of known organisms are used to determine putative relationships among unknown and usually uncultured sequences derived from environmental DNA samples. (Adapted from Figure 6 in Nakagawa T, Ishibashi J, Maruyama A, Yamanaka T, Morimoto Y, Kimura H, Urabe T, Fukui M. Analysis of dissimilatory sulfite reductase and 16S rRNA gene fragments from deep-sea hydrothermal sites of the Suiyo Seamount, Izu-Bonin Arc, Western Pacific. *Appl Environ Microbiol* 70:393–403, 2004.)

Speciation is part of the theoretical basis of the BSC. Most of the aspects of speciation have their genesis in this concept and its development. For most organisms, speciation occurs because there is heritable variation on which natural selection can act and there is some sort of isolation of populations. There can be considerable variation in morphology, physiology, and behavior among individuals found in widely dispersed populations. Variation that is observed due to geography is an indicator of differences in environmental selection acting on the local genotypes. Two patterns of geographic variation are the *cline* and *geographic isolate*.

The cline is a continuous pattern of change in traits and genotypes that results from the mixing of populations along some continuum or along some environmental gradient. This can occur across continental scales for widely dispersed species or over longitudinal gradients with much reduced geographic scales (e.g., rivers). Geographic isolates are populations that have been isolated by some barrier and are not able to share genes with other populations of the same species. Usually, the restriction in gene flow is not complete.

When isolation occurs spatially it is termed *allopatric* or *geographic speciation*. Allopatric situations can arise through *geographic barriers* or *founder effects*.

Allopatric speciation occurs when a population is split apart by the establishing of some barrier that prevents or interrupts gene flow between the two isolated populations. Because each population is under different selective pressures they will accrue unique genetic differences. Given enough time, these accrued differences may be large enough that if the barrier is removed, reproduction between the two populations fails because reproduction cannot occur or the hybrids formed have much lower fitness than either population. The differences accumulated will result in various *isolating mechanisms*, which are characteristics that prevent gene exchange from occurring. If diversification continues these mechanisms can become completely exclusive and the two populations are effectively new species. This definition of a species is based on the BSC. Smith (1980) lists the characteristics of species that are susceptible to allopatric speciation. These characteristics include

1. Species that have low reproductive rates
2. Species that produce few offspring
3. Species that have a long life span
4. Species that have late sexual maturity
5. Species that have high competitive ability
6. Species that have high vagility

The second form of geographic speciation is that brought about by *founder effects*. The name gives some indication of how this process occurs. A founder effect is observed when a single gravid female or a relatively few number of individuals or founders, colonize a new area. In contrast to allopatric speciation, organisms that are susceptible to founder effects demonstrate

1. High reproductive rates
2. Early sexual maturity
3. Large numbers of offspring
4. Short life spans
5. Low competitive ability

Founders are often found on the edge of a species range and experience little gene flow from the center of the population. Founder populations find new suitable habitat that is removed from the main population. These populations generally have much lower genetic diversity and in diploid organisms are more homozygous. As with the geographic barriers once a founder population is established in a new location the population can accumulate different adaptive changes and develop isolating mechanisms.

There are other nongeographic speciation scenarios. Principal among these is *sympatric speciation* or speciation without geographic isolation. Sympatric speciation takes place not on the periphery but in the center of a population living in a patchy environment. Sympatric speciation requires a *stable polymorphism* and *assortative mating*. Assortative mating means that organisms that are adapted to a particular patch or niche tend to mate with one another. Sympatric speciation is thought to occur in plants and in insect parasites of both plants and animals.

## Bacterial Speciation

Sympatric speciation may occur in parasite population without the need to invoke reproductive isolation through adaptive polymorphism and habitat preference (Meeûs et al., 1998). Sympatric speciation may be much more prevalent in parasitic species because hosts provide ample opportunities for niche diversification. This is another argument about the effect of scale in evolutionary and ecological processes. The potential habitats within a single host are immense. Parasites living in one part of a host may in reality be "geographically" isolated from other parasites in the same host. A similar argument could be made for both pathogenic and free-living bacteria. Although bacteria may be living in the same environment as measured by our available methods, they may be isolated from other individuals in both space and time.

