Chapter 4
Environmental Systems Biology Approach to Bioremediation

Terry C. Hazen

Abstract Pollution is everywhere. Microbes are also everywhere, and many have the ability to degrade environmental contaminants. Understanding how these microbial communities work to degrade environmental contaminants will enable us to use these microbes to clean up the pollution. Understanding, monitoring, and controlling the environment with biological processes, i.e., an environmental systems biology approach to bioremediation, answer the need which is everywhere. By using an environmental systems approach to bioremediation, we make sure we know of any "fatal flaws" in the approach, get a much better handle on life-cycle cost analysis, and can grade an engineered solution into a natural attenuation solution. The whole is greater than the sum of its parts. By using an environmental systems biology approach to bioremediation and cross-linkage of systems at all levels providing multiple lines of evidence involving environmental observations, laboratory testing, microcosm simulations, hypothesis refinement, field testing and validation, and multiple iterations of this circle, we will be able to make new theories and paradigms for bioremediation of contaminated environments.

4.1 Introduction

Pollution is everywhere (Fig. 4.1). Microbes are also everywhere, and many have the ability to degrade environmental contaminants. Understanding how these microbial communities work to degrade environmental contaminants will enable us to use these microbes to clean up the pollution. Understanding, monitoring, and controlling the environment with biological processes, i.e., an environmental systems biology approach to bioremediation, answer the need which is everywhere. By using an environmental systems approach to bioremediation, we make sure we know of any "fatal flaws" in the approach and get a much better handle on life-cycle cost analysis.

There are three million parts in a Boeing 777 aircraft provided by more than 900 suppliers from 17 countries around the world. Completed genomes provide

T. C. Hazen (✉)
University of Tennessee and Oak Ridge National Laboratory, Knoxville, TN, USA
e-mail: tchazen@utk.edu; hazentc@ornl.gov

C. J. Hurst (ed.), *Understanding Terrestrial Microbial Communities*, Advances in Environmental Microbiology 6, https://doi.org/10.1007/978-3-030-10777-2_4

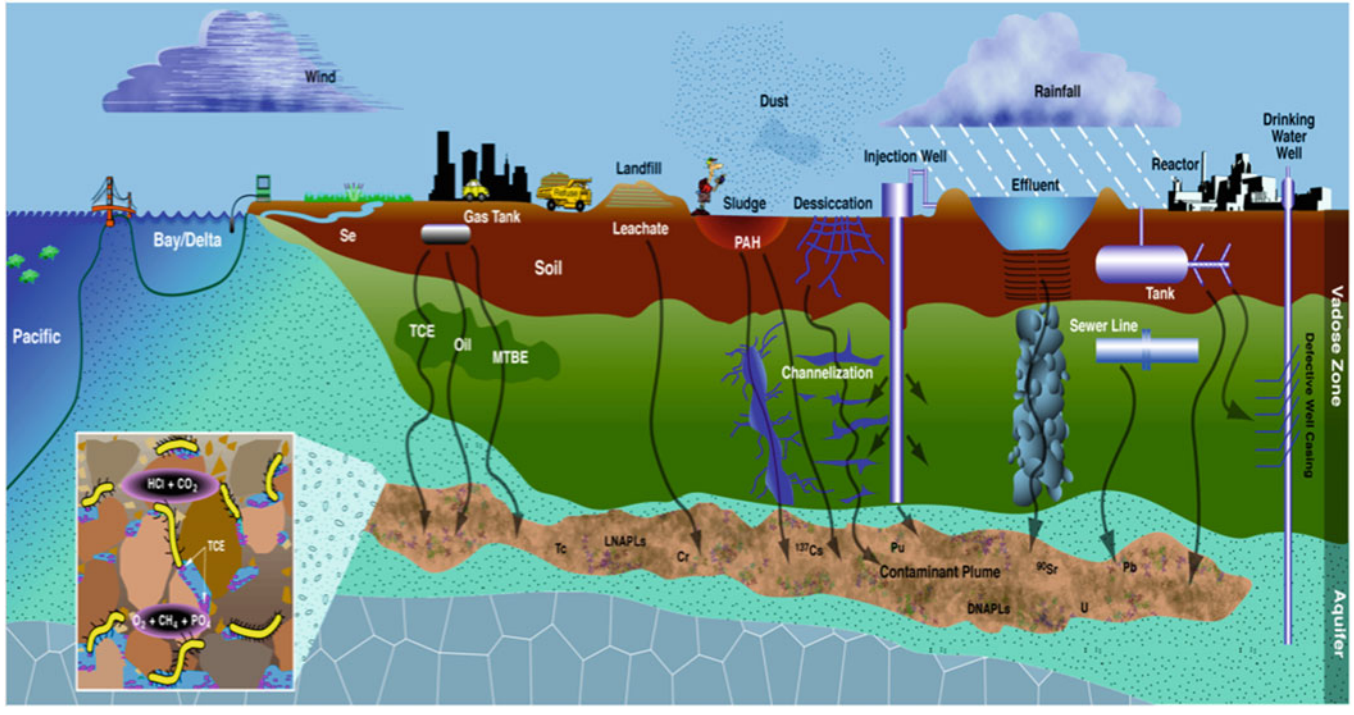

Fig. 4.1 Environmental systems biology (selenium (Se), trichloroethylene (TCE), methyl tert-butyl ether (MTBE), technetium (Tc), light nonaqueous phase liquids (LNAPL), polyaromatic hydrocarbons (PAH), chromium (Cr), cesium (Ce), plutonium (Pu), strontium (Sr), lead (Pb), uranium (U), dense nonaqueous phase liquids (DNAPL))

"parts lists" for many microbes, although the genomic sequence is little more than the blueprint for each part (protein) in the organism. Having a blueprint for the parts of a 777 jet gives few clues as to how each part is made, how it assembles into devices and systems, and much less how it flies which is its essence. For an organism the parts list (genome sequence) does not tell us even the shape of the parts or their function and at a higher level the principles that govern the ability of this large monolith to "fly." The genome sequence is really only a blueprint for each part. The order in which they are made and how that synthesis is controlled is in the parts list, but there is no instruction manual. Genes, proteins, metabolites, and multimolecular assemblies ("molecular machines") interact in an intricate labyrinth of pathways and networks to create, sustain, and reproduce the system we call the living cell—complexity well beyond the engineering and essence of a 777. Systems biology will transform biology from an empirical and descriptive science to a more quantitative and predictive science enabling us to manipulate and use living systems and their components.

Using an environmental systems biology approach to bioremediation requires that we consider interactions at different levels, e.g., the ecosystem (de Lorenzo et al. 2016), community, population, cell, genomics, transcriptomics, glycomics (Kay et al. 2010), lipidomics, fluxomics, proteomics, metabolomics, and phenomics (Hazen and Sayler 2016) (Fig. 4.2). This type of full-spectrum perspective is necessary for successfully using an environmental systems biology approach to bioremediation, because, like the 777 jet, the "whole is greater than the sum of its parts." Engineering success requires that all of the parts function together, and so we must take measurements at various levels in the systems being studied and then build models that interact at those various levels in the systems approach (Fig. 4.3). Systems biology can be a powerful approach to explain environmental phenomenon. However, this methodology intrinsically relies on multiple methods, each of which carries its own advantages and limitations. In order to accurately describe the system in question while not over-interpreting the data produced by these methods, a careful

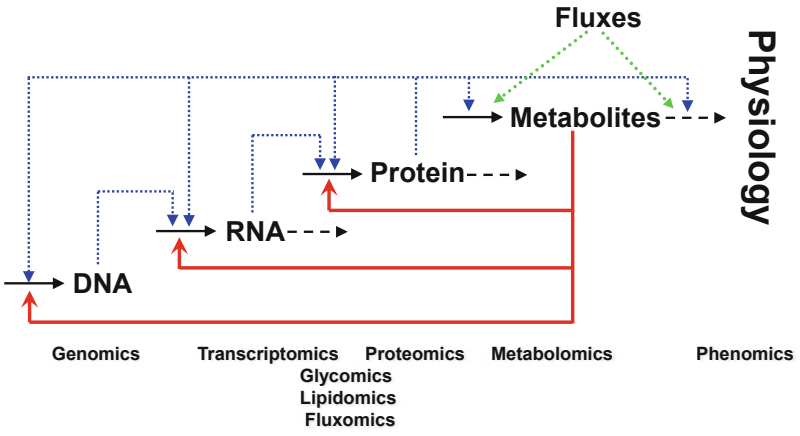

Fig. 4.2 The Omics!

