

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Genomics of Bacteria from an Ancient Marine Origin: Clues to Survival in an Oligotrophic Environment

Luis David Alcaraz¹, Varinia López-Ramírez², Alejandra Moreno-Letelier³,
Luis Herrera-Estrella⁴, Valeria Souza⁵ and Gabriela Olmedo-Alvarez²

¹*Department of Genomics and Health,
Center for Advanced Research in Public Health, Valencia,*

²*Departamento de Ingeniería Genética, Cinvestav Unidad Irapuato,*

³*Division of Biology, Imperial College London, Silwood Park Campus, Ascot,*

⁴*Langebio, Cinvestav, Mexico*

⁵*Departamento de Ecología Evolutiva, Instituto de Ecología,
Universidad Nacional Autónoma de México*

¹*Spain*

^{2,4,5}*Mexico*

³*UK*

1. Introduction

Genomics has certainly changed the way that biology is studied, and has had a substantial impact on many other scientific disciplines as well. Life and Earth have had an interdependent history since the early establishment of the biogeochemical cycles in the Archean. Genomic sequences provide historical information that can be correlated with the geological record. Thus, it is not surprising that comparative genomics aids in understanding current findings in geological sciences. Genomics allows us to explain the evolutionary history of an organism by analyzing and comparing the set of shared genes with all respective relatives. Additionally, by examining the genes that are unique to some strains or taxonomic groups, we can make inferences about their ecology. Since molecular biology was initially developed in bacterial model organisms and we have extensive knowledge about the enzymes that participate in different biochemical pathways, inferences and functional predictions can be made about numerous sequenced genes. Moreover, due to the energetic and evolutionary costs of preserving a gene in bacteria, where specialists tend to have small size genomes, unique genes may aid in exploiting a given niche. Using comparative genomics, we have undertaken the study of the diversity, evolutionary relationships, and adaptations of bacteria to the oligotrophic (low nutrient content) conditions of the unique ecosystem of Cuatro Ciénegas, Coahuila, Mexico. In this chapter, we review how genomics can be used to describe and understand microbial diversity, evolution, and ecology in different environments as well as the relationship with geological data by analyzing current studies, and our own work as a case study (Fig. 1).

2. Bacterial comparative genomics: Clues to evolution, geological history, and ecology

Since the early 1980's, science has had technology to determine the precise nature of DNA, which is the molecule found in all genomes. The era of genomics began with the sequencing of the first bacterial genome, *Haemophilus influenzae*, in 1995. The sequencing of complete genomes has since gained speed and precision, and with the costs of this technology continually decreasing, the sequencing of genomes today that contain millions of base pairs of DNA is possible in just a few days. Over 1,800 bacterial genomes have been sequenced to date, with several hundred more currently in progress. More impressively, the very large genomes of several plants and animals, including the human genome with more than 3 billion bases of DNA, have also been completed. In order to understand how particular genomes give rise to different organisms with very different traits, the ability to compare and contrast features of this wide range of genomes is needed. Moreover, an understanding of the evolutionary history and ecology of the organisms for which we have genomes is required.

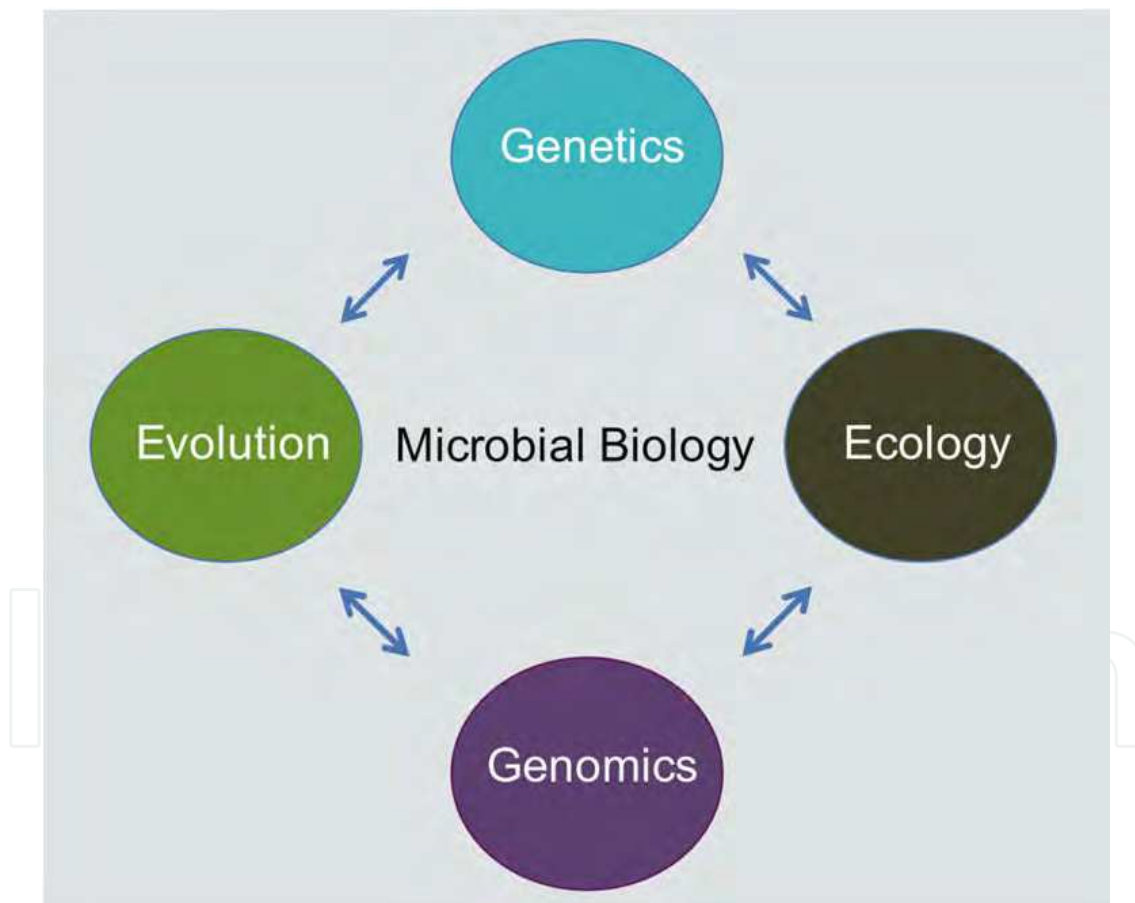


Fig. 1. Integration of Genetics, Genomics, Evolution, and Ecology. Genomics has had a great impact on Biology, particularly in the areas of Genetics, Ecology, and Evolution. Therefore, none of these areas can be studied separately. For microbial Biology, genomic studies have revealed a plethora of microbes that had eluded previous experimental strategies, which has provided the opportunity to understand their influence on the planet's history and present function in the environment.

2.1 How genomic sequences provide clues for understanding the diversity and origin of microorganisms

Classical microbiological classification systems made use of a culturable microbial diversity that only accounted for 1% of the total estimated microbial diversity (Amann et al. 1995; Khan et al. 2010), and due to the historical and clinical importance, it was a classification system that was highly biased towards clinical pathogenic strains. Although the 1% of culturable bacteria is an over-sold idea, it is clear that new culturing media should be developed and tested in order to capture more diversity within a given environment. To date, the most extensively used method for estimating the diversity of bacterial species is sequencing and comparing the 16S rRNA gene, which is an approach that has been used since the early 1970s (Woese & Fox 1977). Moreover, this method has been used in the current culture-free approaches, such as the metagenome surveys, by means of the Next Generation Sequencing (NGS) systems (Metzker 2010) in order to sequence DNA in particular environments (Fig. 2).

