Microbiology (2003), **149**, 3565–3573 DOI 10.1099/mic.0.26457-0

Gradual evolution in bacteria: evidence from *Bacillus* systematics

Michael Feldgarden,† Noah Byrd‡ and Frederick M. Cohan

Department of Biology, Wesleyan University, Middletown, CT 06459-0170, USA

Correspondence
Frederick M. Cohan
fcohan@wesleyan.edu

Received 2 May 2003
Revised 11 September 2003
Accepted 16 September 2003

The bacterial genome projects have suggested a central role for horizontal transfer in bacterial adaptation, but it is difficult to rule out an adaptive role for ordinary genetic change in existing genes. The bacterial systematics literature can readily address the importance of gene acquisition in adaptive evolution, since phenotypic characterization typically assesses presence versus absence of metabolic capabilities, and metabolic gains and losses are most likely due to horizontal transfer and/or gene loss. Bacterial systematists have not geared their studies toward quantitative differences in metabolic capabilities, which are more likely to involve adjustments of existing genes. Here, quantitative variation in metabolism within and between three closely related *Bacillus* taxa has been assayed. While these taxa show no qualitative (i.e. presence versus absence) differences in resource utilization, they are quantitatively different in utilization of 8 % of 95 resources tested. Moreover, 93 % of the resources tested showed significant quantitative variation among strains within a single taxon. These results suggest that ordinary genetic changes in existing genes may play an important role in adaptation. If these results are typical, future genomically based assays of quantitative variation in phenotype (e.g. microarray analysis of mRNA concentrations) may identify hundreds of genes whose expression has been modified. A protocol is presented for identifying those modifications of gene expression and those gene acquisitions that are most likely to have played a role in adaptive evolution.

INTRODUCTION

The peculiar nature of sex in bacteria opens each bacterium to a world of adaptations. Owing to the promiscuity of bacterial sex, bacteria can take up existing adaptations from widely divergent donor species (Ochman *et al.*, 2000). Moreover, because bacterial recombination is infrequent, sexual promiscuity incurs no significant cost to fitness; uptake of incompatible genes from other species is much too rare to disrupt a species' existing adaptations (Cohan, 1994a).

The bacterial genome projects have suggested a central role for horizontal genetic transfer in bacterial adaptation (Lawrence, 2002; Ochman *et al.*, 2000). For example, much of the capacity for virulence in *Salmonella* stems from horizontally acquired 'pathogenicity islands' (Baumler *et al.*, 1997; Ochman & Groisman, 1996). The potential importance of horizontal transfer is suggested by the observation that a typical bacterial genome has received at least 5–10 % of its genes from extremely divergent donors (Ochman *et al.*, 2000). Lawrence (1997, 1999) has argued that bacteria evolve into new niches principally by acquiring new gene loci and operons from other species.

It is difficult, however, to rule out a role for ordinary genetic change in bacterial adaptation. Like the eukaryotes, bacteria can evolve adaptations by modifying the genes they already have. In some cases, bacteria have evolved adaptations by substituting, through recombination, an existing allele with a homologous allele from another species. Penicillin resistance has been acquired in this way in *Streptococcus* and *Neisseria* (Maynard Smith *et al.*, 1991). Mutations have also played a role in bacterial adaptation. For example, synonymous substitutions have optimized mRNA secondary structure (Katz & Burge, 2003) and amino acid substitutions have minimized the use of energetically costly amino acids (Akashi & Gojobori, 2002; Seligmann, 2003).

In other cases, mutations have allowed a strain to step into a new ecological niche. For example, Sokurenko *et al.* (1998) have shown that single point mutations in the FimH adhesin can enable pathogenesis in *Escherichia coli*. Mutations have also been shown to routinely allow invasion of new niches

†Present address: Department of Ecology and Evolution, SUNY Stony Brook, Stony Brook, NY 11794-5245, USA.

‡Present address: Department of Pediatrics and Neonatology, Box 3179, Duke University Medical Center, Durham, NC 27710, USA.

Results of a principal component analysis for utilization of the 95 substrates examined in this study are available from *Microbiology* Online.

in laboratory microcosms of evolution. Several laboratories have found that a clone and its descendants, when allowed to evolve on their own (and without benefit of recombination), inevitably undergo an adaptive radiation: the original clone evolves into a diversity of populations, each with its own ecological niche (Rainey & Travisano, 1998; Rozen & Lenski, 2000; Treves *et al.*, 1998). Thus, some niche invasions are clearly accessible through mutation. Here we demonstrate a role for systematics in assessing the relative importance of heterologous horizontal transfer versus modification of existing genes in bacterial evolution.

The bacterial systematics literature can readily address the importance of gene acquisition in bacterial evolution, but it is of little help in addressing the importance of mutation and other modifications of existing genes. This is because phenotypic characterization of strains and species has been biased toward assessing presence versus absence of metabolic capabilities. Gains and losses of metabolic capabilities are generally due to heterologous horizontal transfer (as well as deletions of genes) (Lawrence, 1997, 1999; Ochman & Groisman, 1996), although some capabilities can be acquired through point mutation (Hall, 1999; Hall & Malik, 1998; Spiers *et al.*, 2002).

Bacterial systematists have generally not geared their studies toward quantitative comparisons of capabilities – for example, by testing whether strains and species differ in the *rate* of fermentation or growth on various sugars. Such quantitative variation, when it exists, is likely due to changes in existing genes (e.g. by increasing the transcription of an existing operon) (Cooper *et al.*, 2003; Ferea *et al.*, 1999). To our knowledge, assessment of quantitative variation among and within bacterial species has been limited to comparing the percentage of different fatty acids in whole cells (Nakamura, 1998; Roberts *et al.*, 1994, 1996); we know of no extensive systematic studies of quantitative variation in metabolic capabilities in bacteria.