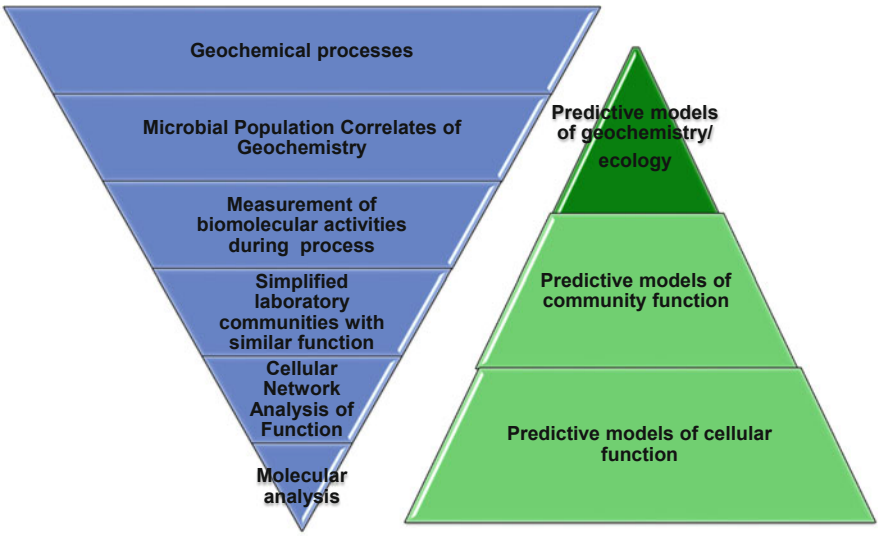

Fig. 4.3 Environmental systems biology

study of the biases of the methods employed in systems biology is warranted (Hazen et al. 2013).

4.2 Geochemical Processes

Setting up an environmental systems approach to bioremediation requires that first we have either a testable hypothesis or a (question) regarding the system to be tested. Examples include:

1. Chromium-contaminated aquifer impacting river (Zhang et al. 2015; Faybishenko et al. 2008)
2. Chlorinated-solvent-contaminated aquifer (Hazen 2010b)
3. Release of CH_4 and CO_2 from either agricultural soil or rain forest (Yao et al. 2018)
4. Biodegradation of an oil spill in the ocean (Atlas and Hazen 2011; Hazen et al. 2010)
5. Biodegradation of oil in soil (Hazen et al. 2003)
6. Combinatorial saccharification of lignocellulose (Woo et al. 2014)
7. Flowback fluid from hydraulic fracking (Trexler et al. 2014)
8. Memory response in bioremediation (Hazen 2018)
10. Landfill biodegradation rates and production of CH_4 (Borglin et al. 2004)

To clarify the boundary system, size must be measured using such assessment approaches. Google Map (GIS, GPS), vessel characteristics (size, e.g., reactor, test

tube fermenter), sampling equipment, and we must obtain any required permits. Knowledge of the system type is also necessary, i.e., whether it is soil, seawater, freshwater, glass, plastic, metal, plant species and strain, animal species and strain as well as operational taxonomic units (OTU), and the involved bacteria, archaea, fungi species or strain, and OTU. The dynamic parameters must also be defined and considered, e.g., direction and rates for (1) physical and chemical characteristics (current, wind, temperature, pH, redox, conductivity, alkalinity, nutrients (nitrogen, phosphorus, carbon, and metals), isotope chemistry, geology (seismic and radar tomography, water availability), (2) meteorological characteristics (rainfall, weather), (3) hydrological characteristics including climate hydraulic residence time, (4) biological characteristics (species, health, sex, age of plant or animal, pretreatment), and (5) whole environment respiration (CO_2/CH_4 flux).

A prime example of biogeochemical effects upon bioremediation would be the emphasis that competing terminal electron acceptors will control how low the redox potential can go and what metals can be either reduced or halorespired in the case of perchloroethylene/trichloroethylene (PCE/TCE) (Fig. 4.4). Presence of certain chemical species in groundwater can also be superb indicators of the dominant terminal electron-accepting (TEA) processes and the relative redox conditions. Many times, bioremediation has been applied to sites that have initially high nitrate, sulfate, or iron, thus resulting in a "stall" which stopped the progression to redox conditions that would allow dehalorespiration of PCE/TCE in groundwater. For bioremediation, monitoring carefully TEA concentrations offers better control of

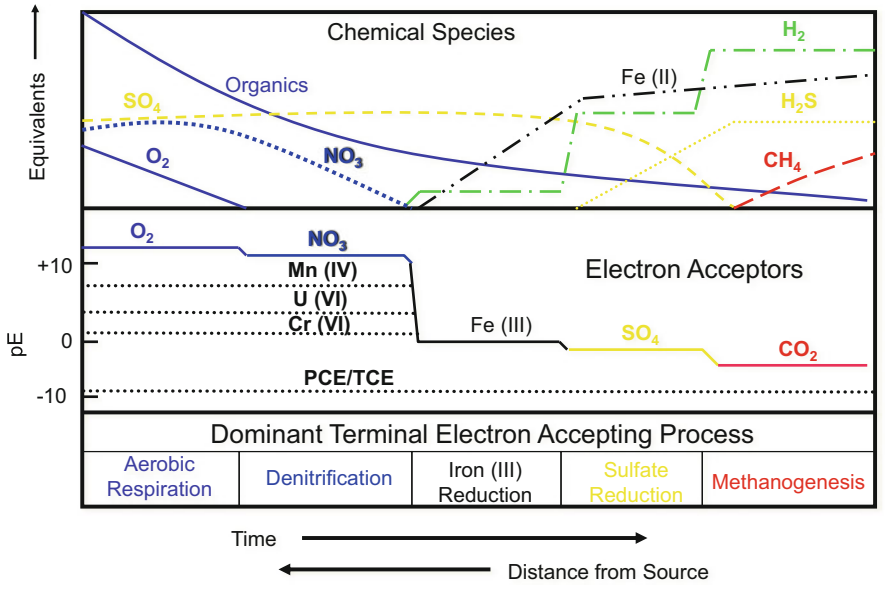

Fig. 4.4 Critical biogeochemistry (sulfate (SO_4), oxygen (O_2), nitrate (NO_3), manganese (Mn), uranium (U), chromium (Cr), perchloroethylene/trichloroethylene (PCE/TCE), iron (Fe), hydrogen (H_2), hydrogen sulfide (H_2S), methane (CH_4), carbon dioxide (CO_2))

electron donor additions and avoids reasons for "stalls" in expected bioremediation processes in situ.