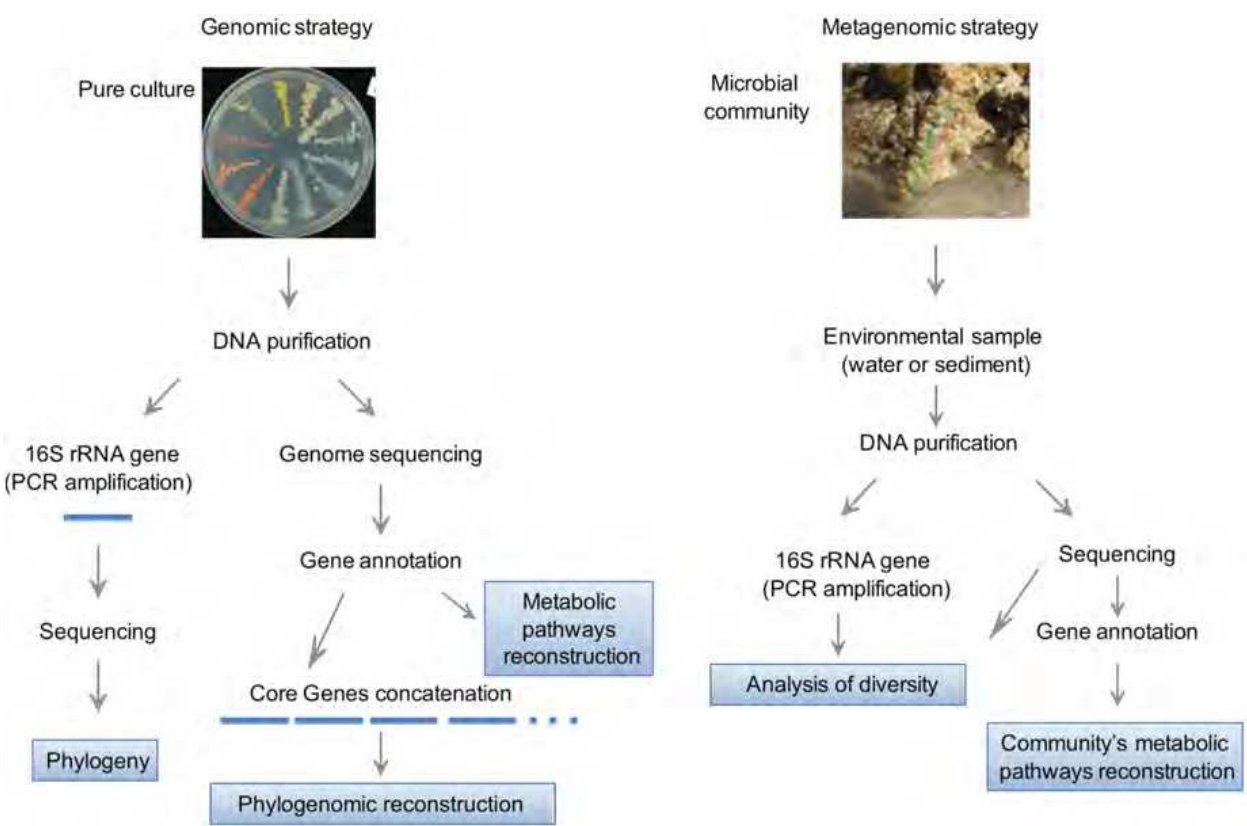


Fig. 2. Sequencing strategies for specific isolates and metagenomes.

The 16S rRNA gene sequencing approach has shed light onto the complex relationships amongst bacteria, and has even predicted the existence of new groups when using culture-free approaches. However, today we can go beyond explaining the evolution of a single gene and try to explore entire genome relationships of bacteria. With the development of NGS technologies, the associated costs are falling, and we can now envision a scenario where a small-sized lab is able to sequence the complete genome of a bacterial strain of interest. With whole genome sequencing, it is possible to perform a deep analysis to explain the relationships of every gene in genomes rather than single gene histories. Moreover, we

can also explain the history and presence of bacteria by understanding the set of genes shared by all relatives, which can reveal a common evolutionary lineage as well as the genes that are exclusive to particular strains. The later genes, which are unique to certain strains or taxonomic groups, can also give us insight on the capabilities of an organism to exploit a given niche and will therefore provide information on environmental features as well.

The numerous genome sequencing projects provide clues for elucidating whole genome relatedness and functions that are exclusive to some groups. On another level of complexity, global scale efforts are conducted to attempt to determine the roles of microbes in the global biogeochemical cycles as well as the discovery and inventory of gene diversity, such as that from the Global Ocean Expedition (Rusch et al. 2007). Other projects focus on a deep understanding of the relationships between our body and the microbes with which we live. This project, known as the Human Microbiome Project (Turnbaugh et al. 2007), attempts to shed light on the close relationship between humans and our microbes. Another challenging project has to do with the sequencing of at least one representative strain of all of the major bacteria groups, known as the Genomic Encyclopedia of Bacteria and Archaea (GEBA) (Nelson et al. 2010), and the ambitious Earth Microbiome Project (<http://www.earthmicrobiome.org/>) will provide a reference framework for conducting comparative genomics and establishing deep relationships amongst microbes.

Some research areas, such as population genetics, are transitioning from an understanding of a very limited number of loci to the analysis of whole genome single nucleotide polymorphisms (SNPs) in order to understand fine differences between the populations of pathogens and free-living bacteria. In contrast, other research areas, such as Molecular Biology, have the power to analyze whole genome regulation by sequencing the entire RNA of an organism that is exposed to a specific condition (transcriptomics) and using the results to predict the function of hypothetical genes. The improvement in the performance of computational resources as well as decreasing costs of these technologies has created tremendous analytical power. It is now possible to rent web services, such as cloud computing, and use highly efficient software online to assemble (Chaisson et al. 2004; Miller et al. 2008; Zerbino & Birney 2008), align (Kurtz et al. 2004), annotate (Aziz et al. 2008; Van Domselaar et al. 2005), and compare the more than one thousand bacterial genomes sequenced to date with in-house datasets. In addition, we are now able to democratically access the experimental and analytical tools to dissect both the evolution and functions of microbial life.

2.2 Core and accessory genomes and the pan-genome concept

In recent years, several groups have started to make whole genome comparisons of close relative strains, which has led to new and unexpected results for what was supposed to be members of the same bacterial species, based on 16S rRNA gene analysis, DNA-DNA re-association experiments, and phenotypic traits (Medini et al. 2005; Reno et al. 2009; Tettelin et al. 2005). Comparative genomic analysis has revealed that within species with the exact same 16S rRNA gene sequence, differences in gene content can account for $\pm 30\%$ of the total coding genes in one strain that are not found in their close relatives. To clearly understand this finding, we can make the comparison that if we look at this in percentages, humans are closer to chimpanzees (1% differences in coding genes) than two bacteria isolates causing the same disease.

From these observations, the genome of a species is now conceptually split into a core genome, which refers to all of the shared genes between analyzed strains, and the

accessory genome, which comprises the strain-specific genes. Core genes are those that are found in all genomes from a given taxonomic group and make up the genes that provide essential functions, such as those for DNA conservation and expression (involved in translation, replication, and transcription) as well as central metabolic pathways. These genes are collectively considered housekeeping genes. Core genes are considered likely components of all isolates of a given species or genus. Since selection at these loci exerts a stabilizing rather than a diversifying influence, variations detected in these genes tend to be neutral, or nearly so, and will accumulate in a consecutive, clock-like manner (see section 9). Thus, these genes should be reliable indicators of evolutionary history. For instance, core 16S rRNA gene sequences are frequently used to reconstruct the phylogenies of microbial species, and metabolic or housekeeping genes are commonly used for genotyping.

Accessory genes are those that may or may not be present in a given strain, since they can typically be associated with mobile elements (i. e. phages or transposons), can encode genes for secondary biochemical pathways, or can code for functions that mediate interactions with the environment. However, there are examples of genes that escape detection as core genes because they have been replaced by genes from a different origin (xenologous genes) that serve the same function but cannot be identified by similarity. This is often the case for divergent alleles. In any case, the genes classified as accessory genes are very interesting from an ecological point of view, since the genes in this category would be those that reveal the function of an organism in its environment. In addition, the range of genetic diversity of a species can be discovered in these genes. Pan-genomes, on the other hand, help us to understand the gene repertoire that has yet to be discovered within the group. The pan-genome is defined as the sum of the core genome and the accessory genes. Pan-genomes are also classified as open, where in an accumulative plot there is still no plateau, and closed, where all of the expected genes are present in the group and newly discovered genes are just a product of chance regardless of how many new strains are sequenced. Pan-genomes help us determine when other genome dynamics are shaping interesting phenotypes. For example, when comparing strains of the *Bacillus cereus-anthraxis* group (Anderson et al. 2005; Ehling-Schulz et al. 2005; Helgason et al. 2004), it was found that they were very clonal, and that their central pathogenic traits were the result of horizontal gene transfer in mobile elements.

A surprising finding in the many available microbial genomes was the high number of genes that had been acquired by exchange of DNA between microbes. The so-called “horizontal genetic transfer” events, confer a remarkable evolutionary potential to many species. The loss or gain of individual genes or large “genomic islands” accounts for the emergence of several specific metabolic, virulence, or drug resistance phenotypes. In fact, another characteristic of accessory genes is that these are often transferred among bacteria through the aid of phages, plasmids, or by transformation with free DNA. In a population, alleles (variant forms of a gene) may be distributed among the members, and individuals may possess either copy of the gene. These alleles (also called homeoalleles) encode enzymes with identical functions but that may have had a different evolution trajectory, and therefore exhibit differences at the gene level. Individual lineages may exhibit different homeoalleles acquired through horizontal gene transfer, but these homeoalleles may have also been lost from the lineage and replaced by other homeoalleles (reviewed in (Andam &

Gogarten 2011)). Importantly, horizontal gene transfer is responsible for the mosaicism often observed in genomes.

3. Using genes to understand the environment: A case study

3.1 The Cuatro Cienegas Basin: Water in a low phosphorous desert ecosystem

The Cuatro Cienegas Basin (CCB) is located in a valley in the central part of the State of Coahuila, Mexico (26°59'N 102°03'W). The Basin is roughly 84 km² in area and an average of 740 m above sea level (Fig. 3). CCB is surrounded by mountains that rise to elevations of approximately 2,500 and 3,000 m. The Geology and Physiography of the region have been thoroughly reviewed in the literature (Minckley 1969). Ancient geologic history indicates that the CCB was at the very nexus of the separation of Pangea, which created what we now know as the Northern hemisphere 220 million years ago. The CCB became isolated from the sea much later, and the subsequent uplifting of the Sierra Madre Oriental occurred approximately 35 million years. The major geological events of the Eocene epoch in northern Mexico corresponded to the genesis and development of the Sierra Madre Oriental, the fold-ranges of the Chihuahua-Coahuila and the Gulf coast plain development (Ferrusquía-Villafranca & González-Guzmán 2005).

Both climatically and geographically, the CCB belongs to the Chihuahuan Desert (Schmidt 1979), which is the second largest desert in North America. The climate is arid, with an average annual precipitation of less than 200 mm, and daytime temperatures in the summer that sometimes exceed 44°C. Despite the dry climate of the CCB, it harbors an extensive system of springs, streams, and pools. Within the valley, spring water flows on the surface and through subsurface channels in karstified alluvium. The main source for the subterranean water in these systems is old water that was deposited there in the late Pleistocene epoch (Wolaver 2008). Water in the CCB is thought to be the relict of a shallow sea that existed 35 million years ago. The oasis water contains low levels of NaCl and carbonates, but is rich in sulfates, magnesium, and calcium.