We will presently demonstrate significant quantitative variation in metabolism both within and between three closely related *Bacillus* taxa, all formerly identified as *Bacillus subtilis*. These taxa, *B. subtilis* subsp. *subtilis*, *B. subtilis* subsp. *spizizenii* and *Bacillus mojavensis*, are sympatric in some but not all habitats (Roberts & Cohan, 1995). Their identification as separate taxa resulted from our discovery that they form distinct sequence clusters for three protein-coding genes (Nakamura *et al.*, 1999; Roberts *et al.*, 1994). As with all bacterial sequence clusters that retain their distinctness in sympatry, these taxa are likely to represent ecologically distinct populations (Dykhuizen, 1998; Palys *et al.*, 1997). The geographical distributions of these taxa also suggest that these groups are ecologically distinct: although these taxa are capable of extremely high rates of dispersal (Roberts & Cohan, 1995), *B. subtilis* subsp. *spizizenii* is the only one of these three taxa to be isolated from Death Valley, CA, USA; also, *B. mojavensis* is the only one of these taxa to be found at high frequencies near our Hohat collection site in the Gobi Desert (Roberts & Cohan, 1995).

While these taxa appear to fill different ecological niches, the nature of the ecological differences is unknown. These taxa are not distinguishable by known differences in any metabolic capability at the level of presence versus absence. Here we compare these taxa, as well as the outgroup *Bacillus licheniformis*, for quantitative differences in the ability to utilize 95 carbon sources; we also analyse these taxa for growth rates in different salt concentrations. We will show that, while the three closely related taxa of our ingroup show no qualitative differences in resource utilization, they are quantitatively different in utilization of 8 % of the resources tested, and they are different in their levels of salt tolerance.

METHODS

Strains. The strains used in this study are listed in Table 1, and they have been described previously (Duncan *et al.*, 1994; Istock *et al.*, 1996; Roberts & Cohan, 1995).

Metabolic measurements. Biolog GP MicroPlates were used to determine the ability of strains to utilize 95 different carbon sources, including two alcohols, four amides, 11 amino acids, five aromatic compounds, 44 carbohydrates, 13 carboxylic acids, three esters, seven phosphorylated compounds and six polymers (Bochner, 1989). Cells were grown in Luria Broth for 12 h at 30 °C, at which point they were washed twice in saline (0·8 % NaCl). Then, 150 µl of the saline suspension was inoculated into each of the 96 wells, and the plates were incubated at 30 °C for 24 h. Each strain was assayed three times on separate dates. Analysis of variance (ANOVA) indicated that there were no date effects. The OD_{595} was determined at the time of inoculation and 24 h later using a Bio-Rad 3550 UV microtitre plate reader. The metabolic capability of each strain with each metabolite was quantified as ΔOD_{595}, the starting OD_{595} subtracted from the final OD_{595}.

We took into account the possibility that apparent differences among strains in their metabolic capabilities might be an artefact of strain differences in their adaptation to the base medium used in all Biolog tests. That is, if some strains were better adapted to Biolog's base medium, these strains might show greater change in OD_{595} on *all* metabolites. We therefore adjusted the raw metabolic capability score (ΔOD_{595}) for each strain (i) on each metabolite (j) by the strain's mean raw score over all metabolites $\overline{x_i}$:

$$\hat{x}_{ij} = x_{ij} - \overline{x_i},$$

where $\hat{x}_{ij}$ and x_{ij} are strain i's adjusted and raw ΔOD_{595} scores, respectively, on metabolite j. This adjusted score quantifies a strain's ability to metabolize a particular substrate relative to its ability to metabolize all of the substrates tested. We have reported analyses of adjusted metabolic data, but have also indicated wherever results based on raw and adjusted scores are qualitatively different.

Salinity tolerance measurements. We assessed salinity tolerance by measuring growth rates in Luria Broth adjusted to a final salt concentration of 1·0, 2·5, 5·0 and 10·0 %. In each well, 4×10^6 cells were added to 0·2 ml of growth medium. The optical density of cultures was read by using a Bio-Rad 3550 UV microtitre plate reader.

Statistical analyses. Analyses were performed using JMP 3.1.5 STATISTICAL DISCOVERY SOFTWARE (SAS Institute, 1996).

Table 1. Strains used in this study

All strains were collected from desert soil. Strains from Tucson were isolated by Conrad Istock; all other strains were isolated by F. M. Cohan.

Strain	Source	Strain	Source
B. subtilis* subsp. *subtilis		***B. subtilis* subsp. *spizizenii***	
RO-A-4	Rosamond, CA, USA	T-G-1-16	Tucson, AZ, USA
RO-J-2	Rosamond, CA, USA	T-T-1-23	Tucson, AZ, USA
RO-NN-1	Rosamond, CA, USA	TU-B-8	Nefta, Tunisia
RS-C-4	Rosamond, CA, USA	TU-D-6	Nefta, Tunisia
RS-D-2	Rosamond, CA, USA	TU-D-8	Nefta, Tunisia
T-89-14	Tucson, AZ, USA	TU-E-6	Nefta, Tunisia
T-89-18	Tucson, AZ, USA	TU-F-6	Nefta, Tunisia
T-89-48	Tucson, AZ, USA	***B. mojavensis***	
T-F-32	Tucson, AZ, USA	RO-B-2	Rosamond, CA, USA
T-G-2-5	Tucson, AZ, USA	RO-C-2	Rosamond, CA, USA
TU-C-7	Nefta, Tunisia	RO-H-1	Rosamond, CA, USA
B. subtilis* subsp. *spizizenii		RO-QQ-2	Rosamond, CA, USA
DV1-B-1	Death Valley, CA, USA	RS-A-2	Rosamond, CA, USA
DV2-A-2	Death Valley, CA, USA	T-G-2-42	Tucson, AZ, USA
DV3-E-3	Death Valley, CA, USA	T-G-3-41	Tucson, AZ, USA
DV4-D-5	Death Valley, CA, USA	***B. licheniformis***	
DV6-A-2-3	Death Valley, CA, USA	DV5-A-1-4	Death Valley, CA, USA
IM-C-216	Hohat, Inner Mongolia	DV5-A-2-2	Death Valley, CA, USA
RO-E-2	Rosamond, CA, USA	DV7-B-2	Death Valley, CA, USA
RO-G-4	Rosamond, CA, USA	DV8-3-1	Death Valley, CA, USA
RO-PP-2	Rosamond, CA, USA	Rr1	Tucson, AZ, USA
RO-SS-1	Rosamond, CA, USA	T-88-14	Tucson, AZ, USA
RS-B-1	Rosamond, CA, USA	T-89-27	Tucson, AZ, USA
T-88-10	Tucson, AZ, USA	T-89-36	Tucson, AZ, USA
T-88-8	Tucson, AZ, USA	T-89-37	Tucson, AZ, USA
T-89-3	Tucson, AZ, USA		