4.3 Microbial Population and Community Correlates with Geochemistry

At this level we must determine key factors that impact communities and populations, a process which includes identifying the stresses and survival pathways (Fig. 4.5). For bioremediation practices, we must analyze the involved communities and their populations so as to better understand the structure and functional relationships and how they control the relevant geochemistries and contaminant degradation. This may be especially important in terms of understanding "fatal flaws," including life-cycle cost analysis, and understanding transitioning to natural attenuation in a treatment train. Measurements can include metagenomes, 16S rRNA, clone libraries, PhyloChip®, GeoChip®, phospholipid fatty acids (PLFA), and proteogenomics. This is, in addition to the other techniques for measuring

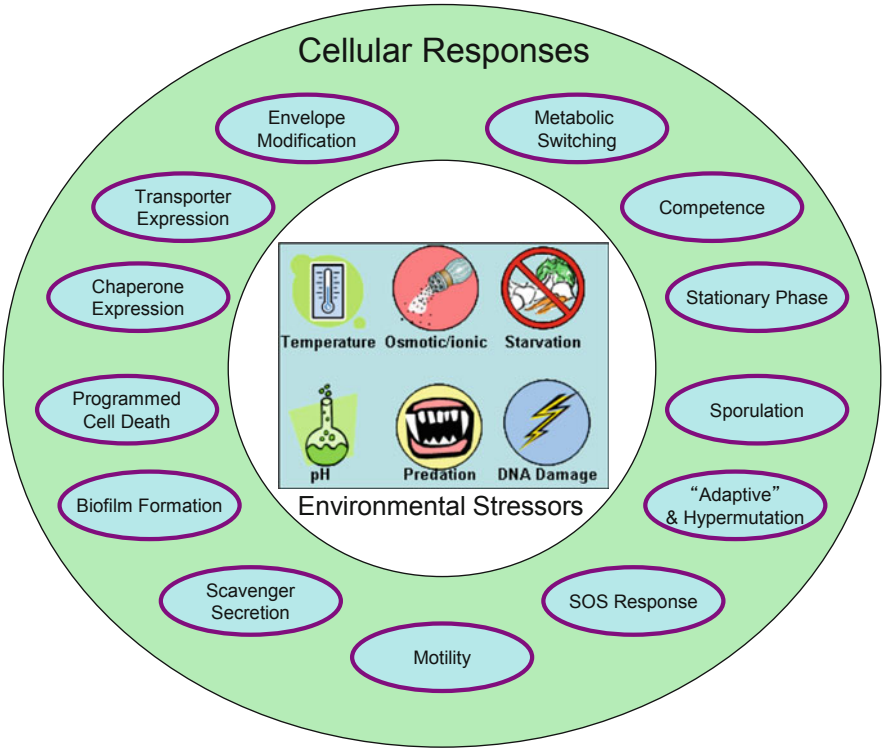

Fig. 4.5 Relevant stress responses

populations and activity, direct counts, fluorescent antibodies, live versus dead cell, fluorescent in situ hybridization (FISH), and enzyme activity probes (EAP).

Prior to the widespread availability of next-generation sequencing, environmental diversity classification was done by amplifying the 16S rRNA gene from DNA extracted from environmental samples that contained prokaryotes. With the current ability to instead sequence all genetic material contained in a sample was born a new possibility of contemporary metagenomics. Whole genome-based metagenomics aims to answer the same questions that 16S-based classification methods did; however there is potentially much more information contained in whole genome sequence information that is not present in 16S sequence information. While metagenomic data can be useful in describing an intangible component of nearly every environment, it should not be used without an understanding of its limitations and biases. It is implied that metagenomic sequences are subject to the same biases that all high-throughput sequencing is subject to. The reader is directed to a review of these biases in (Kircher and Kelso 2010). Outside of sequencing bias, other biases present in metagenomics come from sampling and filtering through the methods chosen for data analysis, which can skew or misrepresent sequence data. In addition to these biases, metagenome data is limited because it is often fragmented and incomplete. Rarely are whole genomes extracted from metagenome data, contiguous sequences are rarely longer than 5 kb (Thomas et al. 2012), and much of the sequence data is from organisms that are themselves not sequenced, uncultured and possibly unculturable, and therefore scarcely characterized.

4.3.1 16S rRNA

We have used 16S rRNA as a standard analytical marker ever since nucleotide sequencing became relatively cheap and fast for determining community structure. The 16S rRNA gene is believed to be highly conserved and has been used for decades as identification at different taxonomic units from Phyla to OTU. So, 16S rRNA analysis does in effect shows the relative community structural profile and has been used to indicate changes in diversity, evenness, and relative densities of certain groups. This trait can be particularly important to see if certain biodegraders are present and also to know if the community structure changes in ways that are either advantageous or detrimental to bioremediation. It cannot tell you anything about the activity, so even if you see that a group of biodegraders is present by 16S rRNA analysis, it does not tell you the group is metabolically active. In addition, 16S rRNA may not be highly conserved in some environments which would lead to erroneous conclusions. However, 16S rRNA analysis has been used to predict geochemistry for bioremediation (Smith et al. 2015).

4.3.2 PhyloChip®

PhyloChip® arrays are a technique which can be used to measure the presence versus absence as well as relative abundances of known prokaryotic species within a sample. The chip was developed by Affymetrix and uses a microarray consisting of 25mer groups which are single-stranded oligonucleotides called probes. Each probe is designed to be complementary to a certain region on the 16S rRNA gene of a particular species. The 16S rRNA gene codes for an RNA component of the small ribosomal subunit. It is a convenient taxonomic marker since it is present in all bacteria and archaea and contains conserved regions allowing for easy design of polymerase chain reaction (PCR) primers, as well as variable regions which allow for discrimination between species. PhyloChip® contains over one million different probes and allows discrimination of over 50,000 OTUs. PhyloChip® has been used to quantify relative abundances of microbial species in deepwater samples from the Gulf of Mexico, with analysis of those results showing that there was indeed a difference in the relative abundances of the microbial populations in samples collected from within an oceanic oil plume versus those which were collected outside of the oil plume. Plume samples were shown to be enriched for taxa (all of which were *Gammaproteobacteria*) known to degrade hydrocarbons. These species could be potential targets for bioremediation (Hazen et al. 2010).

4.3.3 GeoChip®

The GeoChip® functional gene microarray approach is based upon using deoxyribonucleic acid (DNA) microarray evaluations which contain oligonucleotide probes. These gene probes are focused on the biogeochemical cycles of certain metals and important nutrients including carbon, nitrogen, phosphorus, and sulfur. They can also detect resistance to antibiotics, virus, energy production, and the ability to degrade organic contaminants. The GeoChip® is also able to detect gyrB-based phylogenetic markers which represent useful knowledge when trying to detect slow-growing bacteria such as mycobacterium. Advantages to the GeoChip® are that it works for microorganisms not only from environmental origin such as soil, water, and air but also from human and animal sources. Another advantage is that prior knowledge of the microbial community being sampled is not necessary. It is also possible to detect low-abundance microorganisms, which helps to prevent annotation bias. The usage process is relatively quick, with the ability to receive data nearly every day from either DNA or ribonucleic acid (RNA). While the GeoChip® does have advantages, there are also some prevalent weaknesses. For instance, the GeoChip® cannot detect novel gene families because the GeoChip® is only capable of detecting those genes present on the probes. In addition to providing a small pool of information from which to draw information, the information for these functional gene array probes was developed using regions of genes which have

been well conserved throughout time on the 16S rRNA sequence. Problems could arise from focusing only on conserved genes due to natural variations and divergence in gene families, allowing some families to then be missed during probing. Other problems may stem from cross-hybridization problems introduced by match-mismatch probe sets or fluorochrome labeling skewing biases. GeoChip 5.0® contains 167,044 probes which are able to cover 395,894 coding sequences and 1500 gene families. It has been used to show functions that were inhibited during the Deepwater Horizon oil spill and predict geochemistry in a mixed waste site (Lu et al. 2012; He et al. 2018).