Vegetation and fauna in the CCB appear typical of an arid zone. Given the combined conditions of habitat diversity and permanence, as well as the isolation of the basin since historic times, elements of the aquatic fauna have undergone adaptive radiation and speciation that has resulted in many endemic organisms (reviewed in Holsinger & Minckley 1971; Minckley 1969).

The spring-fed ecosystems of the CCB are dominated by microbial mats and living stromatolitic features (see Fig. 3) that are supported by an aquatic sulfur cycle and a terrestrial gypsum-based ecology in large parts of the valley.

The most striking feature of the CCB ecosystem is the very low levels of phosphorus in both the water and soil, which presents an extreme elemental stoichiometry with regards to phosphorus (900:150:1- 15820:157:1 C:N:P ratio, respectively) (Elser 2005) when compared to similar environments. Phosphorus is an essential nutrient for multiple cellular processes, including energy and information. However, it is not an abundant element on the planet and can only be obtained from organic detritus or from tectonics and volcanism. Therefore, the availability of phosphorus is a limiting factor for all life forms. Nevertheless, life persevered and the CCB is characterized by a high endemism in all of the domains of life (Minckley 1969; Scanlan et al. 1993) despite the fact that phosphorus levels are below the level of detection (0.3 µM). In addition, the extremely oligotrophic waters are unable to sustain algal growth, which has caused the microbial mats to be the base of the food web (Elser 2005).

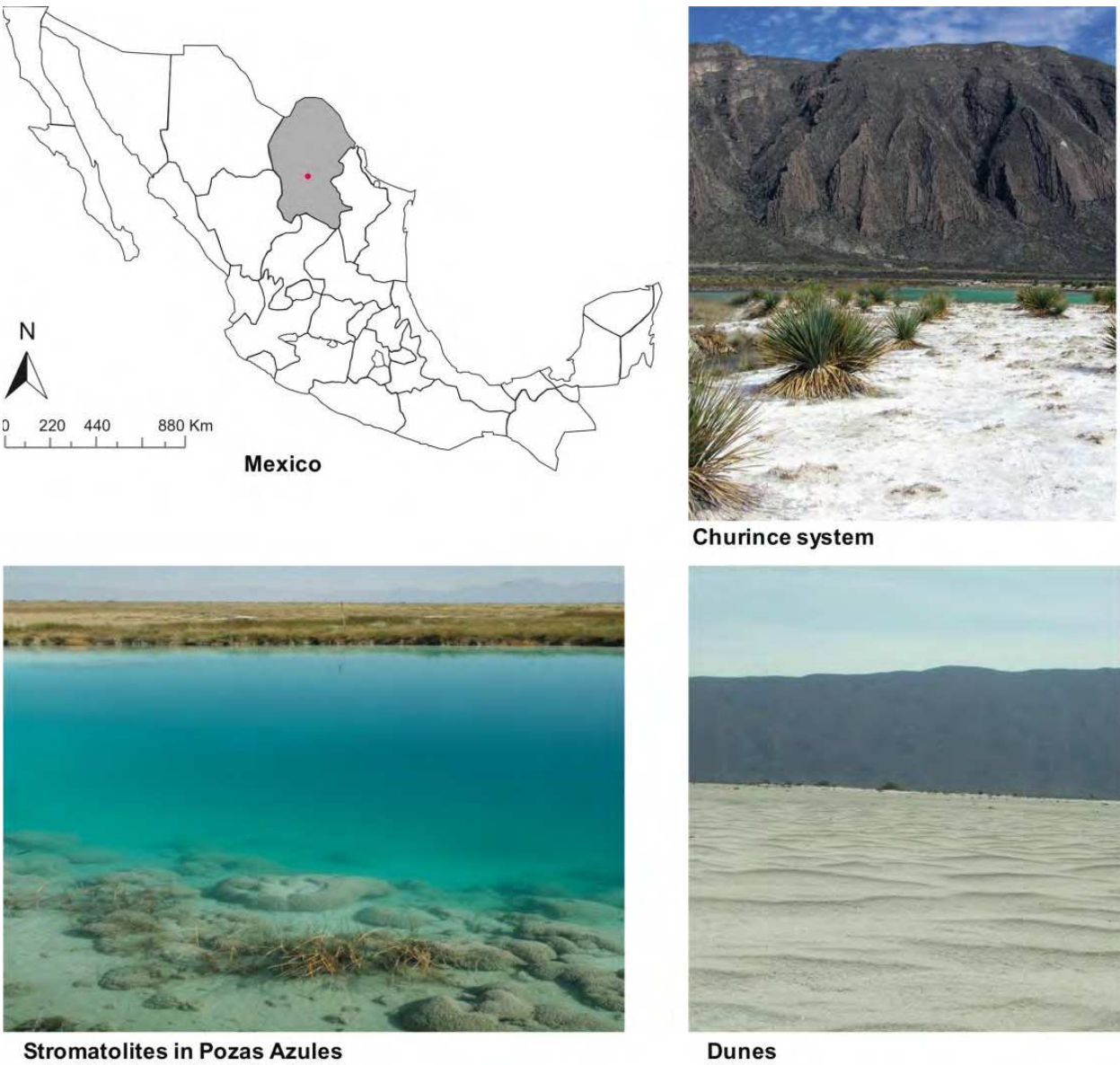


Fig. 3. The Cuatro Ciénegas Basin. (a) Location map of Cuatro Ciénegas, Coahuila. (b) Photograph of the Churince sytem facing San Marcos, one of the mountains in the valley. (c) One of the ponds at Pozas Azules, where live stromatolites can be observed. (d) White gypsum dunes, which are the result of evaporating water, once covered a vast area in the basin.

3.2 Bacterial genomics of two bacterial isolates from the CCB provide clues for the mechanism of bacterial survival in a highly oligotrophic environment

We have taken different approaches to gain knowledge on the diversity and evolution of microbial communities and to understand how microbes in these communities deal with the oligotrophic conditions in the various water bodies of Cuatro Ciénegas. Several microbiological surveys have been carried out, including bacterial isolation and culturing, 16S rRNA gene amplification, genome sequencing, and metagenome sequencing. Characterization of the microbial taxonomic diversity by sequence analysis of the 16S rRNA genes directly from environmental DNA has revealed that nearly half of the phylotypes

from the CCB are closely related to bacteria from marine environments (Souza et al. 2006). This makes sense, given that the geological history of Cuatro Ciénegas suggests that a shallow ocean covered the region and some of the water there is most likely a part of the underground water that feeds different ponds. This has led to the hypothesis that bacteria in the CCB water systems are descendants of marine bacteria from Pre-Cambrian (Souza et al. 2006). Systematic studies have also been used to describe new species of bacteria using current criteria (Anderson et al. 2005; Cerritos et al. 2008; Ehling-Schulz et al. 2005; Escalante et al. 2009; Helgason et al. 2004) as well as by attempting to apply multivariate analysis and physical-chemical analyses in order to determine special relationships amongst the bacteria (Cerritos et al. 2011). At the genomic scale, 2 genomes (Alcaraz et al. 2010; Alcaraz et al. 2008) and 3 metagenomes (Breitbart et al. 2009; Desnues et al. 2008; Dinsdale et al. 2008) had been described, in which genes that allow for adaptation to the oligotrophic environment of Cuatro Ciénegas have been identified. Two additional metagenomes from microbial mats and several more microbial genomes are currently under analysis. The entire dataset that is currently available from this area has helped us identify shared features across different microbial communities that deal with the same constraints, including phosphorous limitation, and has helped us understand the genetic strategies that are available to cope with these conditions.

However, taxonomy has only provided a single molecular marker approach. In order to gather more evidence on whether bacteria from the CCB could be recent migrants from marine environments or ancient creatures from an old sea, it required more than one gene as a tool. Therefore, we sequenced the whole genome of two isolates: *Bacillus coahuilensis* (Alcaraz et al. 2008; Cerritos et al. 2008) and *Bacillus* sp. isolate m3-13 (Alcaraz et al. 2010). On one hand, an analysis of the gene content and metabolic pathways represented by each genome revealed numerous differences (Table 1), starting with genome size, as *B. coahuilensis* turned out to have the smallest genome of all sequenced *Bacillus* spp. (3.65 Mbp) and an incomplete genome in many functions.

Another gene family overrepresented in the genomic datasets from Cuatro Ciénegas includes genes involved in environment sensing mechanisms. For example, all of the two-component (histidine-kinases) as well as the unusual sensitive rhodopsins, such as in the case of *B. coahuilensis*, were found to be overrepresented, indicating that most of the responses of the bacteria that have had their genomes analyzed are environmentally-triggered when compared to generalist and cosmopolitan organisms, such as *B. subtilis*. Moreover, these comparisons also showed an underrepresentation of secondary metabolism genes in the genome of *B. coahuilensis* (Alcaraz et al. 2010; Alcaraz et al. 2008; Desnues et al. 2008) (Fig. 4). Basic capabilities, such as sulfur utilization, were found to be different between the two isolates. *B. coahuilensis* appears not to be able to use inorganic sulfur and depends on organic sources for this element, while *Bacillus* m3-13 is able to use inorganic sulfur. In addition, the two isolates seem to be exposed to different stress, as *B. coahuilensis* has an alkyl hydroperoxide reductase protein coding gene that can aid with oxidative stress. Moreover, this strain also has several sugar transporters and biosynthesis pathways for sugars that help it deal with osmotic stress, such as trehalose, choline, and betaine uptake as well as betaine biosynthesis, all of which are absent in *Bacillus* m3-13.