RESULTS

Metabolism of individual resources

Based on raw, unadjusted metabolic scores, *B. subtilis* subsp. *spizizenii* showed greater (but not necessarily significantly greater) metabolic capability on 79 of 95 metabolites than both *B. subtilis* subsp. *subtilis* and *B. mojavensis* (see Table 2 for data on 22 metabolites; other data not shown). This result suggested the possibility that *B. subtilis* subsp. *spizizenii* might simply be better adapted to Biolog base medium than the other taxa. This also raised the possibility that metabolic differences among other taxa and among strains might be artefacts of differences in level of adaptation to Biolog base medium. To control for differences in adaptation to the base medium, we adjusted each metabolic ΔOD_{595} score for a strain by its average performance on all metabolites (see Methods).

Metabolic measurements for each resource were analysed separately by ANOVA using a nested model: strains were nested within taxa, with taxon and strain both treated as random effects. Because 95 ANOVAs were performed, we reduced the significance values using the sequential Bonferroni adjustment (Rice, 1989). Most carbon sources (88/95) revealed significant differences among strains of the same taxon (data not shown). Using adjusted ΔOD_{595} scores, 12 of 95 carbon sources (13 %) showed significant variation among taxa; using raw ΔOD_{595} scores, 19 of 95 carbon sources (20 %) showed significant variation among taxa (Table 2). Nine of the 95 carbon sources (9 %) showed significant variation among taxa by *both* raw and adjusted analysis – this is our most conservative estimate for the fraction of metabolites for which the taxa are significantly different.

The nine carbon sources with inter-taxon differences were widely distributed across categories of metabolites: taxa were significantly different in metabolism of three of the 11 amino acids, two of the 44 carbohydrates, three of the 13 carboxylic acids and one of the seven phosphorylated compounds.

We next addressed whether the differences among taxa indicated by the ANOVA were due primarily to inclusion of the outgroup, *B. licheniformis*. Of the 22 carbon sources that demonstrated taxon-specific differences (by analysis of either raw or adjusted ΔOD_{595} scores), only two such

Table 2. Mean metabolic capability (ΔOD_{595}) of each taxon, for the 22 metabolites showing significant variation among taxa, based on either raw or adjusted ΔOD_{595} scores

The raw metabolic score is the upper number and the adjusted score is the lower number. The standard errors were calculated based on the variance among strains.

Substrate	*B. subtilis* subsp. *subtilis*	*B. subtilis* subsp. *spizizenii*	*B. mojavensis*	*B. licheniformis*
Amide				
Succinamic acid*	0·083 ± 0·039	0·111 ± 0·041	0·104 ± 0·042	0·137 ± 0·035
	−0·004 ± 0·008	−0·009 ± 0·006	−0·007 ± 0·01	0·180 ± 0·009
Amino acid				
D-Alanine*†	0·052 ± 0·024	0·066 ± 0·023	0·054 ± 0·019	0·112 ± 0·087
	0·035 ± 0·007	−0·055 ± 0·005	−0·057 ± 0·009	−0·008 ± 0·008
Glycyl-L-glutamic acid*†	0·063 ± 0·024	0·098 ± 0·027	0·078 ± 0·034	0·132 ± 0·040
	−0·025 ± 0·006	−0·023 ± 0·005	−0·033 ± 0·008	0·012 ± 0·007
L-Alanyl-glycine*†	0·062 ± 0·024	0·083 ± 0·021	0·067 ± 0·025	0·111 ± 0·049
	0·026 ± 0·005	−0·004 ± 0·004	−0·043 ± 0·006	−0·007 ± 0·006
L-Glutamic acid*	0·079 ± 0·029	0·117 ± 0·032	0·102 ± 0·033	0·142 ± 0·064
	−0·008 ± 0·005	−0·004 ± 0·004	−0·008 ± 0·007	0·220 ± 0·006
Carbohydrate				
Arbutin†	0·126 ± 0·017	0·209 ± 0·012	0·136 ± 0·022	0·143 ± 0·02
	0·039 ± 0·13	0·088 ± 0·009	0·026 ± 0·016	0·023 ± 0·014
D-Mannitol*†	0·112 ± 0·041	0·155 ± 0·054	0·178 ± 0·07	0·093 ± 0·022
	0·025 ± 0·008	0·034 ± 0·006	0·068 ± 0·01	−0·026 ± 0·009
D-Melibiose*	0·089 ± 0·036	0·127 ± 0·036	0·110 ± 0·035	0·12 ± 0·021
	0·001 ± 0·005	0·006 ± 0·003	−0·001 ± 0·006	0·000 ± 0·005
D-Ribose*	0·082 ± 0·031	0·130 ± 0·036	0·130 ± 0·48	0·139 ± 0·032
	−0·006 ± 0·006	0·009 ± 0·004	0·019 ± 0·007	0·020 ± 0·007
D-Sorbitol†	0·144 ± 0·012	0·193 ± 0·009	0·203 ± 0·016	0·149 ± 0·014
	0·057 ± 0·007	0·073 ± 0·005	0·093 ± 0·008	0·030 ± 0·007
Lactulose*†	0·038 ± 0·19	0·043 ± 0·019	0·033 ± 0·01	0·125 ± 0·069
	−0·050 ± 0·009	−0·077 ± 0·006	−0·076 ± 0·011	0·006 ± 0·009
N-Acetyl-D-glucosamine*	0·039 ± 0·024	0·067 ± 0·041	0·046 ± 0·025	0·093 ± 0·032
	−0·049 ± 0·007	−0·054 ± 0·005	−0·064 ± 0·009	−0·026 ± 0·008
N-Acetyl-D-mannosamine*	0·034 ± 0·022	0·062 ± 0·038	0·044 ± 0·023	0·091 ± 0·031
	−0·053 ± 0·006	−0·059 ± 0·005	−0·067 ± 0·008	−0·029 ± 0·007
Sedoheptulosan*	0·065 ± 0·029	0·090 ± 0·035	0·074 ± 0·027	0·105 ± 0·025
	−0·023 ± 0·007	−0·031 ± 0·005	−0·037 ± 0·009	−0·015 ± 0·008
Stachyose*	0·072 ± 0·034	0·094 ± 0·036	0·084 ± 0·032	0·115 ± 0·026
	−0·016 ± 0·007	−0·027 ± 0·005	−0·026 ± 0·009	−0·004 ± 0·008
Xylitol†	0·057 ± 0·025	0·078 ± 0·035	0·065 ± 0·026	0·045 ± 0·016
	−0·031 ± 0·006	−0·042 ± 0·004	−0·046 ± 0·007	−0·074 ± 0·006
Carboxylic acid				
D-Malic acid*	0·045 ± 0·019	0·071 ± 0·025	0·060 ± 0·023	0·054 ± 0·022
	−0·043 ± 0·006	−0·050 ± 0·004	−0·051 ± 0·007	−0·066 ± 0·006
α-Hydroxybutyric acid*	0·055 ± 0·024	0·090 ± 0·033	0·071 ± 0·021	0·075 ± 0·02
	−0·032 ± 0·006	−0·030 ± 0·004	−0·040 ± 0·008	−0·044 ± 0·007
L-Lactic acid*†	0·088 ± 0·043	0·139 ± 0·052	0·124 ± 0·064	0·067 ± 0·025
	0·000 ± 0·008	0·018 ± 0·006	0·014 ± 0·01	−0·053 ± 0·008
L-Malic acid*†	0·084 ± 0·035	0·120 ± 0·033	0·112 ± 0·045	0·145 ± 0·051
	−0·004 ± 0·004	−0·001 ± 0·003	0·002 ± 0·005	0·025 ± 0·004
Pyruvic acid*†	0·153 ± 0·065	0·230 ± 0·068	0·229 ± 0·083	0·173 ± 0·047
	0·065 ± 0·01	0·109 ± 0·007	0·118 ± 0·012	0·053 ± 0·011
Phosphorylated compound				
DL-α-Glycerol phosphate*†	0·033 ± 0·019	0·054 ± 0·026	0·047 ± 0·026	0·026 ± 0·016
	−0·055 ± 0·005	−0·067 ± 0·004	−0·064 ± 0·006	−0·094 ± 0·005