4.3.4 Phospholipid Fatty Acids

PLFAs are a main component of cell membranes, and their analysis can be a useful tool for microbial community analysis. These lipids generally degrade quickly upon cell death, allowing their analysis to only target viable cells, unlike many other microbiology techniques. However, PLFA does come with its own wide set of biases associated with its measurements. The PLFAs are measured and profiled with either a gas chromatography (GC) or a gas chromatography combined with mass spectrometry (GCMS) after they have been extracted and purified. Since this is a direct extraction method representing the fatty acids which doesn't include an intermediate step like PCR or culturing, its usage represents an advantage over several other techniques. Chromatography peaks representing the fatty acids are easily analyzed with proper equipment, without requiring other steps unlike nucleic acid techniques that necessitate sequencing. Fortunately, PLFA analysis is very sensitive and easily reproducible, with the benefit of its being fast and comparatively inexpensive. However, PLFA analysis can require large amounts of sample in order for the obtained data to be statistically significant, and with certain techniques, like fingerprinting, PLFA may require up to ten times more sample size than would analyses based upon examination of fatty acid methyl esters (FAME), which is another type of fatty acid analysis tool. There are four main types of analyses that can be done with the aid of PLFAs: total biomass, physiological indicators, fingerprinting, and taxonomic biomarkers. When examining the Deepwater Horizon spill plume, it was seen that PLFA data supported the 16S rRNA pattern analysis results and provided additional biomass measurements (Hazen et al. 2010).

4.3.5 Functional Gene Clone Libraries

Functional gene clone libraries are a method that can help elucidate gene functionality based upon nucleotide sequence data, as described. This technique is often employed at the population level of systems biology, particularly when the sequence of a gene of interest is known while the corresponding function remains cryptic.

Advantages of this technique include its culture independent nature, which bypasses the need for painstaking microbial isolation. For example, a gene can be PCR-amplified and cloned into a plasmid, such as pUC19, and then transformed into a model heterologous host such as *Escherichia coli*. In doing so, the functionality of the gene can be assayed in vivo independent of isolating the organism. Another advantage is the ability to selectively engineer a regulated promoter to drive the transcription of a gene, thereby providing much needed control of expression in relation to phenotype. A commonly used regulated promoter in *E. coli* is the T5-lac promoter. Limitations of this technique include the fact that functional heterologous expression of a gene in a foreign host may be considered a fortuitous event, at best. For example, some particular gene of interest may require a specific chaperone in order to produce key folds in the protein structure that are critical for enzymatic activity. If the heterologous host lacks the genes encoding for these chaperones, then the resulting protein may not fold correctly and thus yield no enzymatic activity. Even if protein folding takes place independently of chaperones, the enzyme in question may require additional cofactors that would need to be supplied in order for the process to function, such as vitamin B_{12}. For example, the *Dehalococcoides* solely rely on reductive dehalogenases to perform organohalide respiration yet lack the ability to produce the vitamin B_{12} essential to reductive dehalogenase function (Yan et al. 2013). If the heterologously expressed enzyme is indeed functional, the level of its activity may not be reflective of that which occurs in the gene native organism. This can occur because the activity of a gene may be regulated by possibly a single or even multiple regulatory networks in the native host and that regulation perhaps absent in the foreign host. Additionally, expression of the target gene may prove toxic to foreign hosts, which may compromise host cell viability. These limitations must be acknowledged and addressed when employing functional gene cloning in systems biology.

4.4 Measurement of Biomolecular Activities During Process Operations

Analyzing DNA, RNA, and proteins at the cellular level to understand cellular effects in terms of bioremediation can be effective in determining biomolecular activities during a critical remediation process. Measurements in addition to the ones listed in Sects. 4.2 and 4.3 include quantitative PCR (qPCR), stable-isotope probing, enzyme activity probes, metabolomics, proteogenomics, and fluxomics.

4.4.1 Stable-Isotope Probes

Stable-isotope probes (SIP) have been applied in numerous environmental studies (Madsen 2006), and many commercial laboratories offer analytical services (Microbial Insights Inc. (MI), Knoxville, TN). Pombo et al. (2002) used carbon-13 labeled acetate ($CH_3COO–$) which was injected and sequentially extracted from a groundwater zone in order to investigate its fate and transformation in the environment. Carbon-13 was measured in dissolved inorganic carbon (DIC) in groundwater and in phospholipid fatty acids (PLFA) in planktonic microbial biomass. The relative abundance of carbon-13 in DIC was significantly higher than were background levels in groundwater following injection of carbon-13 labeled $CH_3COO–$. This suggested that organic $CH_3COO–$ was transformed to inorganic carbon in the form of bicarbonate ($HCO_3–$). The relative abundance of carbon-13 in PLFA was also significantly higher than background levels in planktonic microbial biomass. This suggested the transformation of $CH_3COO–$ was, to some degree, microbial mediated. Pombo et al. (2002) clearly demonstrated that the fate and transformation of a carbon-13 labeled substrate can be investigated in an environmental system.

4.4.2 Enzyme Activity Probes

Enzyme activity probes (EAPs) are chemicals employed to detect and quantify the activities of specific microorganisms in environmental samples (e.g., soil, water, or sediment). These serve as alternative or surrogate substrates for the protein catalysts (enzymes) which are responsible for the metabolic activities of microorganisms. These surrogate compounds can be transformed by target enzymes into distinct and readily detectable products. Most enzymes cannot function well outside of cells due to ensuing rapid degradation or inactivation of the enzymes. There is a strong relationship between the transformation rate of EAPs and the number of active cells that possess the active form of enzyme. Moreover, enzymes are very selective, and EAPs can only be transformed by specific enzymes. Therefore, EAPs can be used to estimate the numbers of microorganisms which are capable of biodegrading a certain contaminant. They can also be used to determine in situ biodegradation rates of specific contaminants, such as chlorinated solvents.

4.4.3 Metabolomics or Metabolite Expression

Metabolomics or metabolite expression involves using a hydrophilic interaction chromatography technique coupled to tandem mass spectrometry (MS/MS) detection and capillary electrophoresis-mass spectrometry (CE-MS) methods for assaying amino acids, nucleosides, nucleotides, organic acid acetyl-CoAs (CoAs), redox

cofactors and metabolic intermediates of glycolysis, the tricarboxylic acid (TCA) and pentose phosphate pathways, etc.

4.4.4 Proteogenomics

Proteogenomics—the concept of "proteomics" was first coined to describe the complete protein complement expressed by a genome. The definition was specified later to be the protein complement of a given cell at a specified time, including the set of all protein isoforms and protein modifications. Proteomics-based methods have been used in discovery, quantification, and validation of protein-protein interactions. The types of large-scale experiments performed in the field of proteomics require specialized tools, and each tool may be suited to only a particular type of experimental design or question asked. Traditionally, proteomic analyses of complex protein samples involve the resolution of proteins using two-dimensional gel electrophoresis followed by the identification of resolved proteins by use of mass spectrometry. Bias can be caused in both the protein separation and identification processes. This technique has been used for studying contaminated sites (Jiao et al. 2011).