In contrast, isolate m3-13 of *Bacillus* seemed to have a robust metabolism, starting with a complete urea cycle, capacity for taking up inorganic sulfur, and a wide selection of sugar transporters. Theoretically, the m3-13 strain would be able to use chitin, N-acetylglucosamine, maltose, maltodextrin, and sucrose to ferment lactate. It also has genes

Genomic features (differences)	<i>B. coahuilensis</i>	<i>Bacillus m3-13</i>
Genome size	3,640 Mbp	4,294 Mbp
Nitrogen metabolism	Lacks many genes, incomplete urea cycle	Complete urea cycle (Arginine and Ornithine Degradation)
Bacteriorhodopsin	Genes for bacteriorhodopsin and carotene synthesis	Lacks bacteriorhodopsin gene Pigmented, <i>crtB</i> y <i>crtI</i> (fitoene synthase and dehydrogenase)
Phosphate utilization genes	Sulfoquinovose and glycosyl transferase genes , sulfolipid synthesis	Phosphonate ABC transporter Genes for phosphonate utilization (C-P)
Phosphorous recycling	Alkaline phosphatase	Predicted ATPase related to phosphate starvation-inducible protein PhoH
Membrane Transport	Dipeptide-binding ABC transporter, periplasmic substrate-binding component	
Carbohydrates	Trehalose Uptake and Utilization	Chitin and N-acetylglucosamine Maltose and Maltodextrin Sucrose Fermentations: Lactate D-gluconate and ketogluconates metabolism D-ribose utilization Formaldehyde assimilation: Ribulose monophosphate pathway Beta-galactosidase
Cofactors, Vitamins, and Prosthetic Groups		Cobalamin, heme and siroheme, and thiamin biosynthesis
Sulfur metabolism	Sulfolipid synthesis	Inorganic Sulfur Assimilation: Adenylyl-sulfate reductase Ferredoxin--sulfite reductase; Sulfate permease, Pit-type
Stress Response	Osmotic stress: Choline and Betaine Uptake and Betaine Biosynthesis; Alkyl hydroperoxide reductase	
Virulence, Disease and Defense	Arsenical-resistance protein ACR3, Bacitracin export and resistance	Multidrug Resistance Efflux Pumps

Table 1. Genome differences between *B. coahuilensis* and *Bacillus m3-13*.

involved in D-gluconate and ketogluconates metabolism as well as D-ribose utilization. In addition, most of the amino acid biosynthesis pathways are complete. As expected in a place with extreme phosphorous limitation, some notable genetic features of those genomes are related to their ability to utilize phosphonate, which is discussed in the next section.

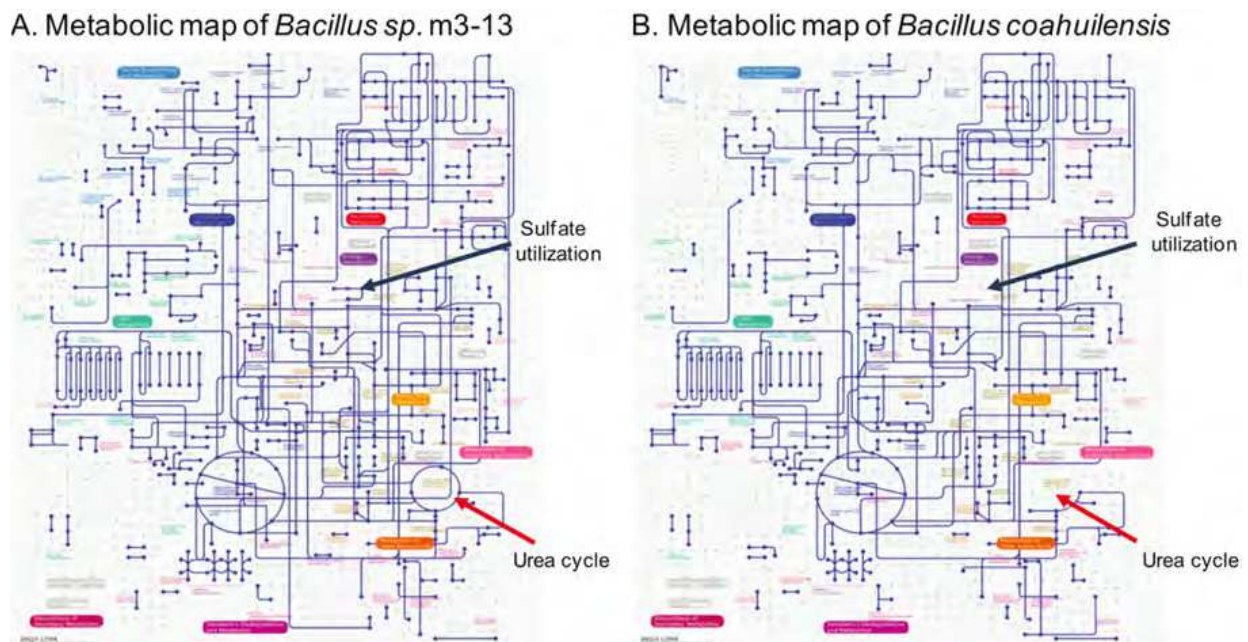


Fig. 4. Metabolic pathways in *Bacillus* m3-13 (A) and *B. coahuilensis* (B). A larger genome and more robust metabolism in m3-13 suggest that this is a generalist bacteria. In contrast, metabolism that lacks several reactions, such as those for inorganic sulfur utilization (blue arrow) and an incomplete urea cycle (red arrow), suggest that these bacteria depend on the community. The reduced genome of *B. coahuilensis* is puzzling for free-living bacteria. These maps were reconstructed from the annotated genome sequences using the KEGG (http://www.genome.ad.jp/kegg-bin/srch_orth/html) database.

4. Strategies of bacteria for dealing with limited phosphorous

Phosphorus is an essential nutrient for multiple processes, such as the synthesis of DNA, RNA, and ATP as well as many other pathways involving phosphorylation (Tetu et al. 2009). However, it is not an abundant element on the planet and can only be obtained from organic detritus or from tectonics and volcanism, and therefore the availability is a limiting factor for all life forms. Since the growth rate and primary productivity are highly dependent on phosphorus (Elser & Hamilton 2007; Elser et al. 2006; Zubkov et al. 2007), bacteria have different mechanisms for the uptake and storage of phosphates in order to cope with this limitation (Adams et al. 2008; Rusch et al. 2007; Tetu et al. 2009). For example, bacteria can use alternative phosphorus sources, use polyphosphates as storage compounds, or employ a highly effective phosphate-recycling mechanism (Fig. 5). Two major phosphate transport systems are involved: the low affinity phosphate inorganic transport (Pit) system, and the high affinity phosphate specific transport (Pst) system (van Veen, 1997). The uptake and assimilation of organic forms of phosphorus, such as phosphonates, require different transporters. Phosphonates are a class of organophosphorus compounds that are characterized by a chemically stable carbon-to-phosphorus (C-P) bond. Although

phosphonates are widespread, only microorganisms are able to cleave this bond. The use of alternative phosphorus sources (phosphonates, phosphites, and hypophosphites) was determined by the presence of the high affinity transporters *pnhD* and *ptxB* as well as the C-P lyase genes *pnhH* and *htxB* (White & Metcalf 2004). Some microorganisms have been shown to accumulate relatively large amounts of polyphosphate, which has been hypothesized to have an important role in the response to changes in nutritional status or environmental conditions. Polyphosphate acts as a reservoir of intracellular phosphate, which is a strategy that seems to be particularly important for mobility and the formation of biofilms (Brown & Kornberg 2004). Amongst the genes induced under phosphorus deprivation are *ppA*, *ppK*, and *ppX* (coding for pyrophosphatase, polyphosphatase kinase, and exopolyphosphatase, respectively), which are involved in polyphosphate metabolism. Finally, extracellular phosphates are recycled by the overexpression of alkaline phosphatases *phoA* and *phoX* (Scanlan et al. 1993). Some bacteria can also utilize phytic acid (inositol hexakisphosphate (IP6)), which is the principal storage form of phosphorus in some plant tissues.