*Variance among taxa significant for raw ΔOD_{595} scores, after adjusting for 95 tests, using the sequential Bonferroni method.
†Variance among taxa significant for adjusted ΔOD_{595} scores, after adjusting for 95 tests, using the sequential Bonferroni method.

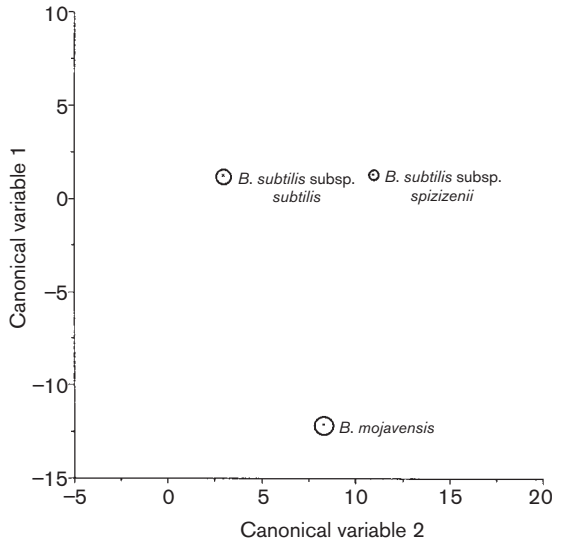

Fig. 1. Canonical analysis of variation among taxa in their ability to metabolize 95 substrates. The canonical variables are each a different linear combination of the 95 metabolic measurements. For each taxon, the point represents the mean over strains, and the circle represents the 95 % confidence interval for the location of the mean.

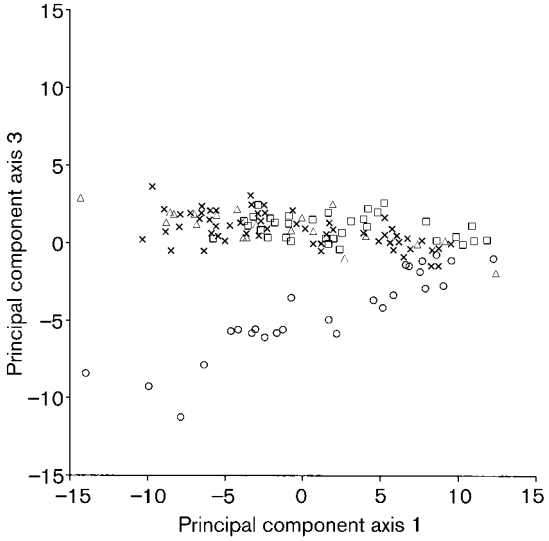

Fig. 2. Principal components analysis of metabolic differences among strains. The principal components are each a linear combination of the 95 metabolic measurements. Strains of *B. subtilis* subsp. *subtilis* (□), *B. subtilis* subsp. *spizizenii* (×), *B. mojavensis* (△) and *B. licheniformis* (○) are indicated. The first three principal components account for 46, 17 and 8 % of the variation, respectively. The coefficients of the first three principal components for each substrate are available as supplementary data from *Microbiology* Online (http://mic.sgmjournals.org).

differences (xylitol and glycyl-L-glutamate) could be attributed entirely to *B. licheniformis* being distinct from the three other taxa. The latter was included among metabolites yielding significant inter-taxon variation for both raw and adjusted ΔOD_{595} scores. Therefore, our conservative estimate for the fraction of metabolites with significant inter-taxon variation within our ingroup is 8 % (8/95).