4.4.5 Fluxomics

Fluxomics studies the rates of metabolic reactions within a biological entity, including how those rates change, and connects them with dynamic physiology. This quantitative approach integrates in vivo measurements and estimates of reaction rates with stoichiometric network models to allow the determination of carbon flux through cellular networks in central metabolism. The combination of cellular network fluxes, termed the fluxome, represents a unique cellular phenotype. This approach can identify metabolic interactions leading to the development and rational design of cellular functions. This can be applied to make informed genetic modifications of industrial organisms and further analyze the resulting changes to optimize the metabolic network that results from those modifications. Two techniques deployed in fluxomics are ^{13}C flux analysis and flux balance analysis (FBA). The latter consists of utilizing the stable isotope of carbon ^{13}C to trace the partitioning of carbon through different pathways followed by either MS or nuclear magnetic resonance analysis (NMR) for identification of the labeled compounds. The former technique uses the stoichiometry of metabolic reactions in conjunction with many biological, chemical, and thermodynamic parameters to create a constrained model of metabolic flux. In ^{13}C isotope labeling, substrates like glucose can be labeled and fed to an organism, distributing the ^{13}C throughout the metabolic network, and the label then reach a steady state in the metabolite pool. Specific labeling patterns in metabolic intermediates can arise as a function of an organism's unique flux

distribution. This is followed by the use of either MS or NMR to measure these patterns and to construct the network flux distribution. Transitioning to the computational side with FBA, this technique consists of constructing models that constrain flux by observed cellular input and output measurements and by the stoichiometric, energy, and mass balance of metabolic reactions. Once constructed, these models can serve to predict fluxes and identify those that may either optimize or enhance a certain desired characteristic (Tang et al. 2007).

4.5 Simplified Laboratory Communities with Similar Function

Microrespiration and mesocosm studies are also an effective technique to determine biodegradation rates or mineralization rates. By mimicking environmental conditions as closely as possible and using samples from the environment that are extremely fresh, this technique can provide a reasonable approximation of the biodegradation or mineralization rates. Unfortunately, since it requires that samples be transferred to laboratory, the "bottle effect" can be enormous, as shown on a comparison of biodegradation of oil on ship studies with lab studies (Liu et al. 2017). It also can sometimes not represent the actual biodegradation rates because electron donors and electron acceptors may not be added to the samples in the right proportions to mimic what actually is going on in the environments (Baelum et al. 2012).

All of the methods used in Sects. 4.2–4.4 can be used in producing simplified lab communities and can be cross-referenced using the systems approach to link with other methods described in Sects. 4.2–4.4. Environmental systems approaches to bioremediation studies often require multiple lines of evidence cross-linked at different system levels (Fig. 4.6).

4.6 Cellular Network Analysis of Function

Phenomics, i.e., phenotype expression and physiology, are integral to understanding cellular network analysis of function. These measurements can include phenotypic microarrays, real-time analyses using Fourier transform infrared spectroscopy (FTIR), nuclear magnetic resonance (NMR), proteogenomics, PLFA, microbial isolation, FISH, EAP, and fluxomics.

4.6.1 High-Throughput Phenotypic Microarray

High-throughput phenotypic microarray (PM) systems are available, such as the OmniLog® by Biolog® Inc., of Hayward, California. Each PM consists of a 96-well plate that contains different variations of a cell culture medium and dye to test a particular phenotype or cell function. It can process up to 50 96-well plates (total of 4800 reactions at a time). Biolog® Inc. has over 1920 (20 plates) phenotypic tests available for gram-negative, gram-positive, yeast, filamentous fungi, and mammalian cells. These 20 different plates can detect respiration with different carbon sources, nitrogen sources, phosphorus and sulfur sources, nutrient supplements, peptide nitrogen sources, osmolytes, pH gradient, antibiotics, chemical sensitivity, and different metals. Furthermore, the user can also create a tailored plate with a colorimetric assay of choice, such as a particular combination of contaminant and electron donor. These systems can be used with isolates or mixed cultures or environmental samples under both aerobic and anaerobic conditions (Borglin et al. 2012).

4.6.2 Fourier Transform Infrared Spectroscopy

FTIR is a routinely used method of infrared spectroscopy and produces a unique molecular fingerprint of a sample based upon the absorption and transmission of infrared light through the chemical bonds contained in the sample. This makes FTIR very useful for identifying unknown materials, determining the quality of a sample, and quantifying the amount of components in a sample. Recent applications of FTIR include biological questions, especially those related to systems biology and bioremediation. Valuably, FTIR may be used as a high-throughput screening technique to identify and classify key small molecules, including metabolites in real-time for examining living bacteria on rocks and soil during active microbial exposure to contaminants (Baelum et al. 2012; Hazen et al. 2010).

All of the methods used in Sects. 4.2–4.5 can be used in producing simplified lab communities and can be cross-referenced using the systems approach to link to measurements in Sects. 4.2–4.5. Environmental systems approaches to bioremediation require multiple lines of evidence cross-linked at different systems levels (Fig. 4.6).

4.7 Molecular Analysis

Ecogenomics, definable as studies of genomes in an environmental context, can be done at the molecular level using 16S rRNA microarrays for community analyses, metagenome sequencing, annotation of sequences for environmental context,

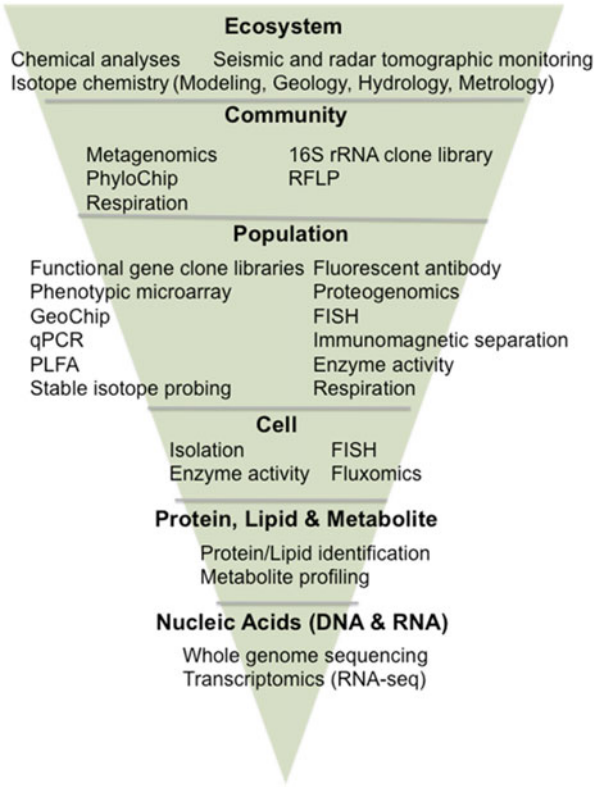

Fig. 4.6 Measurements at different system levels

transcriptomics-gene expression, mRNA expression arrays of one organism or functional group, real-time PCR analyses, protein/lipid identification, metabolite profiling, whole genome sequencing, and transcriptomics (RNA sequencing). Although Sects. 4.2–4.6 give examples of these techniques, it is important to realize that many of these techniques have inherent biases that can be multiplicative and illustrate the reason that conclusions must be considered carefully (Hazen et al. 2013). Figure 4.7 demonstrates where these biases occur and how they might affect the final bioinformatics analysis and conclusions (Hazen et al. 2013).

4.8 Predictive Models of Cellular Function

Predictive models of cellular function can involve annotation of sequences, comparative genomics, integration from biomolecules to ecosystems, and bioinformatics. Ecogenomics has been defined as the study of genomes in an environmental

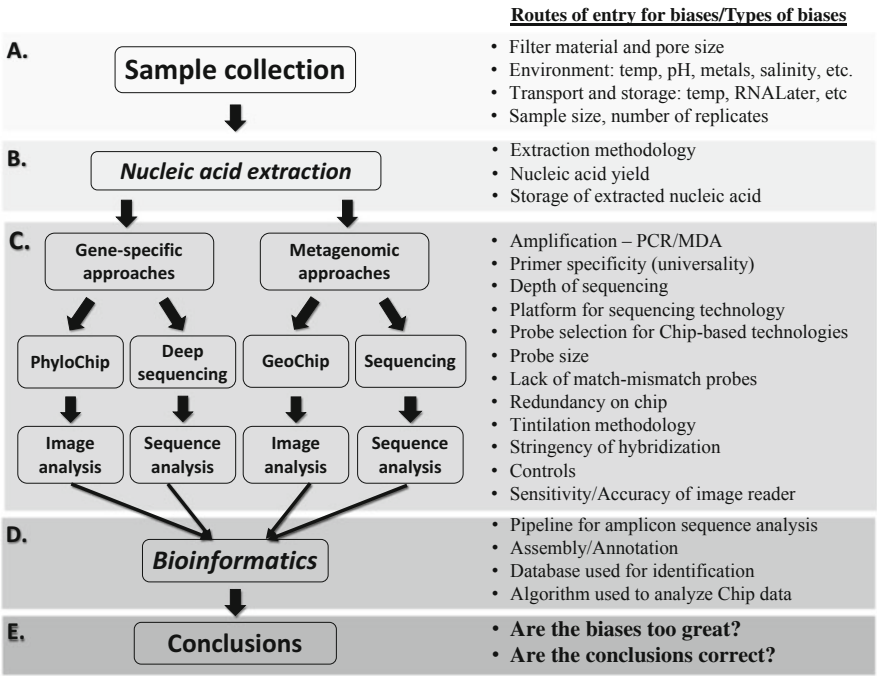

Fig. 4.7 Biases that can be introduced in a genomic pipeline (stabilization solution that stabilizes and protects cellular RNA (RNA later), polymerase chain reaction/multiple displacement amplification (PCR/MDA)) (after Hazen et al. 2013)

context, and Fig. 4.8 illustrates the associated conceptual integration for environmental systems biology (Fig. 4.8) (Deutschbauer et al. 2006).

> Genomics, functional genomics, proteogenomics and systems modeling allows for the analysis of community population structure, functional capabilities, and dynamics. The first step is to obtain DNA extracted from an environmental sample, either after cloning the DNA into a library or by direct sequencing. After the DNA sequence has been assembled, the computational identification of marker genes allows for identification and phylogenetic classification of members of the community and enables the design of probes for subsequent population structure experiments. The assignment of sequence fragments into groups that correspond to a single type of organism (a process called 'binning') is facilitated by identification of marker genes within the fragments, as well as by other characteristics such as G+C content bias and codon usage preferences. Computational genome annotation, consisting of gene prediction and assignment of their possible function using characterized homologs and genomic context, allows for description of the functional capabilities of the community. Knowledge of the genes present also enables functional genomic and proteomic techniques, applied to extracts of protein and RNA transcripts from the sample. These latter studies inform systems modeling, which can be used to interpret and predict the dynamics of the ecosystem and to guide future studies. (Deutschbauer et al. 2006)

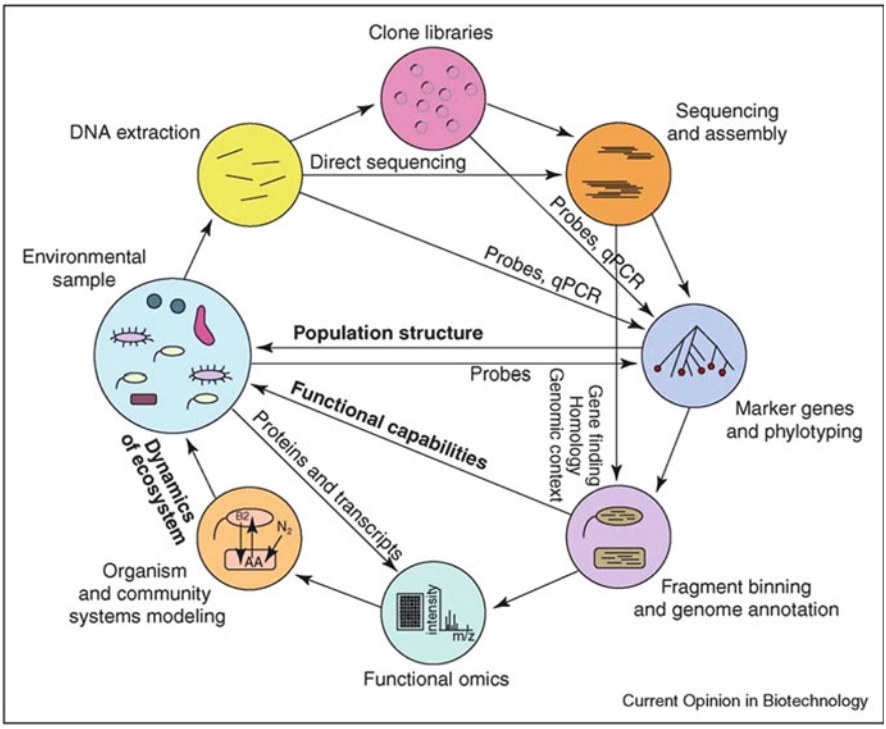

Fig. 4.8 Environmental systems biology using ecogenomics—using a variety of molecular techniques to provide multiple lines of evidence for microbial structure and function at organism, population, and community level (after Deutschbauer et al. 2006)

4.9 Predictive Models of Community Function

Stress response techniques can be used to monitor process control pathways for the purpose of achieving more effective bioremediation (Fig. 4.5) (Hazen et al. 2006). "Although the microbial stress response has been the subject of intensive laboratory investigation, the environmental reflection of the laboratory response to specific stresses has been little explored. However, it is only within an environmental context, in which microorganisms are constantly exposed to multiple changing environmental stresses, that there will be full understanding of microbial adaptive resiliency. Knowledge of the stress response in the environment will facilitate the control of bioremediation and other processes mediated by complex microbial communities" (Hazen et al. 2006). Biostimulation through the addition of nutrient amendments to contaminated environments has recently started to focus on specific stressors that could affect biodegradation and biotransformation processes. Holmes et al. (2004) monitored the nifD gene for nitrogen fixation during acetate stimulation of organic- and nitrogen-poor subsurface sediments. Although nifD expression decreased 100-fold after the addition of ammonium, it had no effect on rates of

toluene degradation or Fe(III) reduction. Thermodynamic analysis of Cr (VI) exposure to sulfate reducers has also been shown to induce an inhibition of growth and energy production that is similar to oxidative stress responses (Chardin et al. 2002). This suggests that commonality in stress responses might provide strategies that can be used to maximize biodegradation and biotransformation processes in situ against specific contaminants without increasing biomass of the target organism. Bioaugmentation (the addition of living cells) for the biodegradation of carbon tetrachloride has also been shown to benefit not only from nutrient balance but also from pH adjustments to avoid pH stress (Dybas et al. 2002). It has been demonstrated that by adding a combination of alkali, acetate, and phosphorus to aid a carbon tetrachloride degrader in a biocurtain strategy, biodegradation of carbon tetrachloride in groundwater passing through the biocurtain could be sustained at 100%.