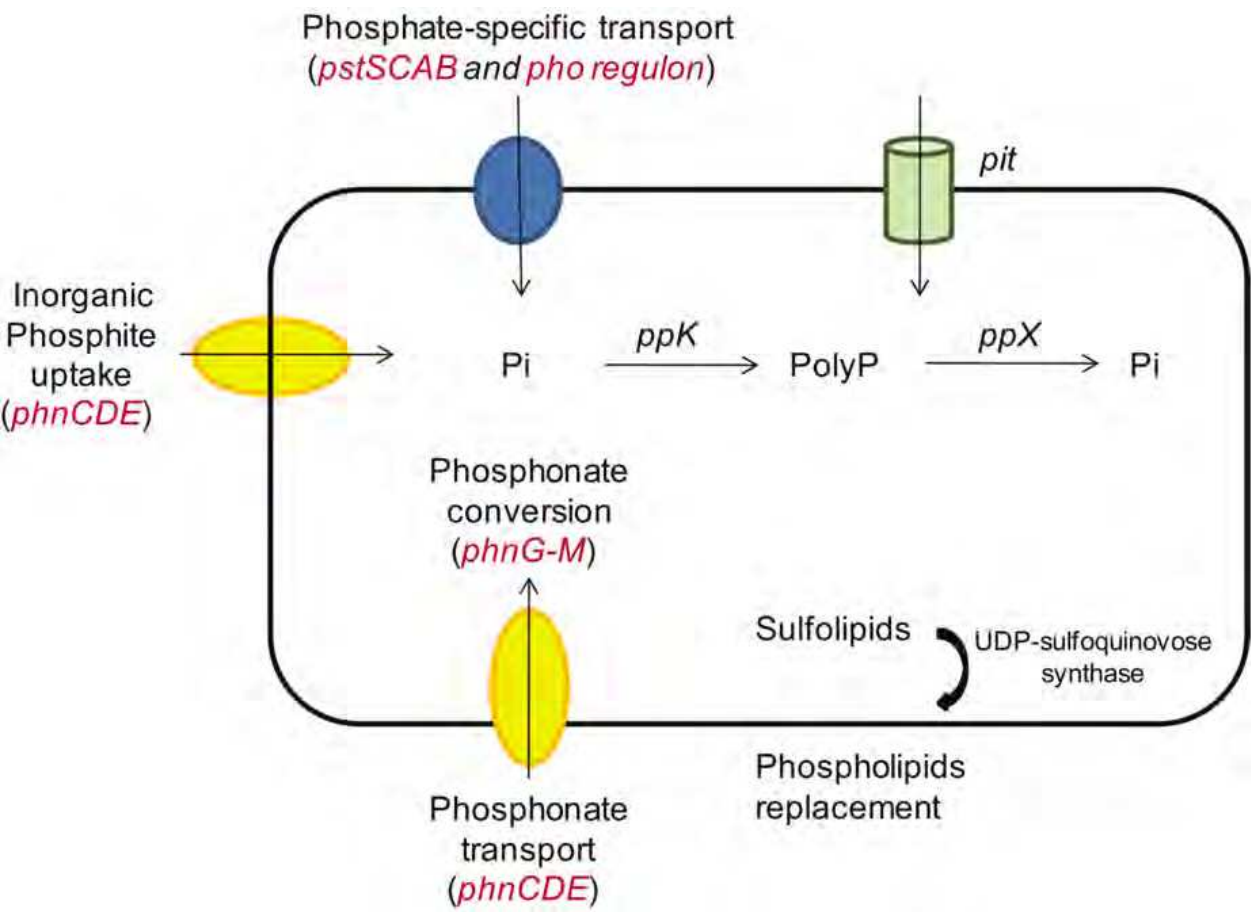


Fig. 5. Strategies for phosphorous utilization by bacteria (adapted from Hirota et al. 2010).

4.1 Phosphorous metabolism in the *Bacillus* genus

Since a low phosphorus concentration is a feature of the CCB basin, we focused on identifying the mechanisms of the *Bacillus* genus as a group, and in particular the isolates from the CCB for dealing with phosphorous. Comparative genomics can reveal the diversity

sulfoquinovose synthase (*sqd1*) and glycosyltransferase (*sqdX*), which are the two key enzymes in the synthesis of sulfolipids, are present in *B. coahuilensis*. Thin layer chromatography and mass spectrometry analysis have confirmed the presence of sulfolipids in *B. coahuilensis*.

The remarkable acquisition of constitutively expressed genes that allow *B. coahuilensis* to replace membrane phospholipids with sulfolipids is in agreement with genomic adaptations to extreme phosphate limitation. However, not all *Bacilli* from the CCB ponds use the same strategy to cope with phosphorous limitations. *B. coahuilensis* and its relatives as well as a few other *Bacilli* species have this gene, but we did not find sulfolipids among several other *Bacillus* spp. from the CCB. Therefore, how do other bacteria strains cope with the limiting phosphorous conditions? As explained above, genomic sequencing of another strain, *Bacillus* isolate m3-13, showed that it possessed *phn* genes that code for phosphonate ABC importers, permeases, and a phosphonate lyase. We hypothesize that this strain may use these genes to take up and assimilate phosphonates. Importantly, both strategies seem to be used by bacteria from the Cuatro Ciénegas as well as by marine bacteria (Fig. 6).

4.3 Mobile genes involved in phosphorous uptake

As discussed above, growth rate and primary productivity are highly dependent on phosphorus, and bacteria have different mechanisms for the uptake and storage of phosphates in order to cope with this limitation. The sulfoquinovose synthesis operon is absent in all other known *Bacillus* spp. genomes. The *B. coahuilensis* genes are closely related to cyanobacterial *sqd1* and *sqdX*, and the operon arrangement is identical to that in *Synechococcus* sp. PC7942, where these genes participate in the synthesis of sulfolipids (Benning 1998). This finding suggests that the adaptation of *B. coahuilensis* to the extremely low phosphorous concentration of the CCB may have included the acquisition of these genes through horizontal gene transfer (Alcaraz et al. 2008).

Another example of horizontal gene transfer involves genes of the *pho* regulon. The high affinity phosphate transport system (*pst*) is thought to be responsible for phosphate uptake under nutrient stress (Qi et al. 1997; Adams et al. 2008). *Pst* is a typical ABC transport system. Unlike the model bacteria *Escherichia coli* or *B. subtilis*, bacteria from the CCB as well as sequenced marine *Bacillus* lack the low affinity phosphate uptake system and must rely solely on the high affinity transport system. We found two types of operon architectures that harbor the *pst* gene and evaluated their phylogenetic congruity with housekeeping genes. We found high divergence of the two types of *pst*-operons in *Firmicutes* and incongruence with species phylogeny. In contrast to what was expected, the *pst* operon of marine *Bacillus* is not monophyletic, even though marine and the CCB *Bacilli* are resolved as a monophyletic group in the core-gene reconstruction. Therefore, the heterogeneous distribution of the different types of the *pst* operon among closely related species suggests horizontal gene transfer (Moreno-Letelier et al. 2011).

5. Carotenes and gene transfer of bacteriorhodopsin: A good combination in a high-radiation environment

B. coahuilensis has a gene encoding Bacteriorhodopsin, which is a situation similar to the abundance of BR genes in marine environmental samples, and suggests an additional adaptation of marine bacteria in the CCB. The phylogeny of *B. coahuilensis* sensory BR showed that its closest relative is the *Anabaena* sp. PCC7120 rhodopsin. Evidence for

horizontal gene transfer of rhodopsins has recently been obtained from whole genome sequencing and metagenomic projects, and is now thought to be a frequent event in marine bacteria in the photic zone and extreme saline environments. The retinal chromophore of rhodopsin is synthesized as a cleavage product of carotenoids; thus, the combination of carotenoid synthesis and rhodopsin genes has been suggested to be sufficient for rhodopsin function (Frigaard et al. 2006). The genome of *B. coahuilensis* also contains genes encoding *crtB* (phytoene synthase) and *crtICA2* (phytoene dehydrogenases) that could be involved in retinal biosynthesis. The high radiation exposure that is prevalent in shallow waters of the CCB could explain the selection pressure responsible for the maintenance and constitutive expression of the *bsr* gene.

6. Core and pan-genome of the *Bacillus* genus

In order to understand the cohesion of the *Bacillus* genus at the genomic level, we used the core and pan-genomes as the working units and took advantage of the large dataset available. We have analyzed the genes comprising the core genome of the *Bacillus* genus. The core genome of the *Bacillus* spp. that was analyzed (see phylogeny in Fig. 6) contained 814 genes. After annotating each gene in the genome of the two isolates, it was possible to classify and assign the different genes to specific functional categories and reconstruct and compare their metabolic pathways. Figure 7 shows the drawing of a metabolic map representation of the *Bacillus* pan-genome and *Bacillus* core genome obtained with the Kyoto Encyclopedia of Genes and Genomes. Using the pan-genome, we are able to understand the variation in functions across a cosmopolitan genus that can survive under harsh conditions, such as the bottom of the sea, hydrothermal vents, hypersaline environments, or even simply the hosts, as is the case for pathogens. The average gene content for the *Bacillus* genus was $4,973 \pm 923$ and the total pan-genome involved a around 75,000 genes clustered in 19,043 gene families. This is a very large number if we consider that the most recent predictions show that the human genome harbors 20,000 genes (Nelson et al. 2011). From these analyses, it is evident that a vast repertoire of functions is encoded in the *Bacillus* genus, and helps to explain the versatility of these bacteria for living and surviving in harsh conditions.

6.1 The sporulation core and accessory genes

The core genome sequence can help to identify and understand relevant, conserved genes for a trait of an entire group, and we therefore analyzed sporulation (Alcaraz et al. 2010), as the *Bacillus* is a group that is defined as endospore-forming genus. To obtain insight into the biology of this group, we described the relatedness within *Bacillus* using whole genome information to reconstruct their evolutionary history by taking advantage of the dataset available from the complete and draft genomes of 20 *Bacilli* isolated from a wide range of environments. Nucleotide metabolism, cell motility, and secretion showed little variability. In contrast, the features that were highly represented within the genomes and varied the most among the different genes were related to repair and transcription. Secondary metabolism, as expected, also exhibited variation among the taxonomic groups. Figure 7 shows a comparison between the functions defined by the core genome versus those defined by the pan-genome. The latter, which covers numerous metabolic pathways and reflects the potential of the whole group, explains why this is a cosmopolitan bacterial genus with the capability of colonizing diverse niches.

