

January 2011

A theory-based pragmatism for discovering and classifying newly divergent bacterial species

Sarah Kopac
Wesleyan University

Frederick M. Cohan
Wesleyan University, fcohan@wesleyan.edu

Follow this and additional works at: <http://wescholar.wesleyan.edu/div3facpubs>

 Part of the [Biodiversity Commons](#), [Ecology and Evolutionary Biology Commons](#), and the [Environmental Microbiology and Microbial Ecology Commons](#)

Recommended Citation

Kopac, Sarah and Cohan, Frederick M., "A theory-based pragmatism for discovering and classifying newly divergent bacterial species" (2011). *Division III Faculty Publications*. Paper 214.
<http://wescholar.wesleyan.edu/div3facpubs/214>

This Contribution to Book is brought to you for free and open access by the Natural Sciences and Mathematics at WesScholar. It has been accepted for inclusion in Division III Faculty Publications by an authorized administrator of WesScholar. For more information, please contact dschnaidt@wesleyan.edu, ljohnson@wesleyan.edu.

Genetics and Evolution of Infectious Diseases

Edited by

Michel Tibayrenc



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Elsevier

32 Jamestown Road London NW1 7BY

30 Corporate Drive, Suite 400, Burlington, MA 01803, USA

First edition 2011

Copyright © 2011 Elsevier Inc. All rights reserved

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangement with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

British Library Cataloguing in Publication Data

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

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

ISBN: 978-0-12-384890-1

For information on all Elsevier publications
visit our website at www.elsevierdirect.com

This book has been manufactured using Print On Demand technology. Each copy is produced to order and is limited to black ink. The online version of this book will show color figures where appropriate.

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

Cover image: Jenny Telleria

2 A Theory-Based Pragmatism for Discovering and Classifying Newly Divergent Bacterial Species

*Sarah Kopac and Frederick M. Cohan**

Department of Biology, Wesleyan University, Middletown, CT, USA

2.1 Introduction

Are bacterial species real? They are real enough to the systematists who classify them, and to the practitioners of microbiology who depend on bacterial classification. Bacterial systematists have routinely identified species as closely related groups that differ in their disease-causing properties, in their ecological roles in biological communities, and in their physiological capacities (Rosselló-Mora and Amann, 2001). To provide this service, systematists have taken a simple and pragmatic approach—to define species as groups (or clusters) of close relatives separated by large gaps in phenotypic and molecular characters (Vandamme et al., 1996; Rosselló-Mora and Amann, 2001). This practical approach has the cachet of approval from no less an evolutionary biologist than Charles Darwin (Darwin, 1859; Mallet, 2008b). Darwin proposed that animal and plant species should be defined as closely related groups that can coexist as phenotypically distinct clusters (Darwin, 1859; Stamos, 2007; Mallet, 2008b), and this is largely the approach taken by bacterial systematists. This cluster-based approach has proved to be remarkably robust, even as the criteria for defining bacterial species have changed over the decades, from being based on phenotype (usually metabolism) to whole-genome similarity (as measured through genome hybridization) to sequence identity (Rosselló-Mora and Amann, 2001). Bacterial systematists have argued about whether the species they recognize are too narrowly or too broadly defined, and whether they are using the best criteria for demarcating species, but they have agreed that species should hold the essential property of being clusters of close relatives with gaps between them (Gevers et al., 2005).

*E-mail: fcohan@wesleyan.edu

However, many microbiologists and most systematists outside of microbiology have understood species to be more than closely related groups separated by gaps (Mayr, 1982; de Queiroz, 2005). They have viewed the species level of taxonomy as having a reality beyond human attempts at classification. Largely under the influence of Ernst Mayr, the property of cohesion has come to be understood to be a quintessential aspect of species (Mayr, 1963, 1982; Templeton, 1989; de Queiroz, 2005). In this view, species are real because they are the largest groups whose diversity is constrained by a force of cohesion. In the case of the highly sexual animals and plants, the force constraining diversity within species is understood to be genetic exchange. In Mayr's Biological Species Concept, speciation requires certain unusual circumstances that allow newly divergent populations to break free of cohesion by recurrent, high-frequency genetic exchange; speciation is therefore understood to be rare (Mayr, 1963). Recently, zoologists have questioned whether animal species are really cohesive across their geographic ranges, and whether cohesion by genetic exchange actually prevents speciation (Mallet, 2008a). This controversy has raised our doubts as to whether bacterial species are cohesive (Cohan, 2010), an issue to which we shall return.

Many concepts of species have been developed since Mayr's Biological Species Concept, and most have in common certain quintessential features, all related to cohesion (de Queiroz, 2005; Cohan, 2010): the diversity within a species is limited by a force of cohesion, species are invented only once, and different species are ecologically distinct and irreversibly separate. In what we might consider Mayrian concepts of species, these essential properties have been extended to other groups where genetic exchange is rare or absent, such as the bacteria (Templeton, 1989; Cohan, 1994). With our colleagues, we have developed the "ecotype" theory of bacterial species, in which the diversity within an ecotype is constrained by recurrent forces of cohesion such as periodic selection or genetic drift (Cohan, 1994, 2010; Ward, 1998; Cohan and Perry, 2007).

Models of species cohesion depend on homogeneity among members of a species. In the case of animals and plants, cohesion across populations by genetic exchange is widely thought to require homogeneity of reproductive features, such that genes can be successfully exchanged (Mayr, 1963; Templeton, 1989). Likewise, the ecotype model is premised on the existence of ecotypes whose members are *ecologically* homogeneous and interchangeable, such that one competitively superior adaptive mutant can replace all other members of the ecotype (Cohan, 1994).

There is an important pragmatic reason for bacterial species to be demarcated as cohesive and ecologically homogeneous units. The animal ecologist Evelyn Hutchinson saw species as groups that should be homogeneous in their physiological, biochemical, morphological, and ecological characteristics (Hutchinson, 1968). He noted that species so defined have the useful property that the characteristics of any individual classified to a species could be easily predicted. While we believe Hutchinson was overly optimistic about the homogeneity of animal and plant species, microbiologists could probably agree that a taxonomy based on homogeneity, if possible, would be extremely beneficial (Cohan and Perry, 2007). For example, under this approach, the total membership of a pathogenic species would have the

same disease-causing properties, the same tissue tropisms, the same transmission properties, and the same host range, while organisms with significantly different properties would be recognized as different species.

It is widely understood that the species recognized by bacterial systematics are far from satisfying the Hutchinsonian property of homogeneity. The named species have long been known to be metabolically and ecologically diverse (Barrett et al., 1986; Cohan et al., 2006; Smith et al., 2006; Walk et al., 2007; Dykhuizen et al., 2008; Hunt et al., 2008; Koepfel et al., 2008; Manning et al., 2008; Walk et al., 2009; Connor et al., 2010). One goal of this chapter is to propose a method to identify and classify ecologically homogeneous groups we may define as Hutchinsonian species, if they exist.

