

THE SPECIES CONCEPT AND MICROORGANISMS: THE PAST AND PRESENT OF MICROBIAL TAXONOMY

Microorganisms are fantastically diverse, both genetically and phenotypically. The amount and depth of genetic diversity represented among the named microbial species is greater than that represented among the animals and plants combined. If the distinctions and testing in use today could be applied to all microorganisms (even those that cannot be grown in the laboratory), the number of microbial species easily could reach into the billions. Moreover, the current statistical approaches all tend to underestimate microbial diversity. Linear metrics, in particular, are likely to lead to very large errors because they fail to account for population size, age of clades, and variation in rates of nucleotide substitution. Hence, microbial diversity is vast, and the currently available techniques for estimating the extent of this diversity probably under represent it.

A microbial species concept could provide a way to organize and make sense of this diversity and enable communication of information, but it is a controversial topic. The species categories are obviously of interest to microbial taxonomists and systematists, but the most concerned and vocal parties in the debate are the users of the species designations: the researchers, medical and clinical professionals, forensic scientists, and other scientists who handle microorganisms on a day-to-day basis. These individuals need to unravel bacterial diversity in order to understand and manage ecosystems, human infections, or possible cases of bioterrorism. Species categories, in short, help us to understand how things are and how they got that way. Some other way of bringing meaningful organization to the microbial world, a scheme that avoids the term and essence of "species," might also be useful.

THE CURRENT CONVENTIONS FOR DESIGNATING MICROBIAL SPECIES – STRENGTHS AND LIMITATIONS.

Today, species are defined by pragmatic, arbitrary, and sometimes artificial methods in a formalized polyphasic approach based on 16S rRNA sequences, DNA-DNA hybridization, morphology, physiology, and chemotaxonomy. This system is functional in many ways, and the authorized species categories provide a common dialect that scientists and practitioners can use to convey information. Taxonomists can incorporate new knowledge into a species definition at any stage as new techniques become available, and there are mechanisms for incorporating novel data that may be beneficial to the users of species descriptions.

There are, however, certain difficulties inherent in the current model for designating species. The fixed rules and cut-offs in use today are inadequate, particularly the arbitrary requirements that a species' genome have no more than 70% DNA-DNA relatedness or 97% 16S rRNA identity to its nearest relative of another species. There is also a problem of over-classification among the phenotypes and pathotypes that are of high interest to microbiology and medicine. Gene transfer between microorganisms can create conflicts between genotypic and phenotypic classifications of the same organism. Finally, the microbial species definition is drawn from a limited sample of organisms, but it is extrapolated to the entirety of the microbial world. Hence, only a few microbial lineages predominate and shape our view of all microorganisms. It is unknown how well our understanding of microbial species fits organisms that have yet to be cultivated in the lab.

THE IMPACT OF GENOMICS TECHNIQUES ON SPECIES DESIGNATIONS

Genomic techniques, which include a spectrum of methods that use all or part of an organism's genome to characterize that organism, were originally met with slow acceptance among classical taxonomists, but certain genotypic assays that reveal the metabolic potential of an organism have since been widely adopted. These assays have also been used to reveal how metabolic potential varies among strains and species. It should be noted that *metabolic potential* (the ability to carry out a process) is not the same as *phenotype* (the performance of a process), and genomic techniques are not yet predictive of what, exactly, that organism does in its natural environment. In other words, genomic methods can tell us whether a given gene is present and whether it is expressed under laboratory conditions, but they cannot divine activity in nature. A general acceptance of genomics by the taxonomy community is still pending.

It is recommended that a subcommittee be established within the International Committee on Systematics of Prokaryotes (ICSP) to consider a paradigm shift in the species definition to incorporate the emerging genomic data and to revise and improve use of phenotypic tests.

Genomics has the potential to investigate the relationship between clade structure (depictions of genetic relationships) of phylogenetic trees for

THE HISTORY OF MICROBIAL SPECIES DEFINITIONS, CONVENTIONS (STACKEBRANDT)

One of the most exciting developments in biology in the past 100 years has been the transformation of bacterial systematics from a largely subjective area of study with little relevance to the rest of science into a rigorous and objective discipline that now provides a phylogenetic framework that supports research in all other areas of microbiology. The history of bacterial systematics can be divided into four distinct phases: the phase of “early descriptions” between 1872 and 1900; a phase between 1900 and 1955 in which bacterial physiology and ecology were first explored and described; the era from 1955 until 1980, when many new approaches were developed; and the modern era, from 1980 until today, in which modern DNA techniques were incorporated into the species description. Today, the prevailing opinion among microbiologists appears to be shifting away from demarcating bacterial species using arbitrary and artificial definitions and toward a description of species as ecologically or genetically meaningful entities with a shared phylogenetic heritage.