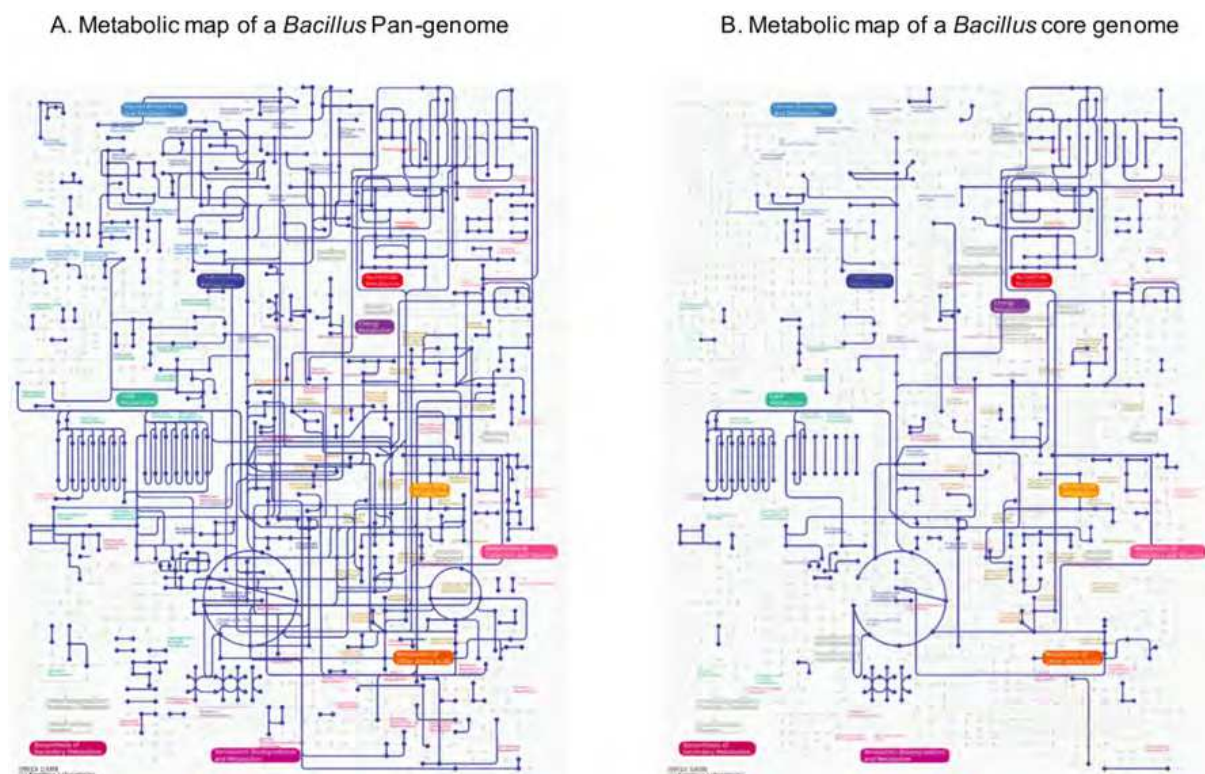


Fig. 7. Comparative metabolic reconstruction for the *Bacillus* pan-genome (A) and the *Bacillus* core genome (B). The combined pool of genes in the sequenced *Bacillus* genomes reflects the greater metabolic potential of the *Bacillus* genus. (B) The metabolic map of the core genome, which is represented by all of the conserved genes among the sequenced *Bacillus* genomes, reflects the basic housekeeping functions of these genera. The apparent absence of some enzymatic pathways that can be considered essential may be explained by the replacement of genes coding these capacities with xenologous genes. These maps were reconstructed from the annotated genome sequences using the KEGG (http://www.genome.ad.jp/kegg-bin/srch_orth/html) database.

When analyzing the core genome for genes involved in sporulation, we found that less than 52 of the 200 genes were conserved across 20 other *Bacillus* genomes. These genes are known to be essential for completion of the sporulation process in *B. subtilis*, but the fact that only a quarter of them are present in the other strains suggests the adaptability of genes for the same process, and particularly for the signaling circuit that responds to diverse environmental cues. Our study also identified the variable genes involved in general metabolism and allowed for the clustering of genes that made sense of their evolution and ecology.

The “accessory” sporulation genes are strain-specific and reflect the genomic flexibility of the group that allows *Bacillus* to colonize different environments that require different sensors to trigger the developmental response of the bacteria. This comparative strategy allows for the identification of the variable genes involved in the process and allows for the clustering of groups on the basis of their evolution and ecology.

7. Genomics in the aid of phylogeny

The use of the whole core genome to reconstruct the group’s phylogeny helped us understand the evolutionary relationships (Alcaraz et al. 2010) (Fig. 6). As stated above, the

core genome is thought to be faithful to the evolutionary history of the analyzed strains, and thus many individual genes' phylogenies are dissected to determine which can be used to define a natural group. Some other studies have analyzed what genes are necessary for a strain to be considered part of the same genus. For example, the cosmopolitan groups are either pathogenic, free-living, or extremophiles, such as *Pseudomonas* (Sarkar & Guttman 2004) and *Bacillus* (Alcaraz et al. 2010). Although the 16S rRNA gene has been traditionally used to draw a phylogeny, the resolution is often lost when working within a genus. Many genes can be concatenated and used to build a phylogeny. In fact, all genes shared by different genomes can be used to reconstruct a phylogeny; in the case of *Bacillus*, we used the 814 genes that constitute the core genome to build a *Bacillus* phylogeny. Most of these genes code for the expected housekeeping genes involved in basic cellular functions, as explained above, which fall into the traditional markers for phylogeny. However, we also included hypothetical conserved genes as well as nontraditional categories, such as genes involved in metabolism and transport mechanisms. The use of the whole core genome to reconstruct a group's phylogeny resolves the evolutionary relationships using the most available information. However, there are drawbacks of a low number of representatives of a given genus and the constant release of new genomes, which will eventually help fill in the gaps in the phylogenetic diversity and will help to understand which genes define a group.

8. Molecular clock

Traditionally, it has been difficult to set absolute dates for diversification of lineage events throughout Earth's history. Before DNA sequencing became widely available, the ages of divergence of the major groups of organisms depended solely on fossil information. Since the fossil record is incomplete, the history of lineages that lacked a rich fossil record, such as bacteria, could not be reconstructed (Kuo & Ochman 2009). When the first amino acid sequences of proteins were analyzed, they seemed to change in a rather constant fashion, which led to the hypothesis that a molecular clock existed in protein evolution. Therefore, molecular dating is based on the assumption that if most variation is neutral, then mutations will become fixed in a lineage at a constant rate that is equal to the mutation rate (Bromham & Penny 2003). One direct consequence of the evolution of a sequence at a rate that is relatively constant is that the genetic difference between any two species is proportional to the time that passed since the species last shared a common ancestor.

If the molecular clock hypothesis holds true, it serves as an extremely useful method for estimating evolutionary timescales. This is of particular value when studying organisms such as bacteria, which have left few traces of their biological history in the fossil record. Phylogenetic trees can be reconstructed that shed light onto the Earth's past. A calibration point is needed to obtain an absolute estimate of the age of a clade, and this can be set by using a known date of divergence, such as fossil information and samples from historical sources as well as paleontological, geological, atmospheric, and climatologic records. Based on that calibration point, the absolute date of a divergence can be estimated based on the number of mutations or substitutions (k) per total length of the sequence (n) per unit of time (Fig. 8), and can then be extrapolated to other parts of the phylogeny (Li 1997). This entire methodology assumes that the rates are equal in all branches of a phylogeny, which is not always the case (Ayala 1999). To deal with rate heterogeneity, several methods have been developed in the past few years that make molecular dating more accurate (Drummond & Rambaut 2007).

Geological evidence suggests that the CCB was a shallow marine environment for most of Earth's history, and when CCB became isolated due to the uplift of the Coahuila block, the bacteria that remained became relics of the diversity from the ancient sea. Thus, the high diversity of CCB would be a product of two things: the diversity already present at the time of isolation, and the new community assemblages that arose in this unique aquatic environment (Moreno-Letelier et al. 2011; Souza et al. 2006). *Firmicutes*, an abundant and widespread genera within CCB (Alcaraz et al. 2010; Alcaraz et al. 2008; Cerritos et al. 2011; Moreno-Letelier et al. 2011; Souza et al. 2006) are the focal group of our studies, in part because there are several endemic species within the site, and because the whole genome has been sequenced for both *Exiguobacterium* and *Bacillus* isolates (*Bacillus coahuilensis* m4-4 (Alcaraz et al. 2008); *Bacillus* sp. m3-13 (Alcaraz et al. 2010); and several unpublished data (from drafts *Bacillus* sp. p15.4, *Exiguobacterium* EPVM and 11-28). We have reconstructed the phylogenetic relationships and estimated the divergence times of aerobic *Firmicutes* from CCB and other similar habitats, in order to determine whether the diversity of the basin is a product of a recent adaptive radiation or the bacteria that remained became relics of the diversity we observe today. For our phylogenetic studies, calibration points were obtained from geologic events. The maximum age of the tree can not exceed the estimate of the origin of life on Earth, which is estimated to be approximately 4000 million years ago (Nisbet & Sleep 2001). The node of aerobic *Firmicutes* is set to have an age of 2300 Ma, the date of the Great Oxidation Event (Battistuzzi et al. 2004; Papineau 2010). CCB bacteria and their sister species were also constrained to have a minimum age of 35 Ma in the case of CCB bacteria and marine bacteria. This age corresponds to the final retreat of the Western Interior Seaway and the uplift of the Sierra Madre Oriental that finally isolated Cuatro Ciénegas from the ocean (Ferrusquía-Villafranca & González-Guzmán 2005).