Through three unique evolutionary mechanisms, bacteria can and do take up novel DNA and incorporate some or all of this material into their own genomes. These mechanisms are *conjugation*, *transformation*, and *transduction*. Each of these mechanisms is discussed briefly, but extensive reviews of the mechanisms are available. Genetic recombination of this exogenous DNA provides a significant source of genetic variation within bacteria. Not all bacteria are capable of recombination. The potential to recombine novel DNA may be a useful tool to separate bacteria into species.

## Mismatch Repair as a Speciation Mechanism

Vulic et al. (1999) described a mechanism for delimiting bacteria into species based on a specific mutation that affects the ability of bacteria to repair mismatched DNA. Many question whether bacterial species concepts are valid because they do not engage in sexual reproduction like other organisms. However, bacteria do engage in



genetic exchange bet
of naked DNA from
for recombination in
groups, although thi

Vulic et al. (199
(MMR), as a mode
increase mutations a
speciation of bacter
with defective repair

To test whether de
a strain of *E. coli* as
only source of varia
strain to reproduce f
gene function, but o
gene lines, they const
recombination. Base
found that the effect
in those lines that ha
the lines being more
repair system. Most
ient barrier (i.e., repl
tions (<10 years und
greater the inferred (
divergence, the highe

The importance o
ened when we con

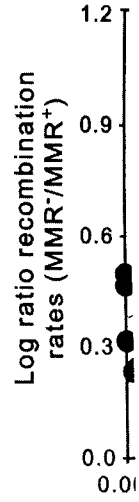


**Figure 3.3** Relationship
(Adapted from Figure 4 in
speciation of bacteria in th
Academy of Sciences, USA

genetic exchange between individuals as mediated by plasmids, viruses, and the uptake of naked DNA from the environment. Based on these mechanisms, a potential exists for recombination in bacteria to be one useful metric for placing bacteria into species groups, although this approach may not be sufficient in all cases.

Vulic et al. (1999) selected one pathway, the methyl-directed mismatch repair (MMR), as a mode of speciation in bacteria. In the MMR pathway, genetic defects increase mutations and recombination rates, and this pathway may be important in speciation of bacteria. Even if speciation does not occur, this and similar pathways with defective repair genes may promote rapid adaptive evolution.

To test whether defects in MMR altered recombination rates, these researchers used a strain of *E. coli* as the founding strain and only source of genetic variation so the only source of variability would be from random mutations. They then allowed the strain to reproduce for nearly 20,000 generations. Some of the lines retained the MMR gene function, but others became defective. From these functional and nonfunctional gene lines, they constructed both donor and recipient genotypes so they could observe recombination. Based on pairwise matings between these independent lines, they found that the effect of mismatch repair systems on recombination rates was greatest in those lines that had evolved nonfunctional repair. This was probably the result of the lines being more sensitive to the recombination-inhibiting effect of a functioning repair system. Most importantly for our discussion, they demonstrated that an incipient barrier (i.e., reproductive barrier) can evolve rapidly during only 20,000 generations (<10 years under their experimental conditions) and influence speciation. The greater the inferred (based on time since evolving and mutation rate) DNA sequence divergence, the higher the rate of recombination (Figure 3.3).

The importance of reproductive isolation in the formation of species is strengthened when we consider that gene flow of any magnitude can swamp genetic

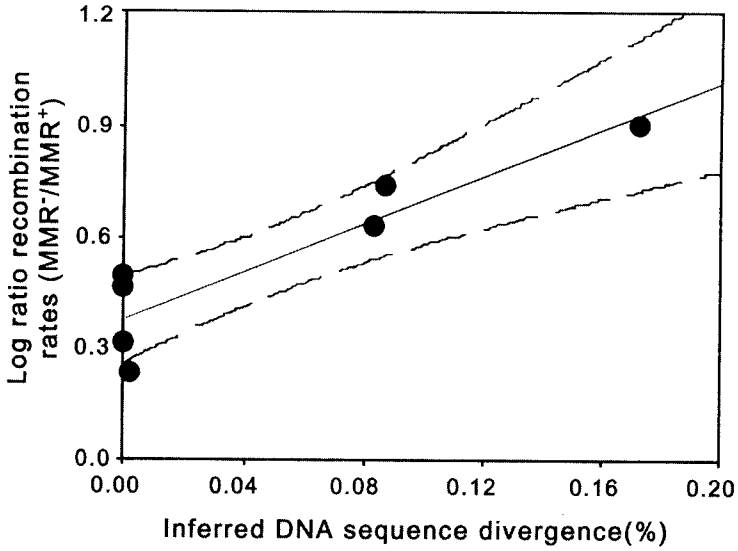


**Figure 3.3** Relationship between the inferred DNA sequence divergence and the rate of recombination. (Adapted from Figure 4 in Vulic M, Lenski RE, Radman M. Mutation, recombination, and incipient speciation of bacteria in the laboratory. *Proc Natl Acad Sci USA* 96:7348–7351, copyright 1999 National Academy of Sciences, USA.)