In 1872, Ferdinand Cohn demonstrated that bacteria could be divided into genera and species using the paradigm proposed for plants and animals by the father of modern taxonomy, Carl Linnaeus. During this early phase of microbial taxonomy, the field was largely dominated by the concerns of medical microbiology: most of the pathogenic bacteria known today were described before the end of the 19th century. At the time, the pattern of properties used to identify new species of bacteria included pathogenic potential, chemical reactions, requirements for growth, and morphology, all of which are still in use today. Bringing order to the bacterial world proved difficult, however. Only two decades after the first bacterial species was described, K. B. Lehmann and R. Neumann denounced the state of bacterial taxonomy as “haphazard and non-scientific” (3).

At the end of the 19th century, bacterial physiology began to have an impact on taxonomy, but systematists still employed a typically “botanical” technique for naming new species; they classified bacteria according to morphology first, and then used physiology to discriminate among the more closely aligned organisms—a mode of classification that did not begin to change until the 1950s. Also during this time, in 1923, the Society of American Bacteriologists (which later became the American Society for Microbiology) presented a report on the characterization and classification of bacterial types that became the basis for Bergey’s Manual, a text that remains the primary reference in bacterial taxonomy even today. By 1955, the field had adopted a pragmatic, arbitrary, and artificial definition for bacterial species: “the type culture together with such other cultures or strains of bacteria that are accepted by bacteriologists as sufficiently closely related” (1). Although this definition was widely accepted, the meaning of “sufficiently closely related” could not be articulated as there was no effective way to determine relatedness at that time.

Between 1955 and the 1980s, bacterial taxonomists developed many new techniques for parsing the bacterial world. Chemotaxonomy, in which the chemical structures of cell constituents are used to differentiate bacteria into relatedness groups, was integrated into species descriptions. In 1961, McCarthy and Bolton presented a means of comparing genetic material through DNA-DNA hybridization (4), a method bacterial systematists rely on to this day to draw

distinctions between closely related species. Numerical phenotypic analysis also emerged during this time, followed by the more sophisticated protein sequence analysis.

In 1965, Zuckerkandl and Pauling evaluated the fitness of various types of biological molecules for deriving the phylogeny of organisms (10). They concluded that the most appropriate molecules are the “semantides,” the molecules that carry genetic information and change slowly over time. In the 1970s, Carl Woese compiled a database of partial rRNA gene sequences and used sequence comparisons to derive a tree of life that put the Bacteria and Eukaryotes on distinctly different branches and uncovered the existence of a third Kingdom, the Archaea. This work expanded the techniques used in protein sequence comparison that were developed in the preceding years. Methodological advances that enabled the cultivation of anaerobes also facilitated progress in developing the tree of life and in adding novel branches to the bacterial and archaeal trees. The combination of molecular, chemotaxonomic, physiological, and other cellular trait analyses also led to new insights into the relatedness among prokaryotic species and revolutionized microbial systematics. (The terms “prokaryotic” and “prokaryote” denote organisms that lack a nucleus, i.e., the Bacteria and Archaea. Although the term is not useful for biological classification, as it denotes the lack of a feature, it is commonly used to designate microscopic organisms that are neither eukaryotes (possessing a nucleus) nor viruses.)

By the 1980s, the list of bacterial names had reached 40,000, a number that many systematists agreed was out of proportion with the sum of bacterial diversity described to date. In 1980, a group of invested microbiologists hewed the list of 40,000 down to a trim 2,500 names designated as “validly published species” (6).

In the following years, nucleic acid analyses, including 16S rRNA sequence analysis, protein-encoding gene sequence analysis, and gene profiling methods, influenced bacterial taxonomy. As new methods were developed, many were integrated into the requirements for defining new species. Today, fewer than 600 new species of bacteria are described every year, in part because of the onerous amount of testing required to ensure a bacterium can be discerned from neighboring species of the same genus.

In 2000, Hagström et al. (2) reported their comparison of 16S rRNA sequence similarities with DNA-DNA reassociation values. They asserted that 97% 16S rDNA sequence identity or lower between two bacteria was sufficiently dissimilar to characterize those bacteria as different species. Later this value was increased to 99% sequence identity in light of new data.