There are reasons to suspect that Hutchinsonian species may be extremely limited in their phylogenetic breadth, that they are perhaps limited to containing as phylogenetically narrow a group as a single cell and its immediate descendants (Doolittle and Zhaxybayeva, 2009). Recent genome comparisons suggest the possibility that, at least in some taxa, extremely close relatives are distinct in their genome content (Welch et al., 2002; Whittam and Bumbaugh, 2002; Tettelin et al., 2005; Lefébure and Stanhope, 2007; Rasko et al., 2008; Touchon et al., 2009; Paul et al., 2010). That is, bacteria may acquire genes by horizontal genetic transfer at such a high rate that the set of homogeneous organisms may be too small to be worth the trouble to recognize as a species entity. The second goal of this chapter is to lay out a protocol for determining the phylogenetic extent of ecological homogeneity.

Approaches to discovering homogeneous, cohesive species of bacteria are handicapped by various features of bacterial ecology and evolution, which make it difficult to recognize the ecological dimensions by which species diverge or the physiological adaptations that underlie the ecological divergence of new species (Cohan and Perry, 2007). This is because we cannot just look at bacteria and infer how they are different ecologically, as we can with birds of different beak size or shape. Also, horizontal genetic transfer is thought to be one mechanism responsible for the formation of new ecotypes (Gogarten et al., 2002; Cohan and Koepfel, 2008; Palenik et al., 2009), and we cannot predict the genes transferred or their donor source. We therefore cannot always anticipate the dimensions of ecological and physiological divergence among new bacterial species, even in groups that are well characterized (Cohan and Perry, 2007).

The discovery of newly divergent bacterial species requires a universal method that is not based on a priori knowledge or intuition about the ecological dimensions of speciation. The approach we outline for discovering the homogeneous, cohesive species of the bacterial world is ecology-blind, where we aim to hypothesize ecotype demarcations from sequence data, confirm the ecological distinctness of ecotypes, and then test for their homogeneity and cohesion, all without a priori knowledge of the ecological dimensions of ecotype distinctness (Cohan and Koepfel, 2008). We will lay out a process to identify groups of organisms that fit into species that are real, in the sense that they are homogeneous and cohesive; we also allow for the possibility that some (perhaps many) groups of bacteria fit only into reified units of close relatives that are neither homogeneous nor cohesive. Finally, we will present a

new pragmatism for bacterial systematics, which will recognize the real, ecologically homogeneous units of bacterial diversity, where practical, and will recognize reified, heterogeneous units of close relatives where necessary.

2.2 Ecological Breadth of Recognized Species

The classification scheme of bacterial systematics focuses on finding species that are significantly different from one another in DNA sequence identity, genome content, and physiology (Rosselló-Mora and Amann, 2001), but places almost no emphasis on ensuring that each individual species is homogeneous in any characteristic (Cohan, 2002; Staley, 2006; Ward et al., 2006). Under this system, two individuals may be in the same species if they show a critical (previously 97%, now 99%) sequence identity in their 16S rRNA genes (Stackebrandt and Ebers, 2006). This degree of genetic diversity allows for enormous ecological (and disease-causing) differences within a species, as illustrated by *Escherichia coli*. Members of *E. coli* may be specialized as pathogens or commensals, and may be specialized to colonize the large intestine or other parts of the body (Touchon et al., 2009; Walk et al., 2009). These are vastly different environments where the bacteria encounter different extracellular secretions, pH, and notable differences in the extracellular matrix, which they must attach to. Moreover, different *E. coli* populations may be specialized to different hosts (Gordon and Lee, 1999) and different outside environments (Walk et al., 2007). The profound ecological and physiological differences among *E. coli* populations are reflected by huge genomic differences, with three divergent populations sharing only 39% of their genes (Welch et al., 2002). Other named species have also been found to contain a high diversity of ecologically, physiologically, and genomically distinct members (Marri et al., 2006; Kettler et al., 2007; Lefébure and Stanhope, 2007; Vernikos et al., 2007; Paul et al., 2010).

How did systematists come to agree to house such a huge amalgam of diversity within the species they recognize? In the case of the animals and plants, humans have developed an “*umwelt*,” a foundation for demarcating natural groups of consequence for survival, through natural selection and cultural evolution (Yoon, 2009). However, the bacteria until recently escaped the attention of human interest in biodiversity, and so systematists of bacteria had to develop a way of seeing and classifying the diversity of bacteria from scratch (Cohan, 2010). Moreover, as we have mentioned, bacterial systematists have not had the advantage of being able to anticipate either the ecological differences between close relatives of bacteria or the physiological differences underlying their ecological divergence.

Bacteriologists were successful from the mid-century on in developing an objectively based *umwelt* for species demarcation. While limited at the time to metabolic and other phenotypic characteristics, “numerical taxonomy” allowed bacterial systematists to develop standard levels of phenotypic diversity within and between species (Sneath and Sokal, 1973; Yoon, 2009) (Figure 2.1A). In principle, species

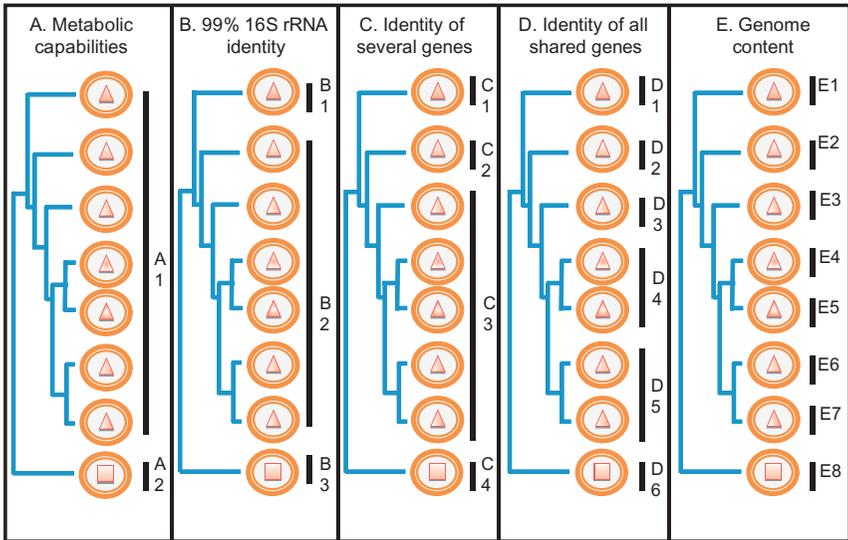


Figure 2.1 Species demarcations under different criteria. Each oval represents a set of closely related cells with identical characteristics of metabolism and ecology, sequences of shared genes, and genome content. Different shapes within the ovals (triangle versus square) represent extremely divergent metabolic capabilities (correlated with ecological function), and variations in shading within a particular shape represent more subtle divergence in metabolism and ecology. The species demarcations under each criterion are indicated by a black vertical bar and a species label (e.g., A1). A. Species were originally defined as groups that differ to a large extent in metabolic capability (indicated by triangle versus square), frequently with much metabolic diversity within each species (indicated here by shading differences within the triangles). B. Defining a species as a group of organisms sharing at least 99% 16S rRNA identity can split the metabolically defined species in the previous panel, as seen here by the splitting of species A1 into B1 and B2. C. Defining species as clusters based on several protein-coding gene sequences can split a 16S-defined species into groups that are each more ecologically homogeneous. This is seen here by the splitting of species B2 into C2 and C3. D. Defining species as clusters based on sequence identity for all shared genes can divide species even further with, for example, species C3 being split into D3, D4, and D5. This may be the most highly resolving method for identifying species based on sequences of shared genes. Within species D4, we can see the possibility that even with this level of resolution for species demarcation, there may still be ecological heterogeneity (indicated by the difference in shading between cells in species D4). Species D5 shows an alternative model where this high level of resolution finds clusters that are ecologically homogeneous, as noted by the same shading patterns among members of D5. E. Defining species by identity of genome content could spuriously split close relatives that are ecologically identical into different species. Note that the two organisms within D5, with the same ecology, are split on the basis of genome content into different species. In this case, E6 and E7 would most likely be different for phage or insertion sequence genes that do not specify ecological niche.

could have been narrowly defined on metabolic grounds, but systematists made a pragmatic, but fateful decision early on to include strains within a species that were heterogeneous in the presence versus absence of many metabolic capabilities. Bacterial species were from the start defined to be extremely diverse in their physiological and, hence, ecological characteristics.

Subsequent incorporation of molecular technology has improved species identification in some important ways (Rosselló-Mora and Amann, 2001; Cohan, 2010). Molecular approaches have provided universal and readily available methods and criteria for species demarcation to all systematists. Using sequence-based criteria, systematists have been able to avoid recognition of polyphyletic groups. Also, because systematics has been based to some extent on whole-genome assays, such as DNA–DNA hybridization, classification has not been deeply affected by recombination across species. Finally, universally applying molecular criteria has led to a pragmatic demarcation scheme that most systematists can agree on (Rosselló-Mora and Amann, 2001) (Figure 2.1B).

Nevertheless, molecular technology has not brought about a refinement in the breadth of diversity subsumed within a recognized taxon. Rather, as each new technology has been embraced, including DNA–DNA hybridization (Wayne et al., 1987), 16S rRNA sequence (Stackebrandt and Ebers, 2006), multilocus sequence analysis (Gevers et al., 2005), and genome-wide average nucleotide identity (ANI) (Konstantinidis and Tiedje, 2005), systematists have attempted to calibrate every new method to yield the existing species taxa (Cohan, 2002; Cohan and Perry, 2007).

Thus, while the approaches of systematics have brought pragmatic solutions for the practice of systematists, we might ask whether these approaches have been pragmatic for microbiologists outside of systematics. The problem is that when systematists reify an amalgam of ecological and functional diversity into a species taxon, other microbiologists tend to assume that each such species constitutes a natural and fundamental unit of biodiversity. This has led to numerous unfortunate consequences for microbiologists outside of systematics. One such consequence is the classification of genes within a recognized species as essential, “core” genes, shared by all “species” members, versus the nonessential, “dispensable” (Tettelin et al., 2005) or “flexible” (Kettler et al., 2007) genes that are shared only by a subset of species members. This dichotomy is false because it is based on the reification of the named species. Also, this gives the impression that those genes held only by one subclade are somehow not essential to the ecology or physiology of that group.

This reification of the core genome may have a real, negative impact on vaccine development. Vaccine development can be based on choosing a target protein that is a member of the core genome (Tettelin et al., 2005). However, if the pathogenic strains of concern constitute only a single ecotype within the species diversity, the choice of vaccine target is unnecessarily restricted to the small core genome of the entire named species, rather than the larger set of genes shared among the pathogen ecotype.

The broad definition of recognized species has led to innumerable errors in population genetic estimation of the critical parameters of evolution. For example, attributing the name *Escherichia coli* to the huge diversity of ecological specialists within the species gives population geneticists the impression that they are dealing

with a group of ecologically interchangeable organisms, such as the members of the fruit fly species *Drosophila pseudoobscura* within a particular habitat. This has led to incorrect application of various algorithms for estimating effective population size and recombination rates (Gordon and Lee, 1999), which assume that the organisms sampled are interchangeable (McVean et al., 2002). In the case of estimating effective population size from sequence diversity, ecological heterogeneity artificially increases the sequence diversity, and thereby the estimate of population size, by conflating the divergence between populations (which is not affected by population size) with divergence within them. Sequence-based estimations of migration rates have also erred by pooling within a taxon a number of ecologically distinct groups (Roberts and Cohan, 1995).

In addition to the errors caused by species reification, the broad brush of systematics has also incurred an opportunity cost for different subfields of microbiology, starting with systematics itself. When a systematist discovers a new species and sees that it can be squeezed into one species taxon, there is no further motivation from systematics to further explore the ecologically distinct clades within the species. Hence, the research in systematics is impoverished by a standard of detail that leaves much of a clade's diversity uncharacterized.

The broad brush also incurs an opportunity cost on epidemiology. In preparation for the next epidemic, epidemiologists might find it useful to identify all the ecologically distinct populations that already exist within a named pathogenic species. We could then prepare for a future epidemic by characterizing, in advance, the disease-causing properties of each population (Cohan and Perry, 2007). Biotechnologists could also take advantage of a more fine-grained systematics of species. After discovering a strain with a valuable enzyme, one could then search for homologs of the enzyme across closely related, ecologically distinct populations, if they were highlighted by taxonomic recognition (Cohan and Perry, 2007; Jensen, 2010).

The molecular revolution has taken us far beyond the early days of systematics, when species demarcation was based entirely on metabolism and other phenotypic traits. Sequencing has now revealed ecologically distinct populations within the recognized species, yet we do not take advantage of this information to refine the demarcations of species. The time has come to incorporate the high resolution of molecular technology into our taxonomy, so that the physiological and ecological diversity we know to exist within the named species can be officially recognized. An important challenge is to develop universal algorithms to analyze sequence data to identify populations that are each ecologically homogeneous and ecologically distinct from one another.

2.3 Models of Bacterial Speciation

In order to integrate ecology into taxonomic classification, we need to take into account the various ways in which bacterial species form and diversity within species is constrained. In the Stable Ecotype model, ecotypes are long-lived, giving

each ecotype ample opportunity to acquire a unique set of neutral mutations in each gene in the genome; also, cohesion results from recurrent periodic selection events within each ecotype lineage (Cohan and Perry, 2007). Ecotypes are founded when a single individual acquires a mutation (or a recombination event) that changes its ecology, through utilizing a new set of resources, thriving under a new set of environmental conditions, or adopting some other change in lifestyle. Since new ecotypes are each founded by a single individual, they start out with zero diversity. A new ecotype is not in direct competition with the members of the parental ecotype because it lives in a different place or uses at least somewhat different resources. For example, a member of a primarily impetigo-causing (skin-infecting) ecotype of *Streptococcus pyogenes* might mutate or acquire a gene that allows it to primarily infect the throat (Bessen, 2009), thus founding a new ecotype. Although the new ecotype may share the same host as the parental ecotype, it is utilizing different host resources, and so the two ecotypes may not experience the same periodic selection events.

Each ecotype remains homogeneous via periodic selection events (Cohan and Perry, 2007). In periodic selection, an individual acquires a genetic change that allows it to be more efficient in its present niche. The individual and its nearly clonal descendants then outcompete all other members of the ecotype and constitute the only lineage that persists into the future (with the exception of a very small level of diversity that survives because of rare recombination) (Cohan, 2005). The result is that the ecotype (surviving through a single lineage) now has extremely little genetic or ecological variation.

How do we know that periodic selection occurs in nature? Periodic selection events are difficult to observe directly in nature. One could try to infer the existence of a history of periodic selection events from an ecotype's low level of sequence diversity (e.g., an ANI of 99%). If one could rule out genetic drift as responsible for limiting the divergence, one might suspect that periodic selection was responsible (since it can purge an ecotype of its sequence diversity regardless of the population size). However, it is difficult to rule out the possibility that an ecotype is simply too young to have accumulated substantial diversity (Koeppel et al., 2008).

It is possible to document periodic selection events when the fitness advantage of an adaptive mutation has transcended beyond its original ecotype, to provide adaptation for another ecotype (the Adapt Globally, Act Locally model) (Majewski and Cohan, 1999). In this case, an adaptive mutation rises to 100% frequency in its original ecotype, becomes transferred to another ecotype by genetic exchange, and then sets in motion a periodic selection event within the recipient ecotype. It is important to keep in mind here that while the adaptive value of the mutation (the allele) transcends beyond its ecotype, the fitness of the mutant (the organism) is limited to the ecological niche of its ecotype; thus, each ecotype has its own periodic selection event, although based on the same mutation (Cohan and Perry, 2007). The genome-wide result of such Adapt Globally, Act Locally periodic selection events is that ecotypes remain divergent throughout their genomes, except that the ecotypes become nearly identical in the gene(s) driving the periodic selection, as well as in linked genes transferred between ecotypes (Koeppel et al., 2008).

There are numerous examples of such anomalous localized identity between otherwise divergent ecotypes. For example, we found that two hot spring *Synechococcus* ecotypes, which are adapted to different temperatures and are 13% divergent throughout the shared parts of their genomes, are nearly identical, within and between ecotypes, in their nitrogen-fixing operon (Bhaya et al., 2007). We interpret this as evidence for a periodic selection caused by acquisition of the nitrogen-fixing genes within each ecotype (Koeppel et al., 2008).

One consequence of many recurrent periodic selection events in each of several long-standing ecotypes is that ecotypes are expected to correspond to sequence clusters for any gene in the genome (Palys et al., 1997). This is because, while diversity in each gene in the genome is recurrently purged within an ecotype, different long-standing ecotypes can accumulate unique mutations in every gene. Assuming the Stable Ecotype model, various algorithms have been developed to find the sequence clusters most likely to correspond to ecotypes (Corander et al., 2008; Hunt et al., 2008; Koeppel et al., 2008; Barraclough et al., 2009).

However, we need to consider some alternative models where ecotypes do not correspond to sequence clusters. In some models, more than one ecotype is subsumed within a sequence cluster; in other models, more than one sequence cluster is contained within a single ecotype (Cohan and Perry, 2007). We believe that all of these models apply under certain circumstances, and that the idealized Stable Ecotype model might occur in only a minority of cases (Cohan, 2010).

In the Speedy Speciation model, cohesion occurs through periodic selection (and/or genetic drift), just as in the Stable Ecotype model (Cohan and Perry, 2007). The difference is that speciation is greatly accelerated as in an adaptive radiation, with the practical consequence that there are many newly divergent species that cannot be distinguished by neutral divergence in a small number of randomly chosen genes (i.e., genes not involved in the adaptive divergence between species, such as those used in most multilocus analyses). Depending on the rate of speciation, species could perhaps be distinguished by neutral sequence variation if the whole genome were sequenced in many isolates. Moreover, sequencing of the whole genome may reveal the genes responsible for ecological divergence.

The Species-Less model is profoundly different from the Stable Ecotype model and all models that assume cohesion within species (Cohan and Perry, 2007). This is a model of rapid speciation, as well as rapid extinction, leading to a high turnover of species. In this case, a species might not persist long enough, from its time of origin to its extinction, to undergo any periodic selection events. In the Species-Less model, each ecotype, while ecologically homogeneous, could not be considered a cohesive unit. Like the case of the Speedy Speciation model, where species are cohesive, the Species-Less model will lead to a diversity of ecotypes that cannot be easily distinguished as sequence clusters.

In the Species-Less model, ecotypes evolve not by becoming more efficient in utilizing their current ecological niche, but instead by evolving to invade a new ecological niche. The Species-Less model may apply to the case for pathogens, where immune-escape mutations may each constitute a new ecotype (Achtman and Wagner, 2008; Cohan, 2010). Also, the Species-Less model may apply in cases

where an environment undergoes a succession process, where organisms at a site must adapt to rapidly changing conditions, for example the successions that occur on mine tailings, with pH and oxidation levels changing predictably and quickly (Remonsellez et al., 2009).

The Nano-Niche model also assumes a high rate of speciation, but here cohesion can occur across ecotypes (Cohan and Perry, 2007). In the Nano-Niche model, closely related ecotypes are subtly and only quantitatively different in their ecology. These “nano-niche ecotypes” use the same set of resources and conditions, but they coexist by using their shared resources and conditions *in different proportions*. Not having any unique resources that might constitute a haven from competition from other ecotypes, each ecotype is vulnerable to extinction from competition with other ecotypes. For a time, the various nano-niche ecotypes may coexist while each has its own private periodic selection events. At some point, however, an extremely competitive adaptive mutant (which bears what we call a speciation-quashing mutation) from one ecotype may extinguish not only the other members of its own ecotype, but also other closely related ecotypes (Cohan, 2005). In the Nano-Niche model, divergence among very closely related ecotypes is limited by these speciation-quashing mutations. Many closely related ecotypes might not last long enough to appear as separate sequence clusters, as based on niche-neutral genes not involved in ecological divergence.

The Nano-Niche model may apply to bacterial ecotypes that adapt over a long time to a given host individual (e.g., a commensal or chronic pathogen). The course of evolutionary adaptation to one human body may bring about multiple periodic selection events in that nano-niche population, but the individual hosts might not be different enough to support unique ecotypes into the indefinite future. Any speciation-quashing mutation that makes an individual not just superior in its own host but also in other hosts would put an end to the speciation among the various nano-niche ecotypes. In the Nano-Niche model, extinction of the nano-niche ecotypes might be too rapid for us to discern them as sequence clusters; also, in this case cohesion (the limit to diversification) occurs across ecologically distinct groups. In our quest to identify ecologically homogeneous groups, we should perhaps be satisfied with finding sets of nano-niche ecotypes whose continued coexistence and divergence we predict will end with a speciation-quashing mutation.

In the Recurrent Niche Invasion model, mobile genetic elements such as plasmids or phage may determine bacterial niches (Cohan and Perry, 2007). For example, in the case of *Rhizobium*, a bacterial lineage may acquire a symbiotic plasmid that adapts it as an endosymbiotic mutualist for a particular set of legume hosts; the lineage may then lose that plasmid and gain another, which adapts it to another set of legume hosts. In the Recurrent Niche Invasion model, the dynamics are particularly interesting when there is no specialization among bacteria to different mutualistic plasmids. In this case, plasmids conferring different ecological niches may come and go among a large group of host bacteria, and any adaptive mutant in the bacterial host population will purge the diversity in the entire bacterial host population. Thus, sequence clusters will correspond to a set of interchangeable bacterial hosts that can each accommodate a diverse set of plasmids. Here, any single

sequence cluster we recognize would actually be adapted to a diversity of plasmid-provided ecological niches. Cohesion extends beyond the individual, plasmid-encoded ecotypes to the whole suite of host organisms that can profit from the set of plasmids (Cohan, 2010).

The Cohesive Recombination model provides another mechanism by which ecologically distinct populations will fail to be recognizable as sequence clusters (Cohan and Perry, 2007). As analyzed quantitatively by Hanage et al. (2006) and Buckley (personal communication), bacteria with the highest rates of recombination may exchange genes so frequently that ecotypes do not accumulate sequence divergence in niche-neutral genes, and so we will not be able to discern the ecotypes as distinct sequence clusters. We note, however, that the rates of genetic exchange in bacteria are never sufficient to hinder or reverse adaptive divergence in niche-specifying genes, as we have previously discussed (Cohan and Koeppl, 2008). Thus, while genetic exchange in rapidly recombining bacteria will not prevent ecotype formation, it may prevent our ability to discover ecotypes using niche-neutral sequence diversity.

In some cases, a single ecotype from a given community may fall into several distinct sequence clusters, as seen in the Geotype plus Boeing model (Cohan and Perry, 2007). Provided that a given taxon has not dispersed frequently, geographically isolated members of a single ecotype may diverge into different clusters (geotypes), even while remaining ecologically interchangeable (Papke and Ward, 2004). This would yield a different sequence cluster in each geographical region, a common phenomenon in the systematics of all organisms that do not readily disperse (Whitaker et al., 2003). In the case of bacteria, geotypes may be a source of confusion for systematists if geotypes have historically been isolated but now with modern human transport, their dispersal has been recently accelerated. In this case (the Geotype plus Boeing model), members of one ecotype isolated from a single site may contain multiple clusters representing the ecotype's various, formerly isolated geotypes from all over the world. In addition, patterns of genetic drift can yield multiple, ecologically interchangeable sequence clusters within a single ecotype.

2.4 Algorithms for Identifying Ecotypes

There are multiple tools available for discerning ecotypes from sequence data, including AdaptML (Hunt et al., 2008), Ecotype Simulation (Koeppl et al., 2008), BAPS (Corander et al., 2008), and GYMC (Barraclough et al., 2009). In contrast to the approaches of bacterial systematics, none of these algorithms assumes a universal criterion for demarcation. Rather, each algorithm uses sequence data from the taxon of focus to identify the appropriate sequence divergence criterion for distinguishing ecotypes.

AdaptML differs from the rest in requiring the habitat of isolation as input data, while the others are blind to ecology (i.e., no information about the ecology or habitat of the strains is taken into account in the analysis) (Hunt et al., 2008). Both

approaches have their advantages (Cohan and Koeppl, 2008). AdaptML is useful when associations with certain habitats are suspected, as this algorithm can simultaneously discover ecotypes and confirm their preferences to habitats specified by the investigator. In contrast, the ecology-blind algorithms do not require the researcher to know anything about the potential environmental differences being analyzed. As a result, multiple ecotypes can be found even in environments which were a priori thought to be homogeneous. However, because Ecotype Simulation does not incorporate ecological data, it cannot confirm the ecological distinctness of ecotypes, and so additional tests must be performed to independently confirm that the clusters are ecologically distinct. Thus far, the clusters found by Ecotype Simulation have consistently been shown to be significantly distinct from one another in their habitat associations; in many cases, the algorithm has hypothesized multiple ecotypes within recognized species (Cohan et al., 2006; Ward et al., 2006; Koeppl et al., 2008; Connor et al., 2010). Likewise, AdaptML has identified ecologically distinct populations within recognized species of *Vibrio* (Hunt et al., 2008). The ecotypes identified by Ecotype Simulation and AdaptML have largely been the same, although in some cases Ecotype Simulation hypothesizes a diversity of ecotypes within one ecotype found by AdaptML (Connor et al., 2010; Melendrez et al., unpublished data). We believe this is because AdaptML can only detect those ecotypes that are different in preferences for habitats anticipated by the investigator. Less frequently, AdaptML hypothesizes multiple ecotypes within one ecotype demarcated by Ecotype Simulation (Connor et al., 2010).

Ecotype Simulation is the only algorithm of its kind that incorporates both periodic selection and speciation. The algorithm is thereby designed to find ecotypes that are each subject to periodic selection, and it estimates how often diversity is purged and how often new ecotypes are formed (Koeppl et al., 2008).

For all of the aforementioned models, the resolution of the analysis depends upon the rates of evolution of the genes analyzed. We have observed that some newly divergent ecotypes may be discerned with protein-coding gene diversity but not with 16S rRNA sequences (Ward et al., 2006; Koeppl et al., 2008). This is because 16S rRNA sequences offer fewer informative sites than protein-coding genes, especially a concatenation of multiple protein-coding genes (Palys et al., 2000) (Figure 2.1B,C). Analysis of a concatenation of all the shared genes (orthologs) among genomes, made possible by whole-genome sequencing, may be the most discerning approach to finding newly divergent ecotypes (Figure 2.1D).

2.5 Confirming the Ecological Distinctness of Ecotypes

If we knew the Stable Ecotype model to apply universally, we could identify ecotypes with confidence as sequence clusters for any gene in the genome (Cohan and Perry, 2007). However, under some models, particularly the Geotype plus Boeing model and models with strong genetic drift, one ecotype may contain multiple sequence clusters. Therefore, each sequence cluster deemed to be an ecotype must be confirmed to be ecologically distinct.

Ecotypes can be confirmed as ecologically distinct when they are significantly different in their habitat associations (Ward et al., 2006; Koeppl et al., 2008). It is important to note that ecotypes need not be absolutely specialized to different habitats (Hunt et al., 2008). Closely related species frequently are at least somewhat ecologically generalized, and coexist by quantitative differences in habitat distinction (Hunt et al., 2008). Various methods are available for confirming quantitative habitat preferences, such as contingency tests (such as Fisher's Exact Test) and AdaptML, designed for bacterial ecotypes (Hunt et al., 2008). We have confirmed the ecological distinctness of putative ecotypes by their habitat associations, including differences in solar exposure, soil texture, rhizospheres, and elevation for soil *Bacillus* (Koeppl et al., 2008; Connor et al., 2010; Kopac et al., unpublished data), temperature and depth in the photic zone in hot spring *Synechococcus* (Ward et al., 2006; Becraft et al., unpublished data), and host range in *Legionella* (Cohan et al., 2006); differences in associations with particles of different size and with season among ecotypes in *Vibrio splendidus* have been identified and confirmed by AdaptML (Hunt et al., 2008). In addition, many ecologists have noted that very closely related sequence clusters (demarcated by eye, rather than by a computer algorithm) are different in their habitat associations (Brisson and Dykhuizen, 2004; Smith et al., 2006; Walk et al., 2007; Manning et al., 2008; Walk et al., 2009).

A disadvantage of the habitat association approach for confirming ecotypes is that it requires the investigator to anticipate the ecological dimensions by which ecotypes diverge. We have previously suggested a different approach where the habitat differences need not be initially known (Cohan and Koeppl, 2008). Each environment from which the focus taxon is isolated may be biotically characterized by identifying the other microbes that live there, using high throughput sequencing of 16S rRNA from each site. Then, environments may be clustered by community type, and one may test whether ecotypes of the focus group differ in their associations with different community types. One may even infer the preferred physical habitats of each ecotype from previous ecological knowledge about the dominant microbial players in an ecotype's preferred community.

The ecological distinctness of hypothesized ecotypes may be confirmed by showing that the groups respond to an environmental perturbation in different ways. In some cases, the natural distribution of putative ecotypes may suggest the direction of response of putative ecotypes to a perturbation. For example, when investigators shaded a Yellowstone hot spring mat, this increased the frequencies of ecotypes normally found in the most-shaded depths of the photic zone (Becraft et al., unpublished data). In other cases, putative ecotypes may respond distinctly to a given perturbation, but with no prediction available for the direction of response. For example, disturbing the thermocline of a lake was shown to affect the vertical distribution of different putative ecotypes in distinct ways (Youngblut and Whitaker, unpublished data). When putative ecotypes respond in distinct ways to an environmental perturbation, whether the directions of response were predicted or not, this indicates the ecological distinctness of the hypothesized ecotypes.

The ecological distinctness of ecotypes may also be confirmed by finding physiological differences between ecotypes that adapt them to their preferred habitats.

For example, bacteria may evolve adjustments of their temperature optima by changing the rigidity of their cell membranes, and such changes may be assayed by fatty acid content (Sikorski and Nevo, 2007; Connor et al., 2010). In other cases, investigators may demonstrate that ecotypes grow fastest under laboratory conditions simulating one component of the preferred habitat. For example, ecotypes associated with hotter regions of a Yellowstone hot spring mat were found to have a higher temperature optimum for growth in the laboratory (Allewalt et al., 2006).

The genomic revolution has allowed identification of ecologically distinguishing features even when we do not have any previous knowledge or intuition about them. One approach involves comparing genome content. For example, differences in genome content showed two ecotypes of hot spring *Synechococcus* to differ in nitrogen storage and phosphonate uptake capacity, revealing that a population more distant from the hot spring source is not simply adapted to cooler temperature; it is also adapted to a lifestyle of scavenging elements (N and P) that are less available downstream (Bhaya et al., 2007). Also, genome-wide comparisons of orthologous genes shared across ecotypes can potentially reveal ecological divergence. Shared genes that have participated in the ecological divergence among ecotypes may show an acceleration of divergence, and Vos has argued that detection of such differences can identify ecotypes (as well as the ecological differences among them) (unpublished data).

2.6 Ecological Homogeneity within Ecotypes

Demonstrating an ecological difference among suspected ecotypes appears to be easy, given the success of the approaches we have discussed. It is quite another matter to confirm the ecological homogeneity within a putative ecotype. When one does not know ahead of time which organisms within a suspected ecotype are likely to be different, attempts to identify ecological differences can be expensive and time-consuming.

Doolittle and colleagues have suggested that ecological homogeneity does not even exist beyond the closest of relatives (in the extreme form of this argument, one cell and its immediate offspring) (Doolittle and Zhaxybayeva, 2009). Suspicion that lineages rapidly change their ecology emerged out of a study of genome size variation among closely related isolates of marine *Vibrio* (Thompson et al., 2004). Even bacteria that were indistinguishable by sequence in a marker gene had diverged in genome size. More recently, owing to the ease and economy of fully sequencing genomes, multiple organisms within various named species have been fully sequenced. Members of a species have thus been shown to differ in genomic content that may change the ecology (Rasko et al., 2008; Touchon et al., 2009; Tenaillon et al., 2010).

An important limitation of these studies is that investigators have rarely chosen to sequence extremely close relatives within a named species. Therefore, discovering differences in genome content of ecological import demonstrates only that there

is ecological diversity within a named species, something that was already well known. What is needed is a survey of extremely closely related organisms that have been hypothesized to constitute a single ecotype, as based on sequence analysis. If hypothesized ecotypes are found to be heterogeneous in genome content of ecological significance, this would indicate that the phylogenetic breadth of ecological homogeneity may indeed be extremely limited (Figure 2.1E).

We emphasize that such tests must be careful to show that genome content differences among close relatives are niche specifying. This is because genome content differences between close relatives appear to reflect mostly the coming and going of phage and insertion sequences, and not the acquisition of new ecological traits (Touchon et al., 2009), demonstrating that heterogeneity of genome content does not necessarily imply heterogeneity of ecology.

Full sequencing of close relatives may be the most effective way to both identify and confirm the ecological distinctness of ecotypes. The concatenation of all shared genes will grant extremely high resolution to algorithms such as Ecotype Simulation. Moreover, as we have noted, genome comparisons can also identify the ecological differences among ecotypes. It will be particularly interesting to find whether this most finely resolving approach will find ecologically homogeneous ecotypes, or whether even among closest relatives, there is ecological heterogeneity.

2.7 Are Bacterial Ecotypes Cohesive?

We began with the issue of the reality of bacterial species, whether there is something biologically unique about the level of species. The concept of cohesion has been argued to provide a key dynamic property of species throughout the biological world—that diversity within a species is limited by certain forces but that divergence above the species level is not (Templeton, 1989; de Queiroz, 2005). The ecotype concept (and particularly the Stable Ecotype model) assumes cohesion within ecotypes, in that diversity within an ecotype is limited by periodic selection and genetic drift, but that divergence between them is not. This cohesion requires ecological homogeneity within an ecotype (Cohan and Perry, 2007).

However, ecological homogeneity is not sufficient to ensure cohesion by forces such as periodic selection and drift, which act recurrently over the lifetime of an ecotype. As we have seen in the Species-Less model, it is possible that new, ecologically homogeneous populations may not persist long enough to encounter a periodic selection event before it goes extinct. In a world with a high turnover of bacterial species, with rapid invention and extinction of ecotypes, the only force limiting the diversity within an ecotype would be its short lifetime before extinction. We have previously proposed that some, perhaps most, bacterial ecotypes within a taxon may not represent species-like cohesive groups, while others may be long-lasting and cohesive, and may even extend over broad geographical areas (Cohan, 2010). We have proposed a phylogenetic test to determine whether newly formed ecotypes are cohesive groups (Cohan, 2010). For the purpose of building

systematics that might aim to identify and classify all the ecological diversity within a taxon, it is probably sufficient to focus on finding the ecologically homogeneous clades, without concern for their cohesiveness.

2.8 Incorporating Ecology into Bacterial Systematics

We suggest that systematics should recognize ecologically homogeneous ecotypes rather than the broadly defined, ecologically heterogeneous amalgams currently recognized. To this end, we lay out a protocol for selecting ecotypes that systematists might have the confidence and motivation to recognize. First, we suggest using sequence data to demarcate ecotypes that appear to represent phylogenetic groups with a history of coexistence as ecologically distinct lineages. Ecotypes could be hypothesized by any of the various universal, sequence-based methods, including AdaptML, Ecotype Simulation, GMYC, and BAPS. If such analyses were to be based on many genes, in the extreme the entire set of shared genes in the genome, more newly divergent ecotypes could be resolved.

Second, the most closely related ecotypes should be confirmed to be ecologically distinct from one another by differences in habitat association or in physiology.

Third, in keeping with an important tradition of bacterial systematics, ecotypes should be confirmed to be phenotypically distinct (Rosselló-Mora and Amann, 2001), and we add that ideally the phenotypic differences should confer the ecological niche specificity of the ecotypes.

Fourth, if possible, an ecotype should be confirmed to be ecologically homogeneous, although as we have pointed out, this may be difficult short of sequencing the full genomes of many members of the ecotype.

Fifth, we suggest that we should not be compelled to recognize every ecotype—only those of interest or consequence. This is because some focus taxa may contain multiple, extremely young, ecologically distinct populations that are unlikely to persist into the future (as in the case of the Nano-Niche model). Here we see that there is a conflict between ensuring homogeneity of ecotypes and recognizing only those of potential interest. Thus, the reform we suggest aims to identify the real, ecologically homogeneous groups where possible, but when impractical, we suggest classifying an ecologically heterogeneous clade as an ecotype, provided that it has been identified by sequence-based algorithms as a putative ecotype and has shown to be ecologically distinct from other closely related ecotypes.

We first consider those cases where a recognized, legacy species is found to contain multiple ecotypes, such as is the case for *Bacillus simplex* (Koeppel et al., 2008), *Vibrio splendidus* (Hunt et al., 2008), and probably many cases where sequence clusters within a pathogenic species are known to differ in host range and/or tissue tropism (Gordon and Cowling, 2003; Smith et al., 2006; Walk et al., 2007; Walk et al., 2009). In these cases, we suggest keeping the existing species binomial in order to maintain stability of the taxonomy, but suggest adding a

trinomial “ecovar” epithet to describe the ecotype taxon. For example, an oak-forest-associated and a grassland-associated ecotype within *Bacillus simplex*, from a canyon near Haifa, Israel (Koeppel et al., 2008), might be named *B. simplex* ecovar Alon and *B. simplex* ecovar Esev (based on the Hebrew words for oak and grass). For ecotypes that are found to be outside the phylogenetic range of existing, recognized species, we suggest naming each ecotype as a species.

We believe that the approach we have laid out is pragmatic both for systematists and for those whose work would benefit from a full accounting of the ecological diversity among close relatives. The proposed system is pragmatic because it identifies the likely ecotypes through universally available and applicable techniques of genomics and DNA sequencing, as well as computer algorithms to recognize the ecotypes from sequence diversity patterns. It also does not reify heterogeneous groups by attempting to apply a universal molecular criterion to all bacterial species. Microbiologists outside of systematists would benefit from systematics that would recognize the most recent products of bacterial speciation. Perhaps most importantly, we will more effectively come to know the unique ecological roles played by each member of a vast and diverse microbial community.

Acknowledgment

This work was sponsored by NSF FIBR grant (EF-0328698) and by research funds from Wesleyan University.

References

- Achtman, M., Wagner, M., 2008. Microbial diversity and the genetic nature of microbial species. *Nat. Rev. Microbiol.* 6, 431–440.
- Allewalt, J.P., Bateson, M.M., Revsbech, N.P., Slack, K., Ward, D.M., 2006. Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the octopus spring microbial mat community of Yellowstone National Park. *Appl. Environ. Microbiol.* 72, 544–550.
- Barraclough, T.G., Hughes, M., Ashford-Hodges, N., Fujisawa, T., 2009. Inferring evolutionarily significant units of bacterial diversity from broad environmental surveys of single-locus data. *Biol. Lett.* 5, 425–428.
- Barrett, E.L., Solanes, R.E., Tang, J.S., Palleroni, N.J., 1986. *Pseudomonas fluorescens* biovar V: its resolution into distinct component groups and the relationship of these groups to other *P. fluorescens* biovars, to *P. putida*, and to psychrotrophic pseudomonads associated with food spoilage. *J. Gen. Microbiol.* 132, 2709–2721.
- Bessen, D.E., 2009. Population biology of the human restricted pathogen, *Streptococcus pyogenes*. *Infect. Genet. Evol.* 9, 581–593.
- Bhaya, D., Grossman, A.R., Steunou, A.S., Khuri, N., Cohan, F.M., Hamamura, N., et al., 2007. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J.* 1, 703–713.

- Brisson, D., Dykhuizen, D.E., 2004. *ospC* diversity in *Borrelia burgdorferi*: different hosts are different niches. *Genetics* 168, 713–722.
- Cohan, F.M., 1994. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *Am. Nat.* 143, 965–986.
- Cohan, F.M., 2002. What are bacterial species? *Annu. Rev. Microbiol.* 56, 457–487.
- Cohan, F.M., 2005. Periodic selection and ecological diversity in bacteria. In: Nurminsky, D. (Ed.), *Selective Sweep*. Landes Bioscience, Georgetown, TX, pp. 78–93.
- Cohan, F.M., 2010. Are species cohesive? A view from bacteriology. In: Walk, S., Feng, P. (Eds.), *Bacterial Population Genetics: A Tribute to Thomas S. Whittam*. ASM Press, Washington, DC.
- Cohan, F.M., Koeppl, A.F., 2008. The origins of ecological diversity in prokaryotes. *Curr. Biol.* 18, R1024–R1034.
- Cohan, F.M., Perry, E.B., 2007. A systematics for discovering the fundamental units of bacterial diversity. *Curr. Biol.* 17, R373–R386.
- Cohan, F.M., Koeppl, A., Krizanc, D., 2006. Sequence-based discovery of ecological diversity within *Legionella*. In: Cianciotto, N.P., Abu Kwaik, Y., Edelstein, P.H., Fields, B.S., Geary, D.F., Harrison, T.G., et al., (Eds.), *Legionella: State of the Art 30 Years after Its Recognition*. ASM Press, Washington, DC, pp. 367–376.
- Connor, N., Sikorski, J., Rooney, A.P., Kopac, S., Koeppl, A.F., Burger, A., et al., 2010. The ecology of speciation in *Bacillus*. *Appl. Environ. Microbiol.* 76, 1349–1358.
- Corander, J., Martinen, P., Siren, J., Tang, J., 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* 9, 539.
- Darwin, C., 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.
- de Queiroz, K., 2005. Ernst Mayr and the modern concept of species. *Proc. Natl. Acad. Sci. U.S.A.* 102 (Suppl. 1), 6600–6607.
- Doolittle, W.F., Zhaxybayeva, O., 2009. On the origin of prokaryotic species. *Genome Res.* 19, 744–756.
- Dykhuizen, D.E., Brisson, D., Sandigursky, S., Wormser, G.P., Nowakowski, J., Nadelman, R.B., et al., 2008. The propensity of different *Borrelia burgdorferi* sensu stricto genotypes to cause disseminated infections in humans. *Am. J. Trop. Med. Hyg.* 78, 806–810.
- Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., et al., 2005. Opinion: re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* 3, 733–739.
- Gogarten, J.P., Doolittle, W.F., Lawrence, J.G., 2002. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238.
- Gordon, D.M., Cowling, A., 2003. The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology* 149, 3575–3586.
- Gordon, D.M., Lee, J., 1999. The genetic structure of enteric bacteria from Australian mammals. *Microbiology* 145 (Pt 10), 2673–2682.
- Hanage, W.P., Fraser, C., Spratt, B.G., 2006. Sequences, sequence clusters and bacterial species. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 1917–1927.
- Hunt, D.E., David, L.A., Gevers, D., Preheim, S.P., Alm, E.J., Polz, M.F., 2008. Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* 320, 1081–1085.
- Hutchinson, G.E., 1968. When are species necessary? In: Lewontin, R.C. (Ed.), *Population Biology and Evolution*. Syracuse University Press, Syracuse, NY, pp. 177–186.
- Jensen, P.R., 2010. Linking species concepts to natural product discovery in the post-genomic era. *J. Ind. Microbiol. Biotechnol.* 37, 219–224.

- Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., et al., 2007. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet.* 3, e231.
- Koeppel, A., Perry, E.B., Sikorski, J., Krizanc, D., Warner, W.A., Ward, D.M., et al., 2008. Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2504–2509.
- Konstantinidis, K.T., Tiedje, J.M., 2005. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572.
- Lefébure, T., Stanhope, M.J., 2007. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biol.* 8, R71.
- Majewski, J., Cohan, F.M., 1999. Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. *Genetics* 152, 1459–1474.
- Mallet, J., 2008a. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 363, 2971–2986.
- Mallet, J., 2008b. Mayr's view of Darwin: was Darwin wrong about speciation? *Biol. J. Linn. Soc.* 95, 3–16.
- Manning, S.D., Motiwala, A.S., Springman, A.C., Qi, W., Lacher, D.W., Ouellette, L.M., et al., 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4868–4873.
- Marri, P.R., Hao, W., Golding, G.B., 2006. Gene gain and gene loss in *Streptococcus*: is it driven by habitat? *Mol. Biol. Evol.* 23, 2379–2391.
- Mayr, E., 1963. *Animal Species and Evolution*. Belknap Press of Harvard University Press, Cambridge, MA.
- Mayr, E., 1982. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Harvard University Press, Cambridge, MA, Chapter 6.
- McVean, G., Awadalla, P., Fearnhead, P., 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160, 1231–1241.
- Palenik, B., Ren, Q., Tai, V., Paulsen, I.T., 2009. Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ. Microbiol.* 11, 349–359.
- Palys, T., Nakamura, L.K., Cohan, F.M., 1997. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int. J. Syst. Bacteriol.* 47, 1145–1156.
- Palys, T., Berger, E., Mitrica, I., Nakamura, L.K., Cohan, F.M., 2000. Protein-coding genes as molecular markers for ecologically distinct populations: the case of two *Bacillus* species. *Int. J. Syst. Evol. Microbiol.* 50 (Pt 3), 1021–1028.
- Papke, P.T., Ward, D.M., 2004. The importance of physical isolation in microbial evolution. *FEMS Microbiol. Ecol.* 48, 293–303.
- Paul, S., Dutta, A., Bag, S.K., Das, S., Dutta, C., 2010. Distinct, ecotype-specific genome and proteome signatures in the marine cyanobacteria *Prochlorococcus*. *BMC Genomics* 11, 103.
- Rasko, D.A., Rosovitz, M.J., Myers, G.S., Mongodin, E.F., Fricke, W.F., Gajer, P., et al., 2008. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* 190, 6881–6893.
- Remonsellez, F., Galleguillos, F., Moreno-Paz, M., Parro, V., Acosta, M., Demergasso, C., 2009. Dynamic of active microorganisms inhabiting a bioleaching industrial heap of low-grade copper sulfide ore monitored by real-time PCR and oligonucleotide prokaryotic acidophile microarray. *Microb. Biotechnol.* 2, 613–624.

- Roberts, M.S., Cohan, F.M., 1995. Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution* 49, 1081–1094.
- Rosselló-Mora, R., Amann, R., 2001. The species concept for prokaryotes. *FEMS Microbiol. Rev.* 25, 39–67.
- Sikorski, J., Nevo, E., 2007. Patterns of thermal adaptation of *Bacillus simplex* to the microclimatically contrasting slopes of “Evolution Canyons” I and II, Israel. *Environ. Microbiol.* 9, 716–726.
- Smith, N.H., Gordon, S.V., de la Rua-Domenech, R., Clifton-Hadley, R.S., Hewinson, R.G., 2006. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nat. Rev. Microbiol.* 4, 670–681.
- Sneath, P., Sokal, R., 1973. *Numerical Taxonomy: the Principles and Practice of Numerical Classification*. W.H. Freeman, San Francisco, CA.
- Stackebrandt, E., Ebers, J., 2006. Taxonomic parameters revisited: tarnished gold standards. *Microbiol. Today* 33, 152–155.
- Staley, J.T., 2006. The bacterial species dilemma and the genomic-phylogenetic species concept. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 1899–1909.
- Stamos, D.N., 2007. *Darwin and the Nature of Species*. State University of New York Press, Albany, NY.
- Templeton, A., 1989. The meaning of species and speciation: a genetic perspective. In: Otte, D., Endler, J. (Eds.), *Speciation and Its Consequences*. Sinauer Associates, Sunderland, MA, pp. 3–27.
- Tenaillon, O., Skurnik, D., Picard, B., Denamur, E., 2010. The population genetics of commensal *Escherichia coli*. *Nat. Rev. Microbiol.* 8, 207–217.
- Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., et al., 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.
- Thompson, J.R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D.E., Benoit, J., et al., 2004. Genotypic diversity within a natural coastal bacterioplankton population. *Science* 307, 1311–1313.
- Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., et al., 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* 5, e1000344.
- Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., Swings, J., 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60, 407–438.
- Vernikos, G.S., Thomson, N.R., Parkhill, J., 2007. Genetic flux over time in the *Salmonella* lineage. *Genome Biol.* 8, R100.
- Walk, S.T., Alm, E.W., Calhoun, L.M., Mladonicky, J.M., Whittam, T.S., 2007. Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. *Environ. Microbiol.* 9, 2274–2288.
- Walk, S.T., Alm, E.W., Gordon, D.M., Ram, J.L., Toranzos, G.A., Tiedje, J.M., et al., 2009. Cryptic lineages of the genus *Escherichia*. *Appl. Environ. Microbiol.* 75, 6534–6544.
- Ward, D.M., 1998. A natural species concept for prokaryotes. *Curr. Opin. Microbiol.* 1, 271–277.
- Ward, D.M., Bateson, M.M., Ferris, M.J., Kühl, M., Wieland, A., Koepfel, A., et al., 2006. Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 1997–2008.

- Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I., et al., 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* 37, 463–464.
- Welch, R.A., Burland, V., Plunkett 3rd, G., Redford, P., Roesch, P., Rasko, D., et al., 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 17020–17024.
- Whitaker, R.J., Grogan, D.W., Taylor, J.W., 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 301, 976–978.
- Whittam, T.S., Bumbaugh, A.C., 2002. Inferences from whole-genome sequences of bacterial pathogens. *Curr. Opin. Genet. Dev.* 12, 719–725.
- Yoon, C.K., 2009. *Naming Nature: The Clash Between Instinct and Science*. Norton, New York, NY.