Overall species differences in metabolism

Through canonical analysis (Dillon, 1984), differences among ingroup taxa in 95 metabolic measurements were projected onto a two-dimensional phenotype space. The canonical analysis shows *B. mojavensis* and the two subspecies of *B. subtilis* to be distinct in their average metabolic measurements (Fig. 1).

A multiple analysis of variance (MANOVA) of the pool of 95 metabolic characters showed the taxa to be significantly distinct in overall metabolic phenotype (Pillai's Trace = 2·69, $F = 16·11$, d.f. = 285, $P < 0·0001$). The MANOVA also shows striking variation among strains within each taxon (Pillai's Trace = 35·25, $F = 1·91$, d.f. = 4180, $P < 0·0001$). A principal components analysis (Fig. 2) shows that the inter-taxon variation leads to considerable overlap of phenotypes among species. Therefore, quantitative differences among taxa are not diagnostic, even while the modal values of the taxa are significantly different.

Salinity tolerance

As salinity increased, growth rates decreased (Table 3), and no strains grew at 10 % salinity. Taxon effects were significant, with *B. mojavensis* growing approximately 10 % faster than either *B. subtilis* subspecies at both 1 and 2·5 % NaCl. There were no significant differences at 5 % NaCl. The *B. subtilis* subspecies were not significantly different in their growth rates at any salinity.

DISCUSSION

Quantitative variation in diagnostic systematics

This study presents the first extensive survey, to our knowledge, of quantitative variation for metabolic capabilities within and between closely related bacterial taxa. The two subspecies of *B. subtilis* and the closely related *B. mojavensis*, as well as the outgroup *B. licheniformis*, all showed wide-ranging quantitative variation, both within and between taxa. Quantitative variation was significant among strains of the same taxon for 93 % of the 95 carbon sources tested, and the ingroup taxa were significantly different in their average rates of metabolism for 8 % of the carbon sources.

Nevertheless, these quantitatively varying metabolic characters do not contribute to diagnostic systematics for the *Bacillus* taxa studied here. Each of the metabolic traits investigated showed considerable overlap among taxa, even for traits whose mean values differed significantly among taxa (Fig. 2). Likewise, *qualitative* variation in metabolic

Table 3. Mean growth rates (doublings per hour) in Luria broth with different salt concentrations

Asterisks indicate those taxa that grew significantly faster or slower than others at a given salt concentration, as determined by the Tukey–Kramer test (*, $P<0{\cdot}05$; **, $P<0{\cdot}005$). Standard errors are based on the variance among strains within the taxon.

Salt concn	Taxon		
	B. subtilis subsp. *subtilis*	*B. subtilis* subsp. *spizizenii*	*B. mojavensis*
1 % NaCl	$1{\cdot}09 \pm 0{\cdot}04$	$1{\cdot}03 \pm 0{\cdot}14$	$1{\cdot}17 \pm 0{\cdot}04$*
2·5 % NaCl	$0{\cdot}95 \pm 0{\cdot}15$	$1{\cdot}01 \pm 0{\cdot}15$	$1{\cdot}16 \pm 0{\cdot}11$**
5 % NaCl	$0{\cdot}46 \pm 0{\cdot}12$	$0{\cdot}51 \pm 0{\cdot}02$	$0{\cdot}43 \pm 0{\cdot}03$

capabilities has also failed to provide diagnostic phenotypic characters for taxa of the *B. subtilis* complex. As yet, the only diagnostic phenotypic characters distinguishing *B. mojavensis* and the *B. subtilis* subspecies are the quantitatively varying levels of fatty acids (Nakamura *et al.*, 1999; Roberts *et al.*, 1994, 1996).

Gradual evolution versus abrupt evolution by horizontal transfer

Comparative genomics has led to increased interest in adaptation via abrupt evolutionary change effected by horizontal transfer, at the expense of interest in incremental change effected by mutations in existing genes (Lawrence, 1999, 2002; Ochman *et al.*, 2000). Lawrence (1999, 2002) has hypothesized that nearly all invasions of new niches, and perhaps much of adaptation within a niche, is effected by horizontal transfer. That bacteria have acquired heterologous genes by horizontal transfer is demonstrated by large differences in genomic content among close relatives, as assayed in three decades of DNA–DNA hybridization studies (Johnson, 1973; Lan & Reeves, 1996), and more recently by sequencing multiple genomes from the same species (Alm *et al.*, 1999; Baba *et al.*, 2002; Beres *et al.*, 2002; Perna *et al.*, 2001; Shirai *et al.*, 2000).

In the case of the *Bacillus* taxa we have investigated, only *B. subtilis* strain 168 has been fully sequenced (Kunst *et al.*, 1997), but DNA–DNA hybridization indicates considerable variation in genome content within and between named taxa (Nakamura *et al.*, 1999; Roberts *et al.*, 1994). Within each of the taxa studied here, pairs of strains are on average $6{\cdot}3 \pm 0{\cdot}5$ % different in the genes they contain (Nakamura *et al.*, 1999; Roberts *et al.*, 1994), and pairs of strains from different taxa differ on average by $37{\cdot}0 \pm 0{\cdot}4$ % of their genomes for the two subspecies of *B. subtilis*, by $59{\cdot}1 \pm 1{\cdot}2$ % for *B. mojavensis* and *B. subtilis* subsp. *subtilis*, and by $58{\cdot}5 \pm 1{\cdot}4$ % for *B. mojavensis* and *B. subtilis* subsp. *spizizenii* (Nakamura *et al.*, 1999; Roberts *et al.*, 1994). We do not know at this point what fraction of these genomic differences are responsible for adaptive differences, but there is clearly room for horizontal transfer to play a role in adaptive divergence.

Nevertheless, the present study suggests that acquisition of genes is not the whole story of phenotypic evolution in bacteria. A sizeable fraction of diverse metabolic traits were shown to vary quantitatively within (93 %) and between (8 %) ingroup taxa, so modest changes in the expression of many genes could have played a large part in adaptive evolution. We cannot be certain exactly how many genetic changes were responsible for variation in the metabolic characters studied, since a single genetic change can affect the expression of multiple genes (Cavalieri *et al.*, 2000; Hamilton, 2002). However, the observation that metabolic capability varied over so many substrates, and over several categories of substrates, suggests that quantitative metabolic changes have occurred frequently in *Bacillus* evolution. Even so, it is unknown whether the metabolic variants we have observed reflect adaptations in nature.

What is the likelihood that incremental change in metabolic capabilities has led to invasion of new ecological niches in bacterial evolution? Experiments in evolutionary microcosms have demonstrated that evolution by mutation alone can yield new, ecologically distinct populations, which can co-exist indefinitely with the parental population. For example, consider the case of the clonal descendants of one *E. coli* cell, evolving in medium with glucose as the only carbon source. Here mutational increases in acetate metabolism founded a new population that could survive by feeding on the acetate secreted as a waste product by the parental glucose-specialized population (Treves *et al.*, 1998). Mutational adjustment of a large number of metabolic capabilities could potentially found an entire ensemble of stably co-existing, ecologically distinct populations.

However, co-existence between populations that are only incrementally different in their metabolic capabilities is severely challenged by temporal variation in nutrient levels (Feldgarden *et al.*, 2003; Lunzer *et al.*, 2002). With fluctuating nutrient levels, one population will almost inevitably out-compete to extinction any other population that uses the same set of resources, albeit in different proportions. In contrast, acquiring a metabolic capability by horizontal transfer can give a population a novel resource not utilized by the parental population, thus preventing extinction by the parental population (Cohan, 2004; Holt, 1987). Populations distinguished by horizontal transfer events are much more likely to co-exist over ecological time.

Over evolutionary time, the co-existence of quantitatively different populations is also challenged by adaptive mutations appearing in each population. As long as two populations use entirely the same set of resources (but in different proportions), an adaptive mutation in one population threatens to out-compete the other to extinction (Cohan, 2004). However, a population that has diverged by acquiring a new metabolic operon instantaneously escapes this risk of extinction. Thus, while there are likely many more opportunities for ecological divergence to originate by mutation than by horizontal transfer, persistence of these populations is more likely when they have diverged through horizontal transfer.

Gradual evolution may have its most important role in improving the level of adaptation in a newly divergent, ecologically distinct population. Many new adaptations have somewhat deleterious side effects, and incremental change may play a role in compensating for novel adaptations that bring a lineage into a new niche (Bouma & Lenski, 1988; Cohan *et al.*, 1994; Schrag *et al.*, 1997). Furthermore, when a population enters a new niche, there may be new opportunities for adaptation by incremental changes. For example, when birds first evolved flight, this created natural selection to take advantage of new feeding opportunities (e.g. involving changes in behaviour and feeding morphology). Likewise, a bacterial population that can live in new microhabitats owing to a horizontally transferred gene may find its established resources available in different proportions. This would precipitate gradual evolution in metabolic capabilities.

Metabolic and genomic diversity within taxa

Each taxon has shown considerable quantitative variation in metabolic rates (Fig. 2) and salt tolerance (Table 3), as well as in genomic content (as revealed by DNA–DNA hybridization) (Nakamura *et al.*, 1999; Roberts *et al.*, 1994, 1996). What are the dynamics that allow such a high degree of variation within a bacterial species?

The level of genetic diversity within an ecologically homogeneous population of bacteria (i.e. an ecotype) is most likely regulated by periodic selection (Cohan, 1994a, 2002b, 2004; Koch, 1974; Levin, 1981). Owing to the extremely low rates of recombination within bacterial populations (Cohan, 2002a, c; Maynard Smith *et al.*, 1993), selection favouring an adaptive mutation is expected to bring the entire genome of the adaptive mutant cell to nearly 100 % frequency (Cohan, 1994b, 2004). This extinguishes variation at all loci within the genome, including coding and regulatory sequences, as well as variation in genomic content (i.e. the set of genes borne by a genome). If all members of a named taxon are actually members of the same ecotype (i.e. the strains are ecologically interchangeable), then the intra-taxon variation would represent only ecologically meaningless variation that appears randomly between periodic selection events, only to be extinguished with the next periodic selection.

Most bacterial species, however, are unlikely to represent a single ecotype. Most species investigated show considerable variation in protein-coding sequences, and this variation typically falls into multiple clusters of sequence similarity (Cohan, 2002b; Feil *et al.*, 1999, 2000), although there are some notable exceptions (*Yersinia pestis* and *Bacillus anthracis*) (Keim & Smith, 2002; Klevytska *et al.*, 2001), owing to systematists' practice of splitting lethal human pathogens (*nomen periculosum*) from extremely close relatives that are less harmful (Stackebrandt *et al.*, 2002). Any long-divergent clusters appearing within a species are unlikely to be members of the same ecotype, since the first periodic selection event would have collapsed the species diversity to a single cluster (Cohan, 2002b; Dykhuizen, 1998; Palys *et al.*, 1997). Moreover, the clusters observed within a named species are typically ecologically distinct (Achtman *et al.*, 2001; Chan *et al.*, 2001; Maiden *et al.*, 1998; Zhu *et al.*, 2001) and their phylogenies are consistent with single ecotypes (Cohan, 2002b).

Thus, a more likely interpretation of the great genomic and phenotypic diversity within a named species is that most of this diversity represents adaptive divergence *among* ecotypes. For example, variation within *B. mojavensis* in the ability to metabolize sorbitol may reflect differences between ecotypes that are adapted to environments with different sorbitol levels. Perhaps only a small fraction of the genomic and phenotypic diversity within a named species represents transient, random divergence within ecotypes.

We suggest a two-pronged approach to investigating the adaptive and non-adaptive origins of genomic and phenotypic variation within a named species. First, strains should be classified into putative ecotypes using one of several sequence-based approaches (Cohan, 2002b; Maiden *et al.*, 1998). The principle is that multilocus sequence clusters are likely to correspond to ecotypes, since periodic selection recurrently purges the diversity within but not between ecotypes. The putative ecotypes then establish a framework for mapping the metabolic and genomic diversity within a species.

Second, we should apply genomic technology to expand greatly the breadth of phenotypic surveys we can accomplish. Recently, microarray assays of whole-genome mRNA concentrations have shown adaptive differences between natural, conspecific populations of fish (Oleksiak *et al.*, 2002). While this approach has not yet been applied to natural populations of bacteria, the magnitude of variation found in the present study suggests that taxa may vary in the expression patterns of hundreds of genes.

We can then attempt to map phenotypic variation in gene expression and metabolism (and perhaps differences in genomic content as well) onto the putative ecotypes inferred from sequence clusters. Variation appearing within a putative ecotype is likely to represent merely the random changes occurring within a population between periodic selection events. However, any differences that correspond

to the putative ecotypes will demand further study. To the extent that differences in gene expression appear to determine niche differences, these differences will provide evidence that the putative ecotypes are really ecologically distinct and will suggest the nature of ecological differences among ecotypes.

ACKNOWLEDGEMENTS

This research was supported by National Science Foundation grant DEB-9815576 and by research funds from Wesleyan University.

REFERENCES

Achtman, M., van der Ende, A., Zhu, P. & 8 other authors (2001). Molecular epidemiology of serogroup a meningitis in Moscow, 1969 to 1997. *Emerg Infect Dis* **7**, 420–427.

Akashi, H. & Gojobori, T. (2002). Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **99**, 3695–3700.

Alm, R. A., Ling, L. S., Moir, D. T. & 20 other authors (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**, 176–180.

Baba, T., Takeuchi, F., Kuroda, M. & 11 other authors (2002). Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet* **359**, 1819–1827.

Baumler, A. J., Gilde, A. J., Tsolis, R. M., van der Velden, A. W., Ahmer, B. M. & Heffron, F. (1997). Contribution of horizontal gene transfer and deletion events to development of distinctive patterns of fimbrial operons during evolution of *Salmonella* serotypes. *J Bacteriol* **179**, 317–322.

Beres, S. B., Sylva, G. L., Barbian, K. D. & 13 other authors (2002). Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proc Natl Acad Sci U S A* **99**, 10078–10083.

Bochner, B. R. (1989). Sleuthing out bacterial identities. *Nature* **339**, 157–158.

Bouma, J. E. & Lenski, R. E. (1988). Evolution of a bacteria/plasmid association. *Nature* **335**, 351–352.

Cavalieri, D., Townsend, J. P. & Hartl, D. L. (2000). Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. *Proc Natl Acad Sci U S A* **97**, 12369–12374.

Chan, M. S., Maiden, M. C. & Spratt, B. G. (2001). Database-driven multi locus sequence typing (MLST) of bacterial pathogens. *Bioinformatics* **17**, 1077–1083.

Cohan, F. M. (1994a). The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *Am Nat* **143**, 965–986.

Cohan, F. M. (1994b). Genetic exchange and evolutionary divergence in prokaryotes. *Trends Ecol Evol* **9**, 175–180.

Cohan, F. M. (2002a). Clonal structure: an overview. In *Encyclopedia of Evolution*, pp. 159–161. Edited by M. Pagel. New York: Oxford University Press.

Cohan, F. M. (2002b). What are bacterial species? *Annu Rev Microbiol* **56**, 457–487.

Cohan, F. M. (2002c). Population structure and clonality of bacteria. In *Encyclopedia of Evolution*, pp. 161–163. Edited by M. Pagel. New York: Oxford University Press.

Cohan, F. M. (2004). Periodic selection and ecological diversity in bacteria. In *Selective Sweep*. Edited by D. Nurminsky. Georgetown, Texas: Landes Bioscience.

Cohan, F. M., King, E. C. & Zawadzki, P. (1994). Amelioration of the deleterious pleiotropic effects of an adaptive mutation in *Bacillus subtilis*. *Evolution* **48**, 81–95.

Cooper, T. F., Rozen, D. E. & Lenski, R. E. (2003). Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc Natl Acad Sci U S A* **100**, 1072–1077.

Dillon, W. R. M. G. (1984). *Multivariate Analysis: Methods and Applications*. Chichester: Wiley.

Duncan, K. E., Ferguson, N., Kimura, K., Zhou, X. & Istock, C. A. (1994). Fine-scale genetic and phenotypic structures in natural populations of *Bacillus subtilis* and *Bacillus licheniformis*: important implications for bacterial evolution and speciation. *Evolution* **48**, 2002–2025.

Dykhuizen, D. E. (1998). Santa Rosalia revisited: why are there so many species of bacteria? *Antonie Van Leeuwenhoek* **73**, 25–33.

Feil, E. J., Maiden, M. C., Achtman, M. & Spratt, B. G. (1999). The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Mol Biol Evol* **16**, 1496–1502.

Feil, E. J., Smith, J. M., Enright, M. C. & Spratt, B. G. (2000). Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics* **154**, 1439–1450.

Feldgarden, M., Brisson, D., Stoebel, D. M. & Dykhuizen, D. E. (2003). Size doesn't matter: microbial selection experiments address ecological phenomena. *Ecology* **84**, 1679–1687.

Ferea, T. L., Botstein, D., Brown, P. O. & Rosenzweig, R. F. (1999). Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc Natl Acad Sci U S A* **96**, 9721–9726.

Hall, B. G. (1999). Experimental evolution of Ebg enzyme provides clues about the evolution of catalysis and to evolutionary potential. *FEMS Microbiol Lett* **174**, 1–8.

Hall, B. G. & Malik, H. S. (1998). Determining the evolutionary potential of a gene. *Mol Biol Evol* **15**, 1055–1061.

Hamilton, B. A. (2002). Variations in abundance: genome-wide responses to genetic variation and vice versa. *Genome Biol* **3**, reviews1029.

Holt, R. D. (1987). On the relationship between niche overlap and competition: the effect of incommensurable niche dimensions. *Oikos* **48**, 110–114.

Istock, C. A., Bell, J. A., Ferguson, N. & Istock, N. L. (1996). Bacterial species and evolution: theoretical and practical perspectives. *J Ind Microbiol* **17**, 137–150.

Johnson, J. (1973). Use of nucleic-acid homologies in the taxonomy of anaerobic bacteria. *Int J Syst Bacteriol* **23**, 308–315.

Katz, L. & Burge, C. B. (2003). Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res* **13**, 2042–2051.

Keim, P. & Smith, K. L. (2002). *Bacillus anthracis* evolution and epidemiology. *Curr Top Microbiol Immunol* **271**, 21–32.

Klevytska, A. M., Price, L. B., Schupp, J. M., Worsham, P. L., Wong, J. & Keim, P. (2001). Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. *J Clin Microbiol* **39**, 3179–3185.

Koch, A. L. (1974). The pertinence of the periodic selection phenomenon to prokaryote evolution. *Genetics* **77**, 127–142.

Kunst, F., Ogasawara, N., Moszer, I. & 148 other authors (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–256.

Lan, R. & Reeves, P. R. (1996). Gene transfer is a major factor in bacterial evolution. *Mol Biol Evol* **13**, 47–55.

Lawrence, J. G. (1997). Selfish operons and speciation by gene transfer. *Trends Microbiol* **5**, 355–359.

Lawrence, J. G. (1999). Gene transfer, speciation, and the evolution of bacterial genomes. *Curr Opin Microbiol* **2**, 519–523.

Lawrence, J. G. (2002). Gene transfer in bacteria: speciation without species? *Theor Popul Biol* **61**, 449–460.

Levin, B. R. (1981). Periodic selection, infectious gene exchange and the genetic structure of *E. coli* populations. *Genetics* **99**, 1–23.

Lunzer, M., Natarajan, A., Dykhuizen, D. E. & Dean, A. M. (2002). Enzyme kinetics, substitutable resources and competition: from biochemistry to frequency-dependent selection in *lac*. *Genetics* **162**, 485–499.

Maiden, M. C., Bygraves, J. A., Feil, E. & 10 other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140–3145.

Maynard Smith, J. M., Dowson, C. G. & Spratt, B. G. (1991). Localized sex in bacteria. *Nature* **349**, 29–31.

Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proc Natl Acad Sci U S A* **90**, 4384–4388.

Nakamura, L. K. (1998). *Bacillus pseudomycoides* sp. nov. *Int J Syst Bacteriol* **48**, 1031–1035.

Nakamura, L. K., Roberts, M. S. & Cohan, F. M. (1999). Relationship of *Bacillus subtilis* clades associated with strains 168 and W23: a proposal for *Bacillus subtilis* subsp. *subtilis* subsp. nov. and *Bacillus subtilis* subsp. *spizizenii* subsp. nov. *Int J Syst Bacteriol* **49**, 1211–1215.

Ochman, H. & Groisman, E. A. (1996). Distribution of pathogenicity islands in *Salmonella* spp. *Infect Immun* **64**, 5410–5412.

Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.

Oleksiak, M. F., Churchill, G. A. & Crawford, D. L. (2002). Variation in gene expression within and among natural populations. *Nat Genet* **32**, 261–266.

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.

Perna, N. T., Plunkett, G., 3rd, Burland, V. & 25 other authors (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157 : H7. *Nature* **409**, 529–533.

Rainey, P. B. & Travisano, M. (1998). Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69–72.

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution* **43**, 223–225.

Roberts, M. S. & Cohan, F. M. (1995). Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution* **49**, 1081–1094.

Roberts, M. S., Nakamura, L. K. & Cohan, F. M. (1994). *Bacillus mojavensis* sp. nov., distinguishable from *Bacillus subtilis* by sexual isolation, divergence in DNA sequence, and differences in fatty acid composition. *Int J Syst Bacteriol* **44**, 256–264.

Roberts, M. S., Nakamura, L. K. & Cohan, F. M. (1996). *Bacillus vallismortis* sp. nov., a close relative of *Bacillus subtilis*, isolated from soil in Death Valley, California. *Int J Syst Bacteriol* **46**, 470–475.

Rozen, D. E. & Lenski, R. E. (2000). Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. *Am Nat* **155**, 24–35.

Schrag, S. J., Perrot, V. & Levin, B. R. (1997). Adaptation to the fitness costs of antibiotic resistance in *Escherichia coli*. *Proc R Soc Lond B Biol Sci* **264**, 1287–1291.

Seligmann, H. (2003). Cost-minimization of amino acid usage. *J Mol Evol* **56**, 151–161.

Shirai, M., Hirakawa, H., Kimoto, M. & 8 other authors (2000). Comparison of whole genome sequences of *Chlamydia pneumoniae* J138 from Japan and CWL029 from USA. *Nucleic Acids Res* **28**, 2311–2314.

Sokurenko, E. V., Chesnokova, V., Dykhuizen, D. E., Ofek, I., Wu, X. R., Krogfelt, K. A., Struve, C., Schembri, M. A. & Hasty, D. L. (1998). Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proc Natl Acad Sci U S A* **95**, 8922–8926.

Spiers, A. J., Kahn, S. G., Bohannon, J., Travisano, M. & Rainey, P. B. (2002). Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics* **161**, 33–46.

Stackebrandt, E., Frederiksen, W., Garrity, G. M. & 10 other authors (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**, 1043–1047.

Treves, D. S., Manning, S. & Adams, J. (1998). Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Mol Biol Evol* **15**, 789–797.

Zhu, P., van der Ende, A., Falush, D. & 13 other authors (2001). Fit genotypes and escape variants of subgroup III *Neisseria meningitidis* during three pandemics of epidemic meningitis. *Proc Natl Acad Sci U S A* **98**, 5234–5239.