In summary, molecular analyses have enabled bacterial taxonomists to decode the phylogeny of the bacterial world and the distinctions between different types of bacteria are being drawn with a finer and finer brush. However, there is still no consensus in the microbiology community about what exactly constitutes a bacterial species, which tests (and results) are required to identify a bacterium as a unique species, or how to classify those bacteria that cannot be cultivated in the lab.

CURRENT VIEWS OF PAN-SPECIES (FRASER-LIGGETT)

What exactly, is a microbial species? Whole genome sequences are beginning to produce results that can help us answer this question by circumscribing the complete repertoire of genes embraced by a given species. These genes, also called the “pan-genome” include the core genome, which are genes present in all strains, and the non-essential genome, representing genes that vary from one strain to another.

Previous work has recognized the phenomenon of the pan-genome. Welch et al. (9), for example, examined the genomes of three strains of *Escherichia coli*, and found that, although the three had 2,996 genes in common (the core genome), 58% of their genes could only be found in one or two of the strains (the non-essential genome). Other reports have also revealed an extensive amount of genomic diversity among strains of a single species.

Clearly, it is not possible to characterize a species by a single genome sequence, but how many genome sequences are necessary? The answer may vary from species to species. In a study carried out by Tettelin et al. (8), eight genome sequences of strains of the pathogen *Streptococcus agalactiae* were studied to identify how many genomes were needed to fully define the pan-genome of that species. The results revealed several areas of rearrangement within the genome of *S. agalactiae* as well as a number of strain-specific genomic islands larger than 5,000 base pairs. Three hundred and fifty eight genes were nonessential (also called strain-specific). A linear regression analysis showed that as the number of examined genomes increased, the number of core genes declined, eventually paring the core genome of this species down to roughly 1,800 genes. More than 35% of the core genes of the *S. agalactiae* genome are housekeeping genes, while the nonessential genes are dominated by hypothetical and unclassified genes. The authors examined other bacterial species for which multiple whole genome sequences were available and found that the notion of the pan-genome continued to hold.

There are two possible models of the pan-genome, either of which may apply to any given bacterial species: the closed model and the open model. In the closed model, applicable to bacteria that maintain an isolated lifestyle, a bacterium possesses a core set of genes and a smaller number of nonessential genes. Organisms that adhere to the open model possess a core set of genes, but the number of nonessential genes is difficult to pin down since these bacteria exist in multiple environments and have more opportunities for genetic exchange with many other species.

core genes (genes shared by all microbes), including 16S rRNA, and the phenotypes of microorganisms. Some genomics data has challenged the notion of species as coherent clusters of genetically and phenotypically allied organisms. Discrepancies identified between genomic information and data gleaned through more long-standing techniques in taxonomy may be attributable to the use of inappropriate culture conditions, which can fail to elicit the activity of an otherwise important gene discovered through genomics. Genomics can also indicate the presence of interrupted genes, although such genes may not be interrupted in all members of a population. Discrepancies like these can be resolved by examining genomic data on populations. Genomic information can also fail to distinguish ecological differences between species (possibly because of the current inability to interpret sequences effectively). In the end, genomics may support the established taxonomy of some groups and refute the organization of others.

Transcriptomic and proteomic data are important complements to genomic data, as they can reveal a great deal about genetic complexity and the control of gene expression. Understanding the expression of proteins can help scientists draw a link between the phenotype of an organism and its role in an ecosystem. Protein expression data can also enable the discovery of new functions carried out by combinations of expressed genes.

The possible congruity between genomics data and other classification techniques needs to be demonstrated. Real-world test cases, in which the results of 16S rRNA phylogeny or Multi Locus Sequence Typing (MLST) are compared with the results from whole genome sequencing, are recommended.

THE SUCCESS OF rRNA AND OTHER GENES FOR PHYLOGENETIC PURPOSES

Sequencing of the 16S ribosomal RNA gene to identify the class, order, and genus of a microorganism is now an integral part of the approach to microbial taxonomy, but this gene is not useful for identifying many microbes at the species level. At the phylum level, it is difficult to organize relative branching orders using 16S rRNA sequence comparisons. The 16S rRNA gene has been useful for identifying cases of misclassification among named species and for establishing phylogenetic associations. At present, 16S rRNA has proven better than all other genes or gene families for making these types of distinctions, but phylogenies based on multiple genes are better for identifying microbial species.

CONCEPTS, PROCESSES, PATTERNS: WHAT IS THE FUTURE OF MICROBIAL TAXONOMY?

The future of microbial taxonomy is being molded by the work carried out today. The question of whether or not microorganisms are clustered in groups with certain likenesses is pivotal. If microbial diversity is organized in some way, then taxonomy needs merely to find a way to reflect that diversity. If microorganisms do not form groups with meaningful traits and evolutionary history, then organizing microbial diversity will be more difficult, if not impossible. It is also possible that some microorganisms fit nicely into clusters while others do not. In this event, microbial taxonomy may have to embrace multiple approaches to classification. Moving forward in the fields of microbial taxonomy and systematics, research needs to address more fully the issues of clustering, community genomics, and the development of novel approaches for classification and characterization.

SPECIES CLUSTERING IN MICROORGANISMS

Do microorganisms cluster in groups with meaningful commonalities? The answer appears to be both “yes” and “no”; some groups do have species-like habits but others do not. Moreover, the appearance of clustering in a microbial phylogeny depends heavily upon the technique used in the analysis, so that in some cases, for example, several strains can appear to form a tight group in one analysis but may be widely divergent in another analysis. Bacterial and archaeal traits must cluster to some degree, since even random processes could produce bunches of organisms with certain similarities.

The concept of microbial species makes sense only in light of evolution—the interplay of mutation and selection. In some cases, the continuum of life has been “pruned,” and bacteria with certain combinations of traits have died off. In trying to define the boundaries of some microbial species, the boundaries between/among clusters are “fuzzy” or difficult to define.

The problems biologists face in trying to delineate species boundaries are not unique to microbes; often, the species concept is also difficult to apply in eukaryotes. Figuring out how to organize the microbial world in a meaningful way may help biology cope with the problems that are also inherent in the classification of multicellular organisms. More work is needed in modeling and analysis of clustering dynamics.

COMMUNITY GENOMICS AND ENVIRONMENTAL SAMPLING

Community genomics will play an important role in refining the microbial species concept. By sampling natural populations, researchers can acquire the kind of comprehensive data needed to define what a species is and what it is not. If scientists are to make the most of these data, however, they must take care to avoid the common pitfalls associated with environmental studies.

Sampling strategy is critical in community genomics, since non-systematic sampling can lead to biased results. Sampling for microbial genomics should be directed by scientists with expertise in both the environment and the community type under study, which may include ecologists and environmental chemists. Because of the limitations of the analytical tools employed, metagenomic studies often underestimate the diversity present despite high sample complexity. Complexity should be considered as much as possible. Some initiatives could focus on selecting simple environments or low diversity communities or on isolating subsets of complex communities by using cell sorting or other validated enrichment techniques, for example. The environments selected for metagenomics studies should also be well characterized with respect to other features. The ecological, physical, and chemical context needs to be understood so that noted differences between samples can be correlated with metabolism.

It is essential to improve the tracking of metadata (such as the sampling location, the physical properties of the environment, and the handling of samples) associated with community genomics.

Time series and spatial series sampling provide rich information and can help to detect long-term trends. They are encouraged in community genomics investigations.

In order to properly gauge expectations, investigators should have some grasp of the population structure of a community prior to carrying out extensive sequencing. In order to provide needed data for the field, it is appropriate to carry out shallow sampling of many environments and deeper, more thorough sampling in a few environments. In shallower sampling, for example, researchers can compare similar niches to uncover whether they bear the same dominant organism. Deeper sam-

METAGENOMICS AND METASPECIES (DOOLITTLE)

The phenetic and genetic complexity of microorganisms is considerable, and microbiologists have come to realize that a single genome sequence is not sufficient to characterize the wealth of organisms that fall within the boundaries of a "species". Metagenomics offers the promise of looking at biodiversity in a different, unbiased way, independent of any ideas about species. Metagenomics approaches treat the genetic content of communities as one large sample of genes and DNA. It remains to be seen whether studying metagenomes rather than single genomes will give rise to some new concept of species.

Any definition of microbial species is predicated on the existence of clusters of microorganisms with phenetic and genetic commonalities. If such clustering exists (and some data indicates clustering may be due, at least in part, to random birth and death processes), there is no reason to believe clusters will always be produced, or that, when they are, they will be equally tight, or marked by cohesive qualities. Recombination also clouds the situation, leading scientists to different conclusions about the phylogeny of an organism depending on which gene is analyzed. In light of these facts, it can be argued that a unifying species concept cannot be developed to justify the adoption of any single species definition.

Without a general species definition and a uniform species concept, science cannot reasonably answer many general questions about microbial species as a category, such as how many are there, how are they distributed globally or what their population sizes are, at least not with the quantity and quality of data currently available. However, there are ways to work around this dilemma and avoid the pitfalls associated with defining microbial species. It is possible that some of these general questions can be reformulated to avoid the idea of species by using metagenomics approaches. For example, questions about biodiversity may be better asked by invoking the concepts of operational taxonomic units, molecular clustering, and ecotypes. Moreover, characterization of isolates could be streamlined by omitting the current practice of linking an isolate to a species category. Researchers could compare genotype, phenotype, and ecotype data on the isolate directly with archived data on all known isolates, then make further predictions about the isolate based on its relationship to those archived characters.

pling can enable scientists to quantify population level heterogeneity, that is, how representative an isolate is of its population. Focusing on the abundant members of a microbial community has a chance of capturing organisms that carry out the dominant processes (such as carbon cycling), but other keystone processes that are carried out by fewer organisms (such as nitrogen fixation) may be missed.

The potential for gene transfer can impede the identification of species in microbial communities. Research is needed to identify and describe the influence of gene transfer on systematics.

BOTTLENECKS IN MICROBIAL FORENSICS

Microbial forensics is one field in which the question of microbial speciation is of fundamental importance, since tracking and linking microorganisms to individuals and locations relies on an intimate knowledge of the ecology, phylogeny, diversity, and evolution of those organisms. Progress in microbial forensics is currently stymied on a number of fronts. Needs in the field include:

- Methods and analyses that do not require a live organism,
- Documented, validated, statistically supported methods,
- Appropriate high-resolution markers and other match criteria that can establish baselines for comparing microorganisms, and
- Research on microbial evolutionary dynamics and sequencing error rates is needed in order to confidently match samples and differentiate unusual events from the normal background of microbial populations.

NEW TECHNIQUES THAT WILL ENABLE MORE IN-DEPTH SAMPLING

Technical advances have always driven progress in microbiology. There are a number of promising techniques on the horizon that will enable more in-depth sampling and a greater understanding of microbial diversity. These include:

- Single molecule sequencing techniques, including single molecule mRNA sequencing,
- Single cell techniques, including protein-protein interaction tools and genomics tools,
- PCR methods that inhibit the formation of hybrid and chimeric products,
- Optical mapping for determining three dimensional genome topology,
- Tools for *in situ* analysis,
- Technologies that increase the read lengths of nucleotide sequencing,
- Techniques for physically separating microbial cells, and

- Polony sequencing, which enables rapid sequencing of large numbers of individuals from a population.

What Are Species of Microorganisms?

If microbial taxonomy is to fulfill the needs of microbiologists, medical professionals, and others, it must provide a framework that organizes microorganisms in meaningful ways. There is disagreement as to whether the current system, in which microbes are assigned species designations based on certain phenotypic and genetic features, should be renovated, or whether a new system should be built from scratch.

What should the term “species” mean?

The concept of microbial species is in need of improvement, but it is not clear that it needs to be replaced. Reproducible genetic, physiologic, and ecologic clusters of bacteria and archaea have been recognized by a variety of methods, but the best way to arrive at predictable, demonstrable “species” is not known. The research needs outlined later in this report should provide insights into whether the current system should be overhauled or completely supplanted by a new system.

Microorganisms may be divided into species categories based on one of two paradigms: a theory-based model or an operational model. In a **theory-based model**, also referred to as a species concept model, categories are constructed based on an explicit and predictive theory of the mechanisms of speciation. This provides a framework for understanding selection and diversity processes that drive the creation of microbial “species.” The mechanisms of speciation must be understood before a meaningful species theory can be developed. A theoretical speciation framework that takes into account the unique features of microorganisms, including population size, genetic exchange potential, generation times, etc. is needed. Environmental boundaries such as pH differences can create microenvironments that divide populations. A theory-based model should take into account the role of these physical and functional boundaries in speciation.

An **operational-based model** assumes that a satisfactory unifying theory of speciation does not exist and that *de facto* “species” (clusters with commonalities) are revealed by empirical, data-driven analyses. This type of system is much like that in use today, in that it relies on a species definition—a set of criteria and cut-offs for delineating groups that have functional and phylogenetic similarities. In moving forward, there is a lack of compelling data, particularly with respect to ecology, to support the deployment of either a theory-based model or an operationally-based model. More population genetics and phylogenetics information is needed in order to determine which paradigm (or what combination of the two paradigms) is more appropriate for microorganisms.

Presently, the species definition is based on the characteristics of a very limited sample of microorganisms of interest to the medical and microbiological communities; the system devised to fit these organisms has now been extrapolated to the entirety of microbial diversity. Any new system devised for categorizing microorganisms needs to be meaningful for all of microbial life—not just the microbes that have been studied the most.