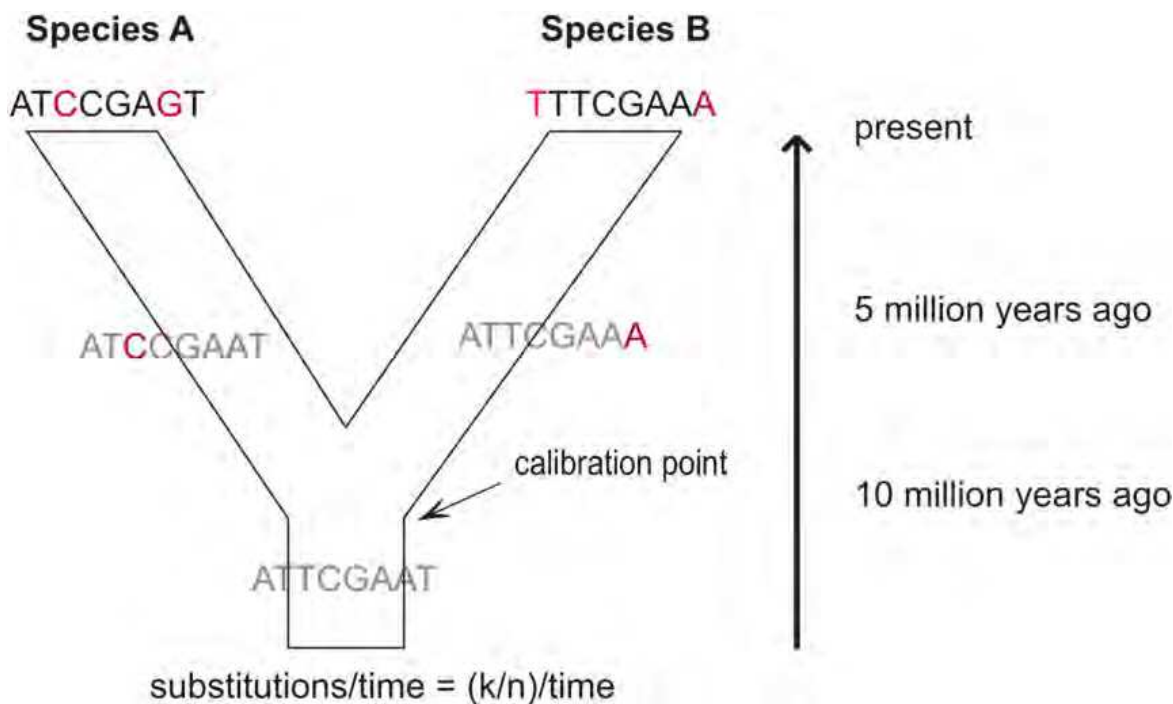


Fig. 8. The concept of the molecular clock based on DNA sequences. A representation of a set of sequences that evolves in a molecular clock fashion, where each branch contains 2 substitutions in 10 million years. For each species $k = 2$ and $n = 8$, which equals 0.025 substitutions per million years.

Our results showed that the diversity of *Firmicutes* in Cuatro Cienegas is not the product of a recent adaptive radiation. We suggest instead that ancient lineages became isolated and continued their own evolutionary path. This is especially true for the representatives of *Bacillus*, where the different taxa have diverged from their sister species at different times (Moreno-Letelier et al, under review). Overall, the Cuatro Cienegas Basin reinforces its geological importance since it is a refuge of ancient bacterial lineages, with traits that are relicts of old times, in some cases even from Precambrian times, and others that have arisen in order to survive in this endangered, extreme, and unique environment.

Questions	Experimental strategies
Have we already uncovered all of the microbial diversity in water, sediments, or mats of any pond?	16S rRNA gene libraries, <i>taxon</i> -specific libraries, metagenomes
Why is there so much microbial diversity? Does genetic diversity correlate with environmental stability or resource availability, or is it a matter of chance and history?	Explore the reasons for diversity: competition, mesocosm experiments in combination with 16S rRNA gene libraries, <i>taxon</i> -specific libraries, metagenomes
In most ponds, microbial mats are in charge of biogeochemical cycles. Therefore, what are the most important functions and are the functions similar in mats from different ponds?	Transcriptome analysis
Organisms from an ancient sea have adapted to the oligotrophic conditions of the ponds. How did they evolve and adapt? How was their genome shaped? What is the role of mobile elements?	Study transposons, plasmids, and phages in genomes and metagenomes
Do communities with extensive genetic diversity also have more functional diversity?	Comparative metagenomics and transcriptomics
Do communities respond differently to environmental change?	Mesocosms, transcriptomics

Table 2. Remaining questions on the CCB microbial diversity and evolution, and the genomic approaches to address them.

9. Future directions

The current approaches that combine genomics, transcriptomics, metagenomics, and proteomics together with classical microbiology will continue to contribute to our understanding of microbial activity and strategies for cell survival and growth of bacteria in oligotrophic ecosystems. Additional explorations of the diversity of the bacteria in these ponds using taxon specific oligonucleotides will allow us to determine whether we’ve actually reached an understanding of diversity with more general approaches, or if some taxa were missed due to the bias of DNA isolation techniques and PCR amplification of 16S rRNA genes with universal primers. Once we confidently identify the members of a community, a future challenge will be to understand the rules of assembly for the organisms within it, including the role of history, competition (for nutrients or antagonic), and

migration. In particular, transcriptomics will help us go beyond the identification of the genetic potential of the organisms afforded by the metagenomic approach and will allow us to actually look at the expression of such genes. The presence of specific mRNAs will be the actual reflection of what the microorganisms are sensing from the environment and how they are responding to them. This will be particularly fruitful once we have full genome information for many of the CCB bacterial isolates, since we will be able to assign mRNAs to specific taxons in the communities.

10. Conclusion

The application of genomic approaches to living systems uncovers the genetic bases of functional variation in nature. The revolution in high-throughput DNA sequencing and gene expression technologies has redefined the notion of a 'model' organism. The interrogation of genomes from animals, plants, microbes, or communities of organisms can identify genetic markers of processes at any scale: ecological, physiological, developmental, transcriptional, and others. Challenges lie ahead in the full interpretation of these datasets as well as understanding the connections between the gene information and the metadata of each particular environment. This will require interdisciplinary work between ecologists, microbial geneticists, biogeochemists, and computational biologists.

11. Acknowledgments

Julio Cruz for the Dune's photograph; Eugenio Reynoso for graphic art. G. Olmedo's laboratory team for their comments to the manuscript. This work was supported by a Multidisciplinario project from Cinvestav to GOA and Fondos Mixtos-Conacyt.

12. References

- Adams, M. M., Gomez-Garcia, M. R., Grossman, A. R. & Bhaya, D. (2008). Phosphorus deprivation responses and phosphonate utilization in a thermophilic *Synechococcus* sp. from microbial mats. *J Bacteriol* 190(24): 8171-8184.
- Alcaraz, L. D., Moreno-Hagelsieb, G., Eguiarte, L. E., Souza, V., Herrera-Estrella, L., et al. (2010). Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics* 11: 332.
- Alcaraz, L. D., Olmedo, G., Bonilla, G., Cerritos, R., Hernandez, G., et al. (2008). The genome of *Bacillus coahuilensis* reveals adaptations essential for survival in the relic of an ancient marine environment. *Proc Natl Acad Sci U S A* 105(15): 5803-5808.
- Amann, R. I., Ludwig, W. & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59(1): 143-169.
- Andam, C. P. & Gogarten, J. P. (2011). Biased gene transfer in microbial evolution. *Nat Rev Microbiol* 9(7): 543-555.
- Anderson, I., Sorokin, A., Kapatral, V., Reznik, G., Bhattacharya, A., et al. (2005). Comparative genome analysis of *Bacillus cereus* group genomes with *Bacillus subtilis*. *FEMS Microbiol Lett* 250(2): 175-184.
- Ayala, F. J. (1999). Molecular clock mirages. *Bioessays* 21(1): 71-75.

- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9: 75.
- Battistuzzi, F. U., Feijao, A. & Hedges, S. B. (2004). A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol* 4: 44.
- Benning, C. (1998). Biosynthesis and Function of the Sulfolipid Sulfoquinovosyl Diacylglycerol. *Annu Rev Plant Physiol Plant Mol Biol* 49: 53-75.
- Breitbart, M., Hoare, A., Nitti, A., Siefert, J., Haynes, M., et al. (2009). Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Ciénegas, Mexico. *Environ Microbiol* 11(1): 16-34.
- Bromham, L. & Penny, D. (2003). The modern molecular clock. *Nat Rev Genet* 4(3): 216-224.
- Brown, M. R. & Kornberg, A. (2004). Inorganic polyphosphate in the origin and survival of species. *Proc Natl Acad Sci U S A* 101(46): 16085-16087.
- Cerritos, R., Eguiarte, L. E., Avitia, M., Siefert, J., Travisano, M., et al. (2011). Diversity of culturable thermo-resistant aquatic bacteria along an environmental gradient in Cuatro Ciénegas, Coahuila, Mexico. *Antonie Van Leeuwenhoek* 99(2): 303-318.
- Cerritos, R., Vinuesa, P., Eguiarte, L. E., Herrera-Estrella, L., Alcaraz-Peraza, L. D., et al. (2008). *Bacillus coahuilensis* sp. nov., a moderately halophilic species from a desiccation lagoon in the Cuatro Ciénegas Valley in Coahuila, Mexico. *Int J Syst Evol Microbiol* 58(Pt 4): 919-923.
- Chaisson, M., Pevzner, P. & Tang, H. (2004). Fragment assembly with short reads. *Bioinformatics* 20(13): 2067-2074.
- Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., et al. (2008). Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* 452(7185): 340-343.
- Dinsdale, E. A., Edwards, R. A., Hall, D., Angly, F., Breitbart, M., et al. (2008). Functional metagenomic profiling of nine biomes. *Nature* 452(7187): 629-632.
- Dormann, P. & Benning, C. (2002). Galactolipids rule in seed plants. *Trends Plant Sci* 7(3): 112-118.
- Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214.
- Ehling-Schulz, M., Svensson, B., Guinebretiere, M. H., Lindback, T., Andersson, M., et al. (2005). Emetic toxin formation of *Bacillus cereus* is restricted to a single evolutionary lineage of closely related strains. *Microbiology* 151(Pt 1): 183-197.
- Elser, J. J. (2005). Effects of phosphorus enrichment and grazing snails on modern stromatolitic microbial communities. *Freshwater Biol* 50: 1808-1825.
- Elser, J. J. & Hamilton, A. (2007). Stoichiometry and the new biology: the future is now. *PLoS Biol* 5(7): e181.
- Elser, J. J., Watts, J., Schampel, J. H. & Farmer, J. (2006). Early Cambrian food webs on a trophic knife-edge? A hypothesis and preliminary data from a modern stromatolite-based ecosystem. *Ecol Lett* 9(3): 295-303.
- Escalante, A. E., Caballero-Mellado, J., Martinez-Aguilar, L., Rodriguez-Verdugo, A., Gonzalez-Gonzalez, A., et al. (2009). *Pseudomonas cuatrociénegasensis* sp. nov., isolated from an evaporating lagoon in the Cuatro Ciénegas valley in Coahuila, Mexico. *Int J Syst Evol Microbiol* 59(Pt 6): 1416-1420.

- Ferrusquía-Villafranca, I. & González-Guzmán, L. (2005). Northern Mexico's Landscape, Part 2, The biotic setting across space and time. *Biodiversity, Ecosystems and Conservation in Northern Mexico*. J. L. Cartron, G. C.-G. R. F. New York, Oxford University Press. 2: 40-51.
- Frigaard, N. U., Martinez, A., Mincer, T. J. & DeLong, E. F. (2006). Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* 439(7078): 847-850.
- Helgason, E., Tourasse, N. J., Meisal, R., Caugant, D. A. & Kolsto, A. B. (2004). Multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. *Appl Environ Microbiol* 70(1): 191-201.
- Hirota, R., Kuroda, A., Kato, J. & Ohtake, H. (2010). Bacterial phosphate metabolism and its application to phosphorus recovery and industrial bioprocesses. *J Biosci Bioeng* 109(5): 423-432.
- Holsinger, J. R. & Minckley, W. L. (1971). A new genus and two new species of subterranean amphipod crustaceans (Gammaridae) from northern Mexico. *Proceedings of the Biological Society of Washington* 83: 25-44.
- Khan, M. M., Pyle, B. H. & Camper, A. K. (2010). Specific and rapid enumeration of viable but nonculturable and viable-culturable gram-negative bacteria by using flow cytometry. *Appl Environ Microbiol* 76(15): 5088-5096.
- Kuo, C. H. & Ochman, H. (2009). Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biol Direct* 4: 35.
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol* 5(2): R12.
- Li, W. H. (1997). *Molecular Evolution* Sunderland, Massachusetts, Sinauer Associates.
- Medini, D., Donati, C., Tettelin, H., Massignani, V. & Rappuoli, R. (2005). The microbial pan-genome. *Curr Opin Genet Dev* 15(6): 589-594.
- Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet* 11(1): 31-46.
- Miller, J. R., Delcher, A. L., Koren, S., Venter, E., Walenz, B. P., et al. (2008). Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24(24): 2818-2824.
- Minckley, W. L. (1969). *Environments of the Bolson of Cuatro Ciénegas, Coahuila, Mexico, with Special Reference to the Aquatic Biota*. El Paso, TX, Texas Western Press,.
- Moreno-Letelier, A., Olmedo, G., Eguiarte, L. E., Martinez-Castilla, L. & Souza, V. (2011). Parallel Evolution and Horizontal Gene Transfer of the *pst* Operon in Firmicutes from Oligotrophic Environments. *Int J Evol Biol* 2011: 781642.
- Nelson, K. E., Mullany, P., Warburton, P. & Allan, E. (2011). *Metagenomics of the Human Body*. K. E. Nelson, E. New York, NY: . 165-173-.
- Nelson, K. E., Weinstock, G. M., Highlander, S. K., Worley, K. C., Creasy, H. H., et al. (2010). A catalog of reference genomes from the human microbiome. *Science* 328(5981): 994-999.
- Nisbet, E. G. & Sleep, N. H. (2001). The habitat and nature of early life. *Nature* 409(6823): 1083-1091.
- Papineau, D. (2010). Global biogeochemical changes at both ends of the proterozoic: insights from phosphorites. *Astrobiology* 10(2): 165-181.

- Qi, Y., Kobayashi, Y. & Hulett, F. M. (1997). The *pst* operon of *Bacillus subtilis* has a phosphate-regulated promoter and is involved in phosphate transport but not in regulation of the *pho* regulon. *J Bacteriol* 179(8): 2534-2539.
- Reno, M. L., Held, N. L., Fields, C. J., Burke, P. V. & Whitaker, R. J. (2009). Biogeography of the *Sulfolobus islandicus* pan-genome. *Proc Natl Acad Sci U S A* 106(21): 8605-8610.
- Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., et al. (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* 5(3): e77.
- Sarkar, S. F. & Guttman, D. S. (2004). Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Appl Environ Microbiol* 70(4): 1999-2012.
- Scanlan, D. J., Mann, N. H. & Carr, N. G. (1993). The response of the picoplanktonic marine cyanobacterium *Synechococcus* species WH7803 to phosphate starvation involves a protein homologous to the periplasmic phosphate-binding protein of *Escherichia coli*. *Mol Microbiol* 10(1): 181-191.
- Schmidt, R. H. J. (1979). A climatic delineation of the 'real' Chihuahuan Desert. *J. Arid Environ* 2: 243-250.
- Souza, V., Espinosa-Asuar, L., Escalante, A. E., Eguiarte, L. E., Farmer, J., et al. (2006). An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert. *Proc Natl Acad Sci U S A* 103(17): 6565-6570.
- Tettelin, H., Maignani, V., Cieslewicz, M. J., Donati, C., Medini, D., 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(39): 13950-13955.
- Tetu, S. G., Brahmsha, B., Johnson, D. A., Tai, V., Phillippy, K., et al. (2009). Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. *ISME J* 3(7): 835-849.
- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., et al. (2007). The human microbiome project. *Nature* 449(7164): 804-810.
- Van Domselaar, G. H., Stothard, P., Shrivastava, S., Cruz, J. A., Guo, A., et al. (2005). BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Res* 33(Web Server issue): W455-459.
- White, A. K. & Metcalf, W. W. (2004). Two C-P lyase operons in *Pseudomonas stutzeri* and their roles in the oxidation of phosphonates, phosphite, and hypophosphite. *J Bacteriol* 186(14): 4730-4739.
- Woese, C. R. & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74(11): 5088-5090.
- Wolaver, B., Sharp, JM, Rodríguez JM. and Ibarra JC (2008). Delineation of Regional Arid Karstic Aquifers: An Integrative Data Approach. *Ground Water* 46(3): 396-413.
- Zerbino, D. R. & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5): 821-829.
- Zubkov, M. V., Mary, I., Woodward, E. M. S., Warwick, P. E., Fuchs, B. M., et al. (2007). Microbial control of phosphate in the nutrient depleted North Atlantic subtropical gyre. *Environmental Microbiology* 9(8): 2079-2089.



Earth and Environmental Sciences

Edited by Dr. Imran Ahmad Dar

ISBN 978-953-307-468-9

Hard cover, 630 pages

Publisher InTech

Published online 07, December, 2011

Published in print edition December, 2011

We are increasingly faced with environmental problems and required to make important decisions. In many cases an understanding of one or more geologic processes is essential to finding the appropriate solution. Earth and Environmental Sciences are by their very nature a dynamic field in which new issues continue to arise and old ones often evolve. The principal aim of this book is to present the reader with a broad overview of Earth and Environmental Sciences. Hopefully, this recent research will provide the reader with a useful foundation for discussing and evaluating specific environmental issues, as well as for developing ideas for problem solving. The book has been divided into nine sections; Geology, Geochemistry, Seismology, Hydrology, Hydrogeology, Mineralogy, Soil, Remote Sensing and Environmental Sciences.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Luis David Alcaraz, Varinia López-Ramírez, Alejandra Moreno-Letelier, Luis Herrera-Estrella, Valeria Souza and Gabriela Olmedo-Alvarez (2011). Genomics of Bacteria from an Ancient Marine Origin: Clues to Survival in an Oligotrophic Environment, Earth and Environmental Sciences, Dr. Imran Ahmad Dar (Ed.), ISBN: 978-953-307-468-9, InTech, Available from: <http://www.intechopen.com/books/earth-and-environmental-sciences/genomics-of-bacteria-from-an-ancient-marine-origin-clues-to-survival-in-an-oligotrophic-environment>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen