Chapter 11
Population and Comparative Genomics Inform Our Understanding of Bacterial Species Diversity in the Soil

Margaret A. Riley

11.1 Introduction

The diversity of soil bacteria is simply staggering. According to Torsvik and colleagues (Torsvik et al. 1990; Torsvik and Ovreas 2002), species diversity in soil samples is so high that our most powerful methods of estimation provide only the crudest measure of its magnitude. Nonetheless, many such estimates exist, and they suggest that a single gram of soil may contain over 10 billion microbial cells and more than 1,800 bacterial species (Torsvik and Ovreas 2002; Gans et al. 2005; Zhang et al. 2008). An equally compelling estimate is provided by Dykhuizen (Dykhuizen and Dean 2004), who examined levels of genetic diversity in soil bacteria and predicted that 30 g of forest soil contains over half a million species!

As our methods of empirically estimating bacterial diversity improve, so to do our methods of mathematically refining these numbers. For example, some analytical methods now take into account the fact that there are very few common soil species and untold numbers of rare species (Youssef and Elshahed 2009; Schloss and Handelsman 2006). These refinements push our estimates of bacterial diversity to over a million species per gram of soil (Gans et al. 2005). To put these numbers into perspective, similar studies of the human gastrointestinal tract predict a mere 400 distinct bacterial phylotypes (Rajilic-Stojanovic et al. 2007). This staggering level of species diversity is even more remarkable when one considers that the number of prokaryotes reported in the National Center for Biotechnology Information molecular database is only about 15,111 (Sayers et al. 2010). Clearly, the soil represents a vast reservoir of untapped bacterial diversity.

How we will classify this newfound diversity remains an open question. As molecular methods, such as whole genome sequencing, are more widely applied

M.A. Riley
University of Massachusetts, 611 N. Pleasant Street, Amherst, MA 01003, USA
e-mail: riley@bio.umass.edu

G. Witzany (ed.), *Biocommunication in Soil Microorganisms*, Soil Biology 23,
DOI 10.1007/978-3-642-14512-4_11,

to characterize bacterial diversity, our ability to make taxonomic sense of what we learn is severely challenged. The focus of this review is to explore how we can employ population and species level comparative genomics to provide a rational basis for identifying, and even naming, evolutionary "lineages." In essence, we want to know whether a functional and useful bacterial species concept emerges from the burgeoning genomic information overload.

11.2 Species Classification by 16s rRNA Comparison

Traditional bacterial species designations were based upon extensive phenotypic characterization of a large number of isolates. Although current methods now require the use of 16s rRNA sequence comparisons to identify the closest relatives of a proposed species, phenotype still remains the primary criterion by which species are identified (Rossello-Mora and Amann 2001).This phylo-phenetic bacterial species concept posits that a bacterial species is "a monophyletic and genomically coherent cluster of individual organisms that show a high degree of overall similarity with respect to many independent characteristics, and is diagnosable by a discriminative phenotypic property" (Rossello-Mora and Amann 2001). Numerous studies have revealed clusters of bacterial isolates that share complex phenotypes and these clusters are often designated as species (Shute et al. 1984; Sneath and Stevens 1985; Barrett and Sneath 1994; Mauchline and Keevil 1991; Kirschner et al. 2001). In fact, Cohan uses the mere existence of these clusters as prima facie evidence of the existence of bacterial species. He notes "Bacterial species exist. . . . bacterial diversity is organized into discrete phenotypic and genotypic clusters, which are separated by large phenotypic and genetic gaps, and these clusters are recognized as species" (Cohan 2002).

The earliest attempts to use molecular data to delineate bacterial species involved the use of DNA–DNA hybridization, in which bacterial species were defined as those isolates sharing at least 70% hybridization under standard conditions (Stackebrandt et al. 2002). Given the enormous range of genetic variation detected in different clearly recognized species, it became obvious that a variability cutoff, such as is imposed with hybridization methods, was not appropriate. Levels of variability will vary over the lifetime of a species and will reflect aspects of its life history, particularly the processes by which it adapts to its habitat.

The use of DNA–DNA hybridization has largely been replaced by the use of 16s rRNA sequences to determine the closest relatives of an isolate, combined with extensive phenotype data. Unfortunately, a disturbingly large number of publications report species diversity based solely upon 16s rRNA sequences (Lindh et al. 2005; Drancourt et al. 2004; Clarridge 2004), an approach which has no sound taxonomic basis.

Recently, investigations into microbial species distinctions have sought to incorporate more robust estimates of molecular diversity into the process of species identification. The assumption is that this molecular diversity will fall into discrete clusters that correspond closely with previously identified phenotype-based clusters. In essence, the question has been raised whether DNA sequence variability can be employed to inform the division of a genus into species, to distinguish among similar species, or to address whether bacterial species exist at all (Godoy et al. 2003; Priest et al. 2004; Baldwin et al. 2005; Hanage et al. 2005; Thompson et al. 2005).

11.3 Whole Genome Comparisons Aid Species Classification

The availability of whole genome sequences for multiple isolates of *Escherichia coli* provided our first glimpse into the dynamic nature of a species genome. Glasner and Perna (2004) and Mau et al. (2006) compared six complete genomes of *E. coli* and revealed a highly conserved genomic backbone of more than 3,000 genes, each with greater than 98% sequence similarity among the isolates. Further, Mau et al. (2006) detected a high level of homologous recombination among these shared genes, confirming earlier studies by Roger Milkman (Milkman et al. 1999; Milkman et al. 2003) that the level of within species recombination for what was then considered a "clonal" species was significant. Even more surprising was the observation that the conserved genomic backbone of *E. coli* was interrupted by hundreds of strain-specific "sequence islands". Edwards et al. (2002) provided a similar comparison of *Salmonella enterica*. As in *E. coli*, a backbone of highly conserved genes was identified, each with an average of greater than 99% sequence similarity and a similar pattern of strain-specific unique sequence islands. This pattern of shared and unique sequences appears to be common among many bacteria (Juhas et al. 2007; Coleman et al. 2006; Waterfield et al. 2003).

Studies with subtractive hybridization and comparative genome hybridization revealed that for *Helicobacter pylori*, *E. coli* and *Staphylococcus aureus*, strains within a species share roughly 75–85% of their genome. A comparison of eight genomes of group B Streptococcus reveals a core of 1,806 genes present in every genome and 907 genes absent in one or more. A similar comparison between five genomes of *Streptococcus pyogenes* revealed a comparable level of genomic diversity and predicted that each new genome added approximately 27 strain-specific genes to the species total (pan) genome. In contrast, eight genomes of *Bacillus anthracis* revealed very few strain-specific genes. In fact, after the addition of four genomes to the comparison, no new unique genes were identified. The general pattern that emerges is that members of a bacterial species share some large fraction of their genomes, but often carry unique, strain-specific sequences. The fraction of the genome shared vs. unique varies greatly from one bacterial species to the next.

11.4 Analyses by Genome-Tree Building

The determination of multiple, complete genome sequences of numerous additional bacteria has enabled a new view of the genomic plasticity of several "well-defined" bacterial species. Such an analysis, referred to as genome-tree building, is based on genome-level phylogenetic analysis. This approach involves scaling-up the traditional tree-building approach and analyzing the phylogenetic trees for multiple gene families (ideally, all families represented in many genomes), in an attempt to derive a consensus, organismal phylogeny (Brown and Volker 2004).

The use of multilocus sequence typing (MLST) is one approach to "genome-tree building" (Maiden et al. 1998; Feil et al. 2003). This method permits the analysis of large numbers of related bacterial isolates, which is essential to the determination of species designations (Feil et al. 2003; Hanage et al. 2005). Such studies have confirmed that species designations based upon phenotypic criteria have a corresponding, underlying MLST-based genotypic clustering (Woodward et al. 2000; Whitaker et al. 2005; Thompson et al. 2005; Godoy et al. 2003).

One of the first multigene-based investigations into the microbial species concept was conducted by Wertz et al. (2003), who sequenced six housekeeping genes from a sample of environmental bacteria representing seven species of Enterobacteriaceae (Gordon and FitzGibbon 1999). Molecular phylogenies for each of the genes were inferred and the branching patterns of the resulting trees compared. In each case, isolates from a species formed a monophyletic group, which corresponded precisely with the clusters identified by phenotypic data, and upon which species distinctions were initially delineated (Holt 1994; Rossello-Mora and Amann 2001).

A molecular-based enteric species phylogeny was inferred from the composite data by concatenating the sequences (Wertz et al. 2003) which contained enough phylogenetic signal to resolve all of the interspecies nodes and thus provided a robust estimate of the enteric phylogeny corresponding with the existing phenotype-based phylogeny. The authors concluded that, at least for this sample of enteric lineages, bacterial species clearly do exist and the same species emerge from phenotypic and genotypic data. More recently, a highly robust phylogenetic tree was constructed for 13 gamma proteobacteria using a concatenated alignment of several hundred conserved orthologous proteins (Lerat et al. 2003). Only two of the proteins had incongruent tree topologies in this analysis.

A similar type of investigation was undertaken with Neisseria, in which a sample of housekeeping genes was obtained from nearly one thousand strains and 11 named species (Hanage et al. 2005). Phylogenetic trees were inferred to investigate whether genotypic clusters can be resolved among these highly recombinogenic bacteria and, if so, the extent to which they correspond to named species. Their analysis suggests that the use of concatenated sequences largely buffered the distorting effect of recombination events and resulted in the resolution of clusters corresponding to the three species most numerous in the sample, *N. meningitidis, N. lactamica* and *N. gonorrhoeae*. Comas et al. (2006) go so far as to propose which

genes should be incorporated into these genome-wide surveys for phylogenetic signal. They define "essential" genes, as opposed to "universal" genes, as providing the greatest phylogenetic signal to noise ratio.

These earlier, human-pathogen focused studies have since been expanded to cover a much broader sampling of bacterial diversity. One such investigation involved an Agrobacterium species complex (Popoff et al. 1984; Mougel et al. 2002; Portier et al. 2006). Genome sequence comparisons among members of the genus highlighted a broad range of intra-species divergence within very closely related but distinct species of Agrobacterium. A subsequent study of the same species compared the sensitivity of recombination to DNA sequence divergence across the species complex. Their data supported earlier claims by Majewski (2001) that "bacterial species experience a degree of sexual isolation from genetically divergent organisms since recombination occurs more frequently within species than between species". A genome-based study of the Actinobacteria, a dominant bacterial phylum in the soil, reveals a surprising level of gene syteny and genome conservation. Prior 16S rRNA gene sequence comparisons had recognized 39 families and 130 genera; all of which share a unique molecular synapomorphy: an insertion in the 23S rRNA gene (Roller et al. 1992). How this ancient phylum has retained such a high degree of genome conservation remains a mystery. Similar outcomes have been observed for Haemophilus, Pseudomonas, and Streptococcus, where it was shown that as little as 9% divergence at recombination marker genes results in a drop of three orders of magnitude in recombination efficiency.

Konstantintidis and Tiedje (2004) compared the gene content of 70 closely related and fully sequenced bacterial genomes to identify whether species boundaries exist and to determine the role of the organism's ecology on its shared gene content. They found that levels of sequence similarity on the order of 94% correspond to the traditional 70% DNA–DNA reassociation standard of the current species definition. Notably, a large fraction, e.g., up to 65%, of the differences in gene content within species are associated with bacteriophage and transposase elements, revealing an important role for these elements during bacterial speciation. Their findings are consistent with a species definition that would include a more homogeneous set of strains than provided by the current definition and one that considers the ecology of the strains in addition to their evolutionary distance.

11.5 The Core Genome Hypothesis

Lan and Reeves were the first to recognize the potential link between the observation of shared vs. unique sequences in bacterial genomes and its implication for discriminating bacterial species (Lan and Reeves 1996). They proposed the core genome hypothesis (CGH), which distinguishes between that fraction of the genome (the core) shared by all members of a species and that fraction found in only a subset of the population (the auxiliary). Core genes encode essential metabolic housekeeping and informational processing functions (Feil 2004). They are

ubiquitous in a species and define the species-specific characteristics. Auxiliary genes may or may not be present in a strain and are generally genes that encode supplementary biochemical pathways, which are associated with phage or other mobile elements, or encode products that serve to interact with the external environment. Thus, auxiliary genes serve in the adaptation of strains to local competitive or environmental pressures (Cohan 2002). They are likely to encode antibiotic resistance, novel metabolic functions, toxin production, and the like (Dobrindt and Reidl 2000; Karlin 2001; White et al. 2001).

The CGH has dramatically influenced how bacteriologists think about the nature of bacterial species. Prior to the CGH, the strongest argument against the recognition of bacterial "species" was the simple observation of horizontal gene transfer (HGT) between bacterial lineages. The fact that bacterial species gene pools may not be tightly closed was enough reason for many bacteriologists to conclude species cohesion could not survive such exchanges. This contradicts the clearly demonstrated fact that bacteria exist in phenotypic clusters, which many bacteriologists recognize as species. Even more compelling, it is becoming clear that these well-defined phenotypic clusters correspond to underlying genotype clusters (Woodward et al. 2000; Whitaker et al. 2005; Thompson et al. 2005; Godoy et al. 2003).

Some have argued it is futile to expect a bacterial species to ever be characterized fully at the genome level, particularly since as more genome sequences are obtained, the pan genomes (i.e., the sum of all genes identified within a species) of numerous species continues to grow (Medini et al. 2005; Tettelin et al. 2005). In fact, some predict that hundreds of thousands of genome sequences are required to fully define certain bacterial species (Medini et al. 2005). Others suggest that the wide range of intra-species variation observed for bacterial species reflects the lack of a universal and meaningful species definition (Feil 2004).

Many ecological and evolutionary factors will impact how many unique genes a species pan genome may encode and how much genetic variation it harbors. There is no "one size fits all" concept that can, or should, be applied. In fact, no existing species definition requires that either the pan genome or the level of genetic variation be known to delineate members of a species.

One of the more commonly accepted species concepts is that proposed by Ernst Mayr, the biological species concept (BSC) (Mayr 1942). Mayr proposed that a biological species is comprised of groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups (Mayr 1942). Although Mayr developed this definition specifically for eukaryotes, it can easily be modified to apply to bacteria. The CGH provides a perfect backdrop from which to articulate this modification. According to the CGH, a bacterial species is comprised of groups of strains that frequently exchange, or could exchange, core genes, but which are relatively restricted from such exchange with other groups. However, it is important to note that at this juncture, the BSC should not be taken to imply any particular process of speciation, merely that the observation of more gene "sharing" (via recombination and/or lateral gene transfer) is observed within vs. between putative bacterial species.

The CGH predicts that a subset of genes, the core, is present in all, or nearly all, individuals within the species. These are the genes that provide the defining characteristics of the species and are assumed to experience primarily purifying selection, to remove deleterious mutations and maintain existing functions. As a species evolves, its core genome will evolve as a complex of coevolved functions. When transferred between species, such genes will most likely experience a selective disadvantage, as this will disrupt coevolved functions. Such transfer will rarely survive. Thus, core genes will diverge as the species diverge.

In contrast, auxiliary genes will be found in only a subset of individuals within a species. The CGH predicts that these genes experience intermittent positive selection, when their function enhances survival in a varied and ever-changing environment. When such genes are exchanged between species, their functions will often provide a selective advantage to the recipient. Frequent successful transfer between species will serve to limit the divergence of auxiliary genes, relative to the core.

The most specific prediction that emerges from the CGH concerns the rate at which core and auxiliary genes accumulate variability. Core genes will, on average, display a neutral rate of evolution, while auxiliary genes will experience a variety of selective pressures, including diversifying selection (acting to increase levels of variation), directional selection (acting to decrease levels of variation), balancing selection (acting to maintain particular alleles in the species), and purifying selection (acting to weed out deleterious mutations). Thus, the average rate of evolution for auxiliary genes could be just about anything, and the variance around this rate should be extreme. These expectations, based upon the neutral theory (Kimura 1968), are quite useful for testing predictions from the CGH (Fay et al. 2002). However, such tests require population-based samples of multiple genomes per species and, unfortunately, most existing species-based genome samples are chosen to represent the diversity of clinical isolates of human pathogens and thus will often underestimate standing levels of genome diversity. The appropriate data is in the pipeline and should soon be available to permit the sort of population genomics required to address this complex and fascinating matter.

11.6 Conclusion

Although we are on the verge of obtaining the type and amount of genotypic data required to examine bacterial species definitions, it is important to note that there is, in fact, no substantive argument to support the hypothesis that bacterial species do not exist. Hence, the real argument remaining is not do they exist, but rather, how can they exist in the face of potentially high levels of HGT. Our job is to develop an understanding of bacterial evolution rich enough to explain this apparent paradox. The CGH provides a set of testable hypothesis from which to launch this exploration.