MICROBIAL FORENSICS (BUDOWLE)

The concept of microbial species is put to work in microbial forensic science. Many nations are faced with the possibility of a biological attack, and discovering, attributing, and prosecuting these cases pose serious challenges for law enforcement, not the least of which is the inscrutable nature of microorganisms themselves. A number of outstanding questions and needs relating to microbial endemism, identity, evolution, and diversity must be addressed in order to better interpret results and gain greater confidence in those interpretations. Scientific validity and rigor are essential for the tools of microbial forensics.

Covert biological attacks are, by definition, more difficult to discover than overt attacks, especially given the high background rates of food-borne illness and other emerging and re-emerging infectious diseases present even in developed nations. Separating endemic cases of disease from the effects of an attack will require an in-depth understanding of the common domestic pathogens and their epidemiology – a significant challenge.

Efforts to compare a biological sample to a piece of evidence can have one of three possible general outcomes: “match,” “excluded as match,” and “inconclusive.” But when the sample and evidence are microorganisms, it may be unclear what constitutes a match and what does not. For example, is a bacterium a match to the evidence if they are the same species? The same strain? In terms of phenotype, at what level of resolution can two microbes be considered identical? Do their entire genomes need to be identical in order to be a match? Given possible errors in sequencing, how realistic is this requirement? If the sample microbe has experienced 50 generations of growth, is it reasonable to expect it to be identical to its parent culture? If not, how different can it be? These questions need to be answered to exploit the full power of microbial forensic analyses.

MICROBIAL FORENSICS

454 SEQUENCING OF MARINE SAMPLES (SOGIN)

It is a little known fact that life in the oceans is dominated by microorganisms. By some estimates, biomass in the world's oceans is 90-98% microbial, but science is only beginning to study the significance of these organisms. A recent investigation by Sogin et al. (7) sought to quantify the diversity of marine bacteria, and their findings indicate that the genetic diversity of bacterial life in the oceans is far more vast than predicted.

The researchers used a tagged sequencing strategy as the cornerstone of their approach. In this method, DNA is harvested from sea water samples and a rapidly evolving, variable region of bacterial 16S ribosomal RNA genes, called V6, is amplified and sequenced in a high-throughput approach called 454 sequencing that enables researchers to construct huge libraries of sequences. Eight deep ocean and hydrothermal vent sites were tested, and a total of 130,000 tagged V6 sequences were compared with a reference database of 44,000 sequences. After identifying the closest 250 reference sequences for a given sequence tag, pair wise alignments were carried out to determine which of the reference sequences were the closest match. Rarefaction analyses of these results indicated that although as many as 3,000 operational taxonomic units (OTUs) could be found in a liter of water, the analyses had not begun to completely capture the diversity present. In other words, the researchers found a huge number of unique sequences in each sample, but they had not even scratched the surface of the diversity present.

A program called DOTUR (Defining Operational Taxonomic Units and estimating species Richness) (5) was used to determine how many clusters of closely related sequences were found in each sample. Three percent or less sequence difference in their 16S rRNA gene is the cutoff used by taxonomists to define the members of a bacterial species. By this ruler, identifying clusters of sequences that are 3% or less divergent as OTUs, about 6,000 taxa could be identified in a liter of sea water. Rarefaction analysis of these data again showed that the bacterial diversity in the samples was enormous – an extrapolation estimated as many as 69,000 OTUs per liter.

In the future, specialized terms to define the criteria used to cluster microorganisms into clusters, such as *genomospecies*. This could facilitate the use of “species” as a designation.

What are some considerations in overhauling the species definition or designing a new system?

There are a number of important questions related to the mechanisms of speciation that need to be addressed before the microbial species concept can be improved or supplanted. They include:

- The dependence of **genetic transfer rates** on niche characteristics. There are enormous differences in the rates of gene transfer between microorganisms living in different environments. For example, free-floating organisms in the marine environment meet up with relatively few exogenous genes as compared with organisms living in close proximity in a biofilm. Any new species paradigm will have to accommodate these differences in exchange rates. Also, the relative contributions of horizontal gene transfer and recombination to speciation must be accommodated but are presently unknown .
- **Focusing on the organisms we can understand.** In developing a new or improved species concept, emphasis should be placed on the groups that are best understood.
- **Accommodating unforeseen organisms.** A new or different framework for designating species would also have to be adaptable to “surprise” organisms that do not follow all the rules known to date. Some bacteria may never be divisible into the classical categories of “species”. The diversity of microorganisms is such that they probably cannot all follow a single set of rules and a pluralism of species concepts will be required to organize the microbial world.
- The selection of **model organisms.** Model organisms must be chosen with an eye to their evolutionary context, since microbes that have emerged late evolutionarily (like *Bacillus anthracis*) may be highly specialized and therefore not representative of the history of related lineages. Also, it is not known how best to sample nature to achieve proper representation to inform and define our concept of speciation.
- **Tolerances.** It is unknown what tolerances should be allowed when assessing what characteristics and criteria will be used to define species.
- The **potential and limitations of genomics** to inform systematics and classification. The full power of these methods is not known. For example, can RNA and protein expression analyses help illuminate speciation? Minor variations in expression may or may not be meaningful.
- The **dominance of “selective advantage.”** If selective advantage is the greatest driver of speciation, then a locus-driven approach to species designations is appropriate. If selective advantage is not dominant, then a genome-driven approach may be appropriate.
- **Different criteria for different lineages.** It is not known how many criteria are required to define a microbial species—it may be that different sets of criteria are required for classification of different lineages.

NEEDED RESEARCH INITIATIVES

Improving our understanding of evolution and speciation is critical to improving microbial taxonomy. The colloquium participants recommended a number of research initiatives to facilitate progress in this field.

SEQUENCING TYPE STRAINS

The international microbial systematics community should coordinate an effort to construct draft genome sequences of each of the roughly 6,500 type strains catalogued by the ICSP. This effort should continue as new species and type strains are described. Acquiring genome sequences for all type strains would significantly advance the integration of genomic information into our understanding of microbial diversity and would enable researchers to map phenotypes to genomes. Once the genomes of the currently described species are sequenced, all new species descriptions should include full genome data as well.

There are a number of considerations and decisions that need to be attended to before a sequencing effort should get underway. For example, it is not known whether some type strains are sufficiently representative of their species to be informative. A method should be developed for validating reference strains as representative of their species.

Another approach to integrating genomic information into our understanding of species could be to gather in-depth sequence data on isolates from several model groups. These groups may include but are not limited to: *E. coli*, *Prochlorococcus*, *Bacillus subtilis*, *Mycobacterium*, *Clostridium*, or *Bacillus cereus*. Sampling strategies should be developed with input from experts in diverse areas of research. Current isolates from multiple natural environments in which the population biology is known must be included.

OTHER RECOMMENDED RESEARCH INITIATIVES

There are a number of areas of research that must be explored to improve our understanding of microbial diversity, how this diversity is organized, and whether (and how) a species concept or definition can be applied to microorganisms. Recommended research initiatives include:

- **Clusters.** Comprehensive, systematic data are needed to uncover whether microorganisms are organized in robust, definable, biologically meaningful clusters.
- The role of **barriers to gene transfer.** Research is needed to discern whether and to what extent barriers to gene transfer like strain-specific phages and sequence divergence contribute to speciation. For example, are barriers to gene transfer permanent or transient and over what time scale are they significant?

- **Multilocus sequence analysis survey (MLSA).** A comparative diversity survey of 1,000 representative isolates representing several named species using MLSA or some other comparative genotypic technique is recommended to establish the extent of recombination among isolates and the degree to which there are natural and meaningful subdivisions among strains.

- **Genomes through time.** It is not known how the strains within a described species evolve over time, but this is a phenomenon that can profoundly affect species designations. Studies are needed to evaluate microbial change over time in the laboratory and in the environment.

- **Linkage between phylogeny and phenotype.** It is not known to what extent genetic relatedness can be used to predict outward characteristics of microorganisms. Research is needed to thoroughly investigate the links between phylogenies constructed from the sequences of core genes and cell phenotype.

- **Improve the definition of core and pan-genomes.** Research is needed to better define the core and pan-genomes and to identify the mechanisms responsible for variation.

- **Ecology.** Research is needed to focus on the ecological pressures responsible for the appearance of clusters of microorganisms.

- **Population genomics.** A combined approach to microbial population genomics that uses both environmental genomics and gene-based analysis of isolates would allow physiological characterization of the microorganisms of interest while avoiding the biases due to solely using cultivation. Research along these lines is needed to estimate selection and provide information about population heterogeneity, size, and age.

- **The biogeography of natural populations.** Biogeography studies are needed to assess the degree of endemism among microorganisms and to determine the relative importance of *in situ* evolution compared to rates of dispersal.

- **Genomic diversity of forensically- and clinically-significant organisms.** In order to correctly attribute the presence of bioterror agents or other disease agents to their source, we need to know more about the diversity of at least a few of these organisms in the environment.

- **Phylogeny-based species descriptions.** It is not known whether phylogeny-based descriptions would be sufficient to delineate species or whether ecological considerations need to figure into a

definition. There is a lack of understanding of the functions of microorganisms *in situ*.

- **Microbial change in new environments.** Understanding what drives change in microorganisms when they are introduced into new environments may contribute to an understanding of clustering and micro clustering.
- **Principles of organization and evolution.** A comparative analysis of low, moderate, and high diversity systems is recommended to uncover whether the principles of organization and evolution at work in tractable, low diversity systems apply across all levels of complexity.

- **Development of integrated bioinformatic tools.** Bioinformatic tools that can evaluate gene content, gene similarities, and phylogeny are needed. A global effort may be needed to accomplish this goal, but proper quality assurance and quality control measures would be needed to ensure reliability.
- **Training.** Training that brings together the fields of microbiology, taxonomy, population biology, and metagenomics is needed.

RECOMMENDATIONS

A subcommittee should be established within the International Committee on Systematics of Prokaryotes to consider a paradigm shift in the species definition to incorporate the emerging genomic data and to revise and use improved phenotypic tests.

The mechanisms of speciation (if speciation exists in microorganisms) must be more thoroughly studied and understood before a meaningful and practical species theory can be developed.

Comprehensive, systematic data are needed to uncover whether microorganisms are organized in robust, definable, biologically meaningful clusters that adhere to the concept of “species.”

There is a lack of compelling data, particularly with respect to ecology, to support the deployment of either a theory-based species model, an operationally-based species model, or a combination of these approaches, to microorganisms. A good deal more population genetics and phylogenetics information is needed in order to determine which paradigm is more appropriate for microorganisms.

The congruence of genomics data and classical taxonomic techniques remains to be demonstrated with rigor. Real-world test cases, in which the results of 16S rRNA phylogeny or MLSA are compared with the results from whole genome sequencing are recommended.

Acquiring draft-quality genome sequences for all type strains would significantly advance the integration of genomic information into our understanding of microbial diversity and would enable researchers to map phenotypes to genomes. The international microbial systematics community should coordinate an effort to construct draft genome sequences of each of the roughly 6,500 type strains catalogued by the ICSP. This effort should continue as new species and type strains are described.

The potential for gene transfer confuses the species situation in microbial communities. The influence of gene transfer barriers such as strain-specific phages and sequence divergence is also unknown. Research is needed to identify and describe the influence of gene transfer and transfer barriers on systematics.

Reference culture collections and repositories are crucial to the study of microbial systematics. A repository for new isolates must be maintained to ensure the ongoing availability of these organisms for future studies.

REFERENCES

Buchanan, R. E. 1955. Taxonomy. *Ann. Rev. Microbiol.* 9:1-20

Hagström, A., J. Pinhassi, and U. L. Zweifel. 2000. Biogeographical diversity among marine bacterioplankton. *Aquat. Microb. Ecol.* 21:231–244.

Lehmann, K. B. and R. O. Neumann. 1896. *Atlas und Grundriss der Bakteriologie und Lehrbuch der Speciellen Bakteriologischen Diagnostik*, 1st edn. München: J. F. Lehmann.

McCarthy, B., J., and E. T. Bolton. 1963. An approach to the measurement of genetic relatedness among organisms. *Proc. Natl. Acad. Sci. USA.* 50: 156-164.

Schloss, P. D., and J. Handelsman. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* 71:1501–1506.

Skerman, V. B. D., V. McGowan, and P. H. A. Sneath, (editors) (1980). Approved lists of bacterial names. *Int. J. Syst. Bacteriol.* 30: 225–420

Sogin, M. L., H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl. 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl. Acad. Sci. USA.* 103: 12115-12120

Tettelin, H., V. Maignani, M. J. Cieslewicz, 43 other authors. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial ‘pan-genome.’ *Proc Natl Acad Sci U S A* 102, 13950–13955.

Welch, R. A., V. Burland, G. Plunkett III, P. Redford, P. Roesch, D. Rasko, E. L. Buckles, S.-R. Liou, A. Boutin, J. Hackett, D. Stroud, G. F. Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz, N. T. Perna, H. L. T. Mobley, M. S. Sonnenberg, and F. R. Blattner. 2002. Extensive mosaic structure revealed by the complete genome sequence of a uropathogenic strain of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 99:17020-17024.

Zuckerandl, E. and L. Pauling. 1965. Molecules as documents of evolutionary history. *J. Theor. Biol.*, 8(2):357-66.