divergence between sexual populations. Understanding speciation in higher organisms requires that we understand the factors that prevent gene flow and allow genetic divergence. An understanding of speciation in bacteria would similarly involve knowing what drives genetic divergence between bacterial populations and what prevents recombination between divergent genomes.

Sniegowski (1998) considered the MMR hypothesis and found what he identified as two serious problems. First, the effects of genetic divergence in laboratory crosses cannot be directly equated to genetic (sexual) isolation between evolving populations in nature. Sniegowski discussed five steps needed for a successful recombination event to occur in a natural bacterial population:

1. Donor DNA must be taken up by the recipient.
2. The DNA must escape the recipient's restriction enzyme system that cleaves foreign DNA.
3. The donor and recipient DNA must form a *heteroduplex*.
4. The heteroduplex must escape the mismatch repair system, which will abort recombination between divergent sequences.
5. The donor gene product must function in the recipient genetic background. Vulic's hypothesis is concerned with the fourth step only.

Second, rates of recombination measured in prokaryotes cannot be directly related to rates of recombination in eukaryotes. Recombination rates in prokaryotes are controlled by the opportunities for recombination to arise. This is affected by microhabitat distributions of potential recombination genomes. Recombination rates in nature are fairly low, $1 \times 10^{-9}$ for *E. coli* and $1 \times 10^{-8}$ for *Bacillus*. It has been estimated that natural levels of recombination are probably too low to constrain adaptive divergence of bacterial populations in niche-adapted species.

## Rapid Speciation?

*Yersinia* is the genus of microbes responsible for keeping the human population from rapid growth during the middle ages. The Black Death, or plague, was caused by a species of this microbe. Based on molecular analysis of the extant species, it appears that the molecular progenitor for both *Y. pseudotuberculosis* and *Y. pestis* existed about 1 million years ago. These two species are genetically different from their nearest relative *Y. enterocolitica*. Although 1 million years seems like a long time, it is estimated that *Y. pseudotuberculosis*/*Y. pestis* and *Y. enterocolitica* formed separate species approximately 100 million years ago. The question is when did *Y. pseudotuberculosis* and *Y. pestis* separate into distinct species? There are no fixed genetic differences between *Y. pestis* and *Y. pseudotuberculosis*. *Y. pestis* contains no genetic variation. All isolates are genetically similar. Based on molecular analysis a common ancestor could have existed as little as 2,500 years ago (1,000- to 6,000-year range). It appears that *Y. pestis* arose as a single clone from *Y. pseudotuberculosis* because there is no variation among any isolates and because of the similarity. Although there are large phenotype and ecological differences between the two species there is very little genetic distance—so are they the same or different species?



## Operons

One distinctive fea
genes that typicall
evolved by the ass
model is distinct f
nism for the gradu
nism for assembly
it is consistent wit
are found in operc
any selective benef

The selfish oper
zontal transfer int
ing for central me
genomes naïve to
provide for unusua
but nonessential n
operons can have a
sification imparted
the potential for d
these values the r
elucidated.

One estimate of
mate, *Escherichia c*
its divergence from
would allow *E. coli*

Differences amc
species boundaries
ably reflect sets of s
occupies a distinct
functions. No phen
uted to the differer
differences can be
distinct functions n
nomenon is slow a
a novel resource re
absence of strong
intense. A function
critical. However, t

## Genome E

During microbial
mization) occurs, i
their genetic inforn

## Operons

One distinctive feature of bacterial genomes is the *operon*, a cluster of co-transcribed genes that typically provide for a single metabolic function. Bacterial operons have evolved by the assembly of previously unlinked ancestral genes. The selfish operon model is distinct from other models in several ways: It provides a plausible mechanism for the gradual assembly of genes into operons; it provides a selection mechanism for assembly of gene clusters and for their maintenance over evolutionary time; it is consistent with the observation that genes providing for nonessential functions are found in operons; and it does not postulate that gene clusters initially provided any selective benefit to host organisms.

The selfish operon model contends that genes assemble into operons after horizontal transfer into naïve genomes. This model therefore predicts that genes providing for central metabolic functions are least likely to be found in operons, because genomes naïve to these functions are rare. Operons are likely to comprise genes that provide for unusual functions, which can effectively invade naïve genomes (i.e., useful but nonessential metabolic functions). To determine if horizontally inherited selfish operons can have a substantial impact on evolutionary history, the potential for diversification imparted by the gain of introgressed selfish operons must be compared with the potential for diversification generated by mutation and adaptation. To compare these values the rate of horizontal gene transfer among extant genomes must be elucidated.

One estimate of horizontal transfer rate is 31 kb every million years. Using this estimate, *Escherichia coli* has gained and lost nearly 3,000 kb of protein-coding DNA since its divergence from the *Salmonella* lineage. Functions provided by some of these genes would allow *E. coli* to explore novel ecological niches in a rapid and effective manner.

Differences among bacteria may be masked by recombination, which obscures species boundaries. The features that discriminate closely related bacterial taxa probably reflect sets of selective pressures inherent in their individual lifestyles; each species occupies a distinct ecological niche that provides selection for essential niche specific functions. No phenotype distinguishing between *E. coli* and *Salmonella* can be attributed to the differentiation of ancestral genes by point mutations, rather all described differences can be attributed to gain or loss of genes. Although genes providing for distinct functions must ultimately evolve through duplication and divergence, this phenomenon is slow and inefficient and would not allow the competitive exploitation of a novel resource required for bacterial speciation. Such a process would require the absence of strong selection. Novel functions probably evolve when selection is not intense. A function may evolve in a niche where the selection for the function is not critical. However, these functions would not be used to exploit a new niche.

## Genome Economization and Speciation

During microbial starvation, a genome-reducing mechanism (i.e., genome economization) occurs, in which prokaryotic cells in exhausted media can lose a part of their genetic information. If this occurs in nonessential genes, the rate of reproduc-

tion may increase. A cell with a smaller genome has a selective advantage over a cell with a larger genome in certain conditions because the cell does not have to replicate or synthesize the extra material. Differences in cell genomes size of 20% have been observed. If we consider that the size of the prokaryote genome is limited to 9.5 Mb, this limit may explain why primitive cells and modern prokaryotes have similar morphological complexity. Nonessential DNA is often located on plasmids or other mobile genetic elements (e.g., antibiotic resistance). Bacteria can be genetically diverse within and among populations, depending on the ecological conditions they are grown under and the amount of DNA they have culled or taken up through selection.

## Hypermutation

Mutator genotypes with increased mutation rates produce rare beneficial mutations more often than wild-type genotype allowing for faster responses to selection. Bacteria may increase their DNA under favorable conditions and this mechanism, which is capable of restoring lost genetic information, provides an advantage for cells in constantly changing environments. Similar DNA often occurs over large phylogenetic distance leading to a widespread horizontal interspecific gene transfer in bacteria (increasing the probability of the selfish operon). Bacteria are the only organisms that have been selected for the ability to take up exogenous DNA actively and recombine it with their genomes. This uptake may be the most important aspect of their evolution. Some barriers to lateral transfer of DNA among taxa do exist and may act as isolating mechanisms and promote speciation.

Prokaryotic genomes can be divided into two parts: exchangeable and non-exchangeable sequences. The latter ones cannot be transferred functionally between species. Change in these genes cause cell death or result in the cell being less competitive and replaced. Prokaryotes can be characterized not by reproductively isolated genomes, but by reproductively isolated sequences. Bacteria diverge when a part of the original exchangeable sequence becomes non-exchangeable due to constantly changing niches. As a result of spatial isolation (from niche changes), the gene flow within the so-called non-exchangeable sequences between two or more populations is interrupted initiating the process of speciation. During this isolation, the exchangeable genes can be transferred between species. This transfer can cause the differences in genomes size within species. This transfer may strengthen sequence isolation through the acquiring of new properties (ecotypes) and contribute to rapid adaptation to novel environments.

It appears that stationary-phase bacteria under stress (e.g., starvation) sometimes produce mutants in response to the stress. Because mutation is random, both deleterious and beneficial mutants arise from multiple molecular mechanisms that may be different from those found in rapidly growing cells. Some studies have shown that *E. coli* collected from many different habitats worldwide increase their mutation rates in response to starvation conditions (Figure 3.4). In contrast, laboratory strains of *E. coli* do not show as elevated a rate of mutation. Given that most bacteria may be living under stressed conditions, especially starvation stress, these mutations may be very important in microbial evolution. Although these increased mutations may result

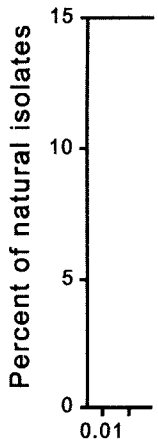


**Figure 3.4** Increase in r
(starvation) as determinec
in new and old starved co
isolates (right of line). (Re
rates in the wild. *Science*

in increased deleteri
bacteria escape loca
conditions.

## Genome Re

*E. coli* and *Haemop*
Inside this subdivisi
very significant wa
*Haemophilus influer*
*E. coli*, a normal re
to many conditions
is an obligate patho

Because of the ver
them to have very d
been defined as copi
having diverged bef
last common ancest
size and a way of li
parasitic life may h
loss of these genes
*H. influenza* has a
frequent enough to
there is little homol

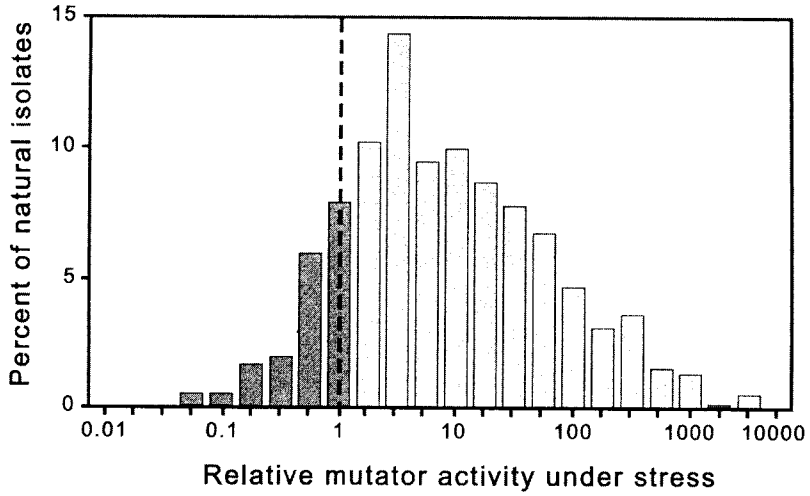


**Figure 3.4** Increase in mutator activity of aging *E. coli* isolates in response to environmental stress (starvation) as determined by measuring the frequency of base substitutions in an RNA polymerase gene in new and old starved colonies. *Dashed line* is the division between new isolates (left of line) and old isolates (right of line). (Reprinted with pemission from Rosenberg SM, Hastings RJ. Modulating mutation rates in the wild. *Science* 300:1382–1383. Copyright 2003 AAAS.)

in increased deleterious mutants, they are probably very important in helping these bacteria escape local extinctions and allowing them to adapt quickly to changing conditions.

## Genome Reduction

*E. coli* and *Haemophilus* both belong to the gamma subdivision of purple bacteria. Inside this subdivision, they appear to be closely related. However, they differ in some very significant ways: genome sizes vary from 4.7 Mb for *E. coli* to 1.8 Mb for *Haemophilus influenzae*. The natural history also is different for the two species. *E. coli*, a normal resident of the gut, can be and often are free living and can adapt to many conditions including changes in salinity and pH to name two. *H. influenzae* is an obligate pathogen requiring specific growth conditions.

Because of the very different preferred habitats of these two species we would expect them to have very different evolutionary ecologies. *Paralous* genes (Fitch, 1970) have been defined as copies issued from a duplication of an ancestral gene, with each copy having diverged before any speciation event. Based on analysis of paralous genes the last common ancestor to *E. coli* and *H. influenzae* was an organism having a genome size and a way of life similar to present-day *E. coli*. The progressive adaptation to parasitic life may have made certain genes dispensable. Alternatively the accidental loss of these genes could have been the stimulus for adopting such a way of life. *H. influenza* has a few genes not found in *E. coli*. Recombination events may be frequent enough to break up large-scale chromosomal arrangements and explain why there is little homology preserved between these two bacteria.