References

Baldwin A, Mahenthiralingam E et al (2005) Multilocus sequence typing scheme that provides both species and strain differentiation for the Burkholderia cepacia complex. J Clin Microbiol 43(9):4665–4673

Barrett SJ, Sneath PH (1994) A numerical phenotypic taxonomic study of the genus Neisseria. Microbiology 140(Pt 10):2867–2891

Brown JR, Volker C (2004) Phylogeny of gamma-proteobacteria: resolution of one branch of the universal tree? Bioessays 26(5):463–468

Clarridge JE 3rd (2004) Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev 17(4):840–862, table of contents

Cohan FM (2002) Sexual isolation and speciation in bacteria. Genetica 116(2–3):359–370

Coleman ML, Sullivan MB et al (2006) Genomic islands and the ecology and evolution of Prochlorococcus. Science 311(5768):1768–1770

Comas I, Moya A et al (2006) The evolutionary origin of Xanthomonadales genomes and the nature of the horizontal gene transfer process. Mol Biol Evol 23(11):2049–2057

Dobrindt U, Reidl J (2000) Pathogenicity islands and phage conversion: evolutionary aspects of bacterial pathogenesis. Int J Med Microbiol 290(6):519–527

Drancourt M, Berger P et al (2004) Systematic 16S rRNA gene sequencing of atypical clinical isolates identified 27 new bacterial species associated with humans. J Clin Microbiol 42 (5):2197–2202

Dykhuizen DE, Dean AM (2004) Evolution of specialists in an experimental microcosm. Genetics 167(4):2015–2026

Edwards SV, Fertil B et al (2002) A genomic schism in birds revealed by phylogenetic analysis of DNA strings. Syst Biol 51(4):599–613

Fay JC, Wyckoff GJ et al (2002) Testing the neutral theory of molecular evolution with genomic data from Drosophila. Nature 415(6875):1024–1026

Feil EJ (2004) Small change: keeping pace with microevolution. Nat Rev Microbiol 2(6):483–495

Feil EJ, Cooper JE et al (2003) How clonal is Staphylococcus aureus? J Bacteriol 185 (11):3307–3316

Gans J, Wolinsky M et al (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science 309(5739):1387–1390

Glasner JD, Perna NT (2004) Comparative genomics of E. coli. Microbiol Today 31

Godoy D, Randle G et al (2003) Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, Burkholderia pseudomallei and Burkholderia mallei. J Clin Microbiol 41(5):2068–2079

Gordon DM, FitzGibbon F (1999) The distribution of enteric bacteria from Australian mammals: host and geographical effects. Microbiology 145(Pt 10):2663–2671

Hanage WP, Kaijalainen T et al (2005) Using multilocus sequence data to define the pneumococcus. J Bacteriol 187(17):6223–6230

Holt JG, NR K et al (1994) Williams and Wilkins. Baltimore, Bergey's Manual of Determinative Bacteriology: 787

Juhas M, Crook DW et al (2007) Novel type IV secretion system involved in propagation of genomic islands. J Bacteriol 189(3):761–771

Karlin S (2001) Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. Trends Microbiol 9(7):335–343

Kimura M (1968) Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet Res 11(3):247–269

Kirschner C, Maquelin K et al (2001) Classification and identification of enterococci: a comparative phenotypic, genotypic, and vibrational spectroscopic study. J Clin Microbiol 39(5):1763–1770

Konstantinidis KT, Tiedje JM (2004) Trends between gene content and genome size in prokaryotic species with larger genomes. Proc Natl Acad Sci USA 101(9):3160–3165