4.10 Predictive Models of Geochemistry and Ecology

The success of any bioremediation application will be highly dependent on careful advance planning of the overall project, including consideration of the characterization, analysis, and monitoring that are to be done before and during the field deployment. The overall planning of the remediation needs to consider a number of steps from conceptual modeling to demobilization and report writing. For any field remediation, the first step is to form a conceptual model of the contaminant plume in the environment and how that environment effects that plume. The uncertainties in that conceptual model will provide the defining drivers for the characterization and monitoring needs. For example, characteristics of an aquifer will have a profound impact on its remediation strategy. The largest part of the expense of any remediation project is the characterization and monitoring. For example, hydraulic conductivities can have a severe effect on your ability to deliver nutrients to the subsurface and can be the most limiting part of the environment. Fortunately, new advances in geophysics and hydraulic push technology such as Geoprobe® have enabled us to characterize sites in a fraction of the time and cost. Once we have established the hydrology and basic geochemistry at a site and used that data to refine our conceptual model, a base line characterization of the microbiology is essential to establish that the right microorganisms are present, that they can be stimulated, and that no undesirable reactions associated with the stimulants or daughter products from the stimulation will occur. This usually requires some treatability and soil compatibility studies as well as monitoring of microbial community structure and function to establish the base conditions prior to stimulation. For example, some metals like arsenic actually increase solubility under the same redox potentials that precipitate chromium and uranium.

Perhaps the best documented and most widely used model for hydrocarbon bioremediation has been the BIOPLUME® model (Borden and Bedient 1986). This model, now in its forth version, uses a series of simultaneous equations to

simulate growth, decay, and transport of microorganisms, oxygen, and hydrocarbons. Rifai et al. (1987) later modified this model (BIOPLUME II®) to incorporate the USGS two-dimensional method of characteristic model (Konikow and Bredeheoft 1978). The original BIOPLUME® model was used to simulate PAH biodegradation at a Texas Superfund site (Borden and Bedient 1986). BIOPLUME II® has been used to model biodegradation of aviation fuel at the US Coast Guard Station at Traverse City, Michigan (Rifai et al. 1988), and to characterize benzene biodegradation over 3 years in another shallow aquifer (Chiang et al. 1989; Choi et al. 2009). Travis and Rosenberg (Travis and Rosenberg 1997) used a numerical simulation model to successfully predict aerobic bioremediation of chlorinated solvents in the groundwater and vadose zone using methane biostimulation at the US Department of Energy's Savannah River Site near Aiken, South Carolina. Their model also used a series of simultaneous equations for microbial growth, and nutrient limitations, in addition to modeling contaminant, microbe, and nutrient transport. Their model predicted the amount of TCE that was biodegraded during a 14-month, full-scale demonstration and was validated by five other methods (Hazen et al. 1994). Other models that are in use these days are BIOSCREEN® (USEPA 2018b), BIOCHLOR® (USEPA 2018a), REMChlor® (USEPA 2018c), REMFuel® (USEPA 2018d), and Matrix Diffusion Toolkit® (GSI_Environmental_Inc 2018). Models like these are becoming increasingly important as our need to understand the terrestrial subsurface "black box" of bioremediation increases in response to increased emphasis on intrinsic in situ bioremediation as a final solution (Hazen 2010a, b, c).

Intrinsic bioremediation is developing rapidly as an important alternative or end-game approach for many contaminated environments. This strategy of a monitored natural attenuation (MNA) through characterization, treatability studies, risk assessment, modeling, and verification monitoring of contaminated environments was first proposed by John Wilson of the US Environmental Protection Agency's Robert S. Kerr Lab in the early 1990s, following which the development and regulatory acceptance of this strategy have increased exponentially. Certainly, much of this rapid deployment of intrinsic bioremediation has been due to the crushing financial burden that environmental cleanup represents and our need to use more risk-based cleanup goals for the thousands of new contaminated sites identified every year. The MNA strategy carries with it a burden of proof for (1) risk to health and the environment and (2) a model that will accurately predict the unengineered bioremediation of the environment (Hazen 2010a, b, c).

One of the most recent models to use a more environmental systems biology approach was the Structured Learning in Microbial Ecology (SLiME) model. In developing the SLiME model, we used data from our 100-well survey campaign at the Department of Energy's (DOE) Oak Ridge-contaminated field site from groundwater and also from similar measurements from the *Deepwater Horizon* spill response phase of the deepwater plume (Smith et al. 2015). Although these two sites represented completely different contaminants, knowledge of the microbial community structure enabled predictions for the contaminate concentrations at both sites (Fig. 4.9).

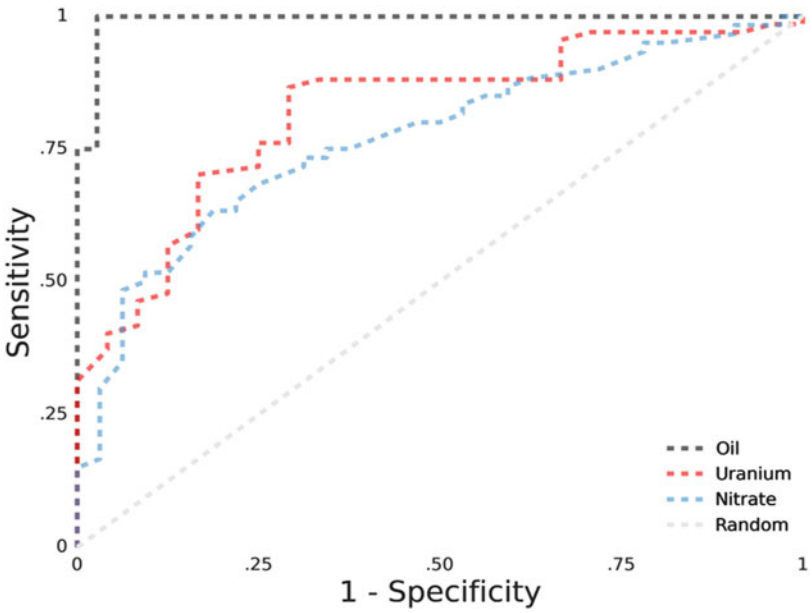

Fig. 4.9 Structured Learning in Microbial Ecology (SLiME) identified biological features associated with the presence of oil, uranium, and nitrate, i.e., predicting contaminants using microbial community structure in situ (after Smith et al. 2015)

4.11 Conclusion and Research Needs

Once all of the above aspects have been thoroughly considered, then a *Field Test Plan* can be developed for the given bioremediation. Environmental systems biology *Field Test Plans* must incorporate the following components: (1) hypotheses to be tested and why, (2) boundaries of environment system, (3) measurements to be made, (4) sampling (protocols, holding times, storage, etc.), (5) resources needed (costs, equipment, people, time, etc.), (6) the Field Test Plan summarizing the best possible scenario of 2–5 to accomplish 1 (including permits, safety, protocols, responsibilities, budgets, priorities, expected outcomes, reports and publications, data management plan, and science of opportunity). After these components have been considered, then (7) feedback to the Field Test Plan involves mobilization, implementation, demobilization, final analyses, and use of models for the preparation of reports and publications (Hazen and Sayler 2016). One campaign for which we used an environmental systems biology approach was at DOE's Oak Ridge-contaminated field site that was studied for over 2 years, and the project illustrated the successful management of a large-scale environmental systems approach (Smith et al. 2015) (Fig. 4.10).

By using an environmental systems biology approach to bioremediation and linked systems at all levels, providing multiple lines of evidence involving

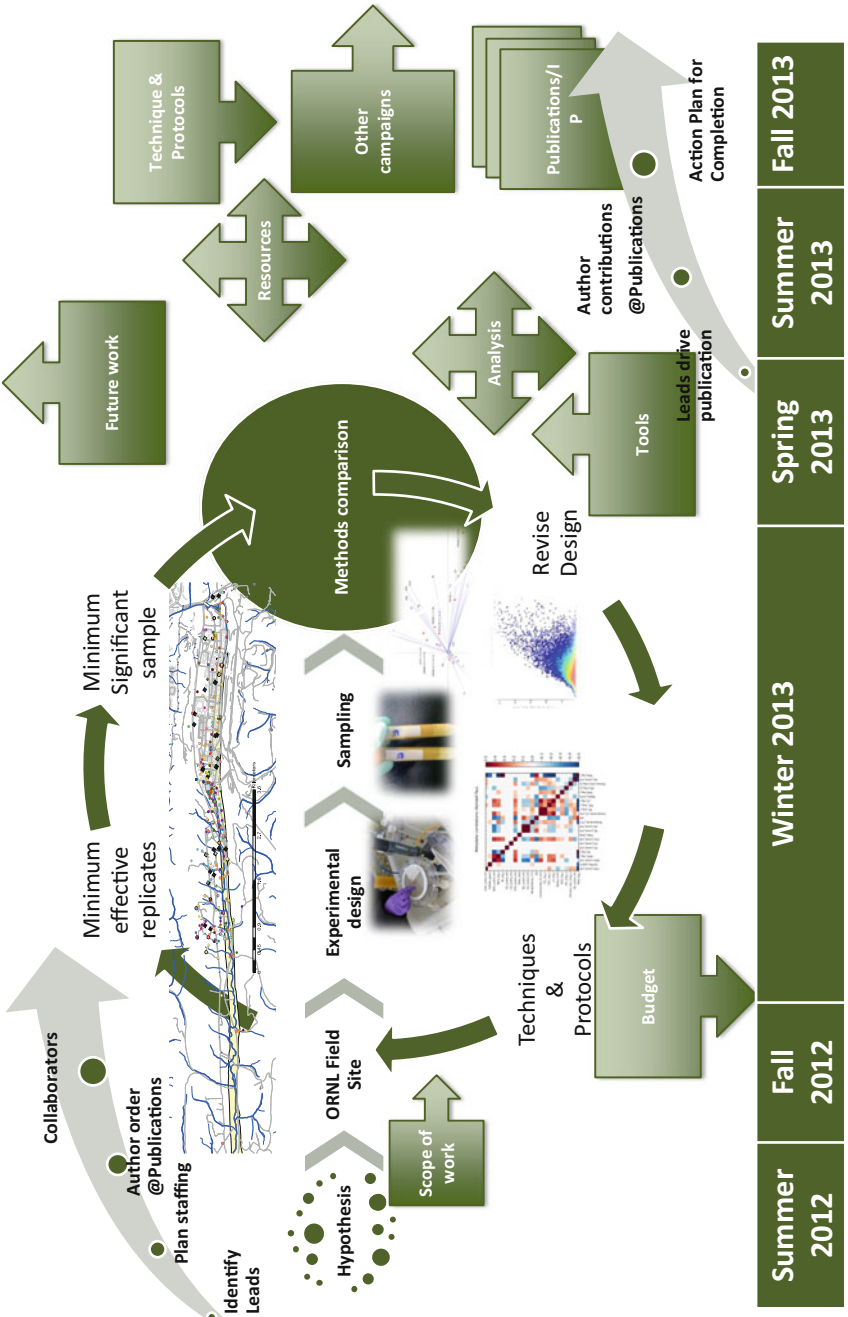

Fig. 4.10 100-well survey timeline

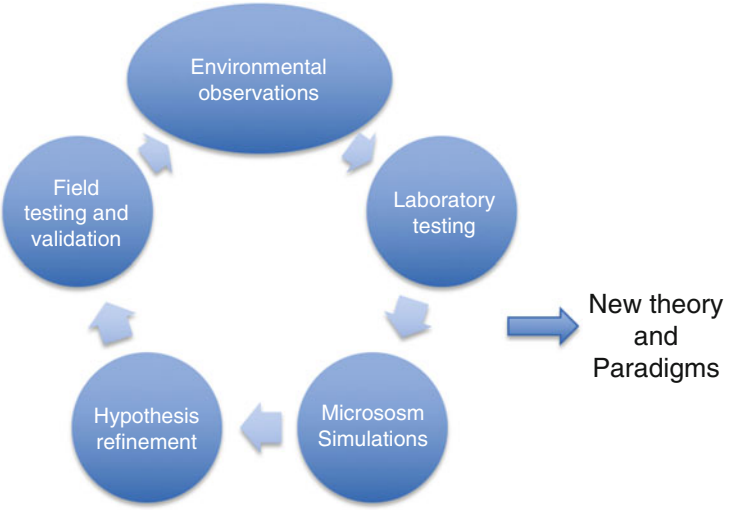

Fig. 4.11 Iterations for using an environmental systems biology approach to bioremediation

environmental observations, laboratory testing, microcosm simulations, hypothesis refinement, field testing and validation, and iteration of this circle, we will be able to make new theories and paradigms for bioremediation of contaminated environments (Fig. 4.11).

Compliance with Ethical Standards

Conflict of Interest Terry C. Hazen declares that he has no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by the author.

References

Atlas RM, Hazen TC (2011) Oil biodegradation and bioremediation: a tale of the two worst spills in US history. Environ Sci Technol 45(16):6709–6715. https://doi.org/10.1021/es2013227

Baelum J, Borglin S, Chakraborty R et al (2012) Deep-sea bacteria enriched by oil and dispersant from the Deepwater Horizon spill. Environ Microbiol 14(9):2405–2416. https://doi.org/10.1111/j.1462-2920.2012.02780.x

Borden RC, Bedient PB (1986) Transport of dissolved hydrocarbons influenced by reaeration and oxygen limited biodegradation 1. Theoretical development. Water Resour Res 22:1973–1982

Borglin SE, Hazen TC, Oldenburg CM et al (2004) Comparison of aerobic and anaerobic biotreatment of municipal solid waste. J Air Waste Manage Assoc 54(7):815–822

Borglin S, Joyner D, DeAngelis KM et al (2012) Application of phenotypic microarrays to environmental microbiology. Curr Opin Biotechnol 23(1):41–48. https://doi.org/10.1016/j.copbio.2011.12.006

Chardin B, Dolla A, Chaspoul F et al (2002) Bioremediation of chromate: thermodynamic analysis of the effects of Cr(VI) on sulfate-reducing bacteria. Appl Environ Microbiol 60(3):352–360. https://doi.org/10.1007/s00253-002-1091-8

Chiang CY, Salanitro JP, Chai EY, Colthart JD, Klein CL (1989) Aerobic biodegradation of benzene, toluene, and xylene in sandy aquifer, and data analysis and computer modeling. Ground Water 27:823–834

Choi NC, Choi JW, Kim SB et al (2009) Two-dimensional modelling of benzene transport and biodegradation in a laboratory-scale aquifer. Environ Technol 30(1):53–62

de Lorenzo V, Marliere P, Sole R (2016) Bioremediation at a global scale: from the test tube to planet Earth. Microb Biotechnol 9(5):618–625. https://doi.org/10.1111/1751-7915.12399

Deutschbauer AM, Chivian D, Arkin AP (2006) Genomics for environmental microbiology. Curr Opin Biotechnol 17(3):229–235. https://doi.org/10.1016/j.copbio.2006.04.003

Dybas MJ, Hyndman DW, Heine R et al (2002) Development, operation, and long-term performance of a full-scale biocurtain utilizing bioaugmentation. Environ Sci Technol 36 (16):3635–3644. https://doi.org/10.1021/es0114557

Faybishenko B, Hazen TC, Long PE et al (2008) In situ long-term reductive bioimmobilization of Cr(VI) in groundwater using hydrogen release compound. Environ Sci Technol 42 (