Lan R, Reeves PR (1996) Gene transfer is a major factor in bacterial evolution. Mol Biol Evol 13 (1):47–55
Lerat E, Daubin V et al (2003) From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-Proteobacteria. PLoS Biol 1(1):E19
Lindh JM, Terenius O et al (2005) 16S rRNA gene-based identification of midgut bacteria from field-caught Anopheles gambiae sensu lato and A. funestus mosquitoes reveals new species related to known insect symbionts. Appl Environ Microbiol 71(11):7217–7223
Maiden MC, Bygraves JA et al (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA 95(6):3140–3145
Majewski J (2001) Sexual isolation in bacteria. FEMS Microbiol Lett 199(2):161–169
Mau B, Glasner JD et al (2006) Genome-wide detection and analysis of homologous recombination among sequenced strains of Escherichia coli. Genome Biol 7(5):R44
Mauchline WS, Keevil CW (1991) Development of the BIOLOG substrate utilization system for identification of Legionella spp. Appl Environ Microbiol 57(11):3345–3349
Mayr E (1942) Systematics and the origin of species. Columbia University Press, New York
Medini D, Donati C et al (2005) The microbial pan-genome. Curr Opin Genet Dev 15(6):589–594
Milkman R, Jaeger E et al (2003) Molecular evolution of the Escherichia coli chromosome. VI. Two regions of high effective recombination. Genetics 163(2):475–483
Milkman R, Raleigh EA et al (1999) Molecular evolution of the Escherichia coli chromosome. V. Recombination patterns among strains of diverse origin. Genetics 153(2):539–554
Mougel C, Thioulouse J et al (2002) A mathematical method for determining genome divergence and species delineation using AFLP. Int J Syst Evol Microbiol 52(Pt 2):573–586
Popoff MY, Kersters K et al (1984) Taxonomic position of Agrobacterium strains of hospital origin. Ann Microbiol (Paris) 135A(3):427–442
Portier P, Fischer-Le Saux M et al (2006) Identification of genomic species in Agrobacterium biovar 1 by AFLP genomic markers. Appl Environ Microbiol 72(11):7123–7131
Priest FG, Barker M et al (2004) Population structure and evolution of the Bacillus cereus group. J Bacteriol 186(23):7959–7970
Rajilic-Stojanovic M, Smidt H et al (2007) Diversity of the human gastrointestinal tract microbiota revisited. Environ Microbiol 9(9):2125–2136
Roller C, Ludwig W et al (1992) Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes. J Gen Microbiol 138 (6):1167–1175
Rossello-Mora R, Amann R (2001) The species concept for prokaryotes. FEMS Microbiol Rev 25 (1):39–67
Sayers EW, Barrett T et al (2010) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 38:D5–D16
Schloss PD, Handelsman J (2006) Toward a census of bacteria in soil. PLoS Comput Biol 2(7):e92
Shute LA, Gutteridge CS et al (1984) Curie-point pyrolysis mass spectrometry applied to characterization and identification of selected Bacillus species. J Gen Microbiol 130(2):343–355
Sneath PH, Stevens M (1985) A numerical taxonomic study of Actinobacillus, Pasteurella and Yersinia. J Gen Microbiol 131(10):2711–2738
Stackebrandt E, Frederiksen W et al (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 52(Pt 3):1043–1047
Tettelin H, Masignani V et al (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci USA 102 (39):13950–13955
Thompson JR, Pacocha S et al (2005) Genotypic diversity within a natural coastal bacterioplankton population. Science 307(5713):1311–1313
Torsvik V, Goksoyr J et al (1990) High diversity in DNA of soil bacteria. Appl Environ Microbiol 56(3):782–787

Torsvik V, Ovreas L (2002) Microbial diversity and function in soil: from genes to ecosystems. Curr Opin Microbiol 5(3):240–245

Waterfield NR, Daborn PJ et al (2003) The insecticidal toxin makes caterpillars floppy 2 (Mcf2) shows similarity to HrmA, an avirulence protein from a plant pathogen. FEMS Microbiol Lett 229(2):265–270

Wertz JE, Goldstone C et al (2003) A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. J Evol Biol 16(6):1236–1248

Whitaker RJ, Grogan DW et al (2005) Recombination shapes the natural population structure of the hyperthermophilic archaeon Sulfolobus islandicus. Mol Biol Evol 22(12):2354–2361

White PA, McIver CJ et al (2001) Integrons and gene cassettes in the enterobacteriaceae. Antimicrob Agents Chemother 45(9):2658–2661

Woodward MJ, Sojka M et al (2000) The role of SEF14 and SEF17 fimbriae in the adherence of Salmonella enterica serotype Enteritidis to inanimate surfaces. J Med Microbiol 49(5):481–487

Youssef NH, Elshahed MS (2009) Diversity rankings among bacterial lineages in soil. ISME J 3 (3):305–313

Zhang H, Xie X et al (2008) Soil bacteria augment Arabidopsis photosynthesis by decreasing glucose sensing and abscisic acid levels in planta. Plant J 56(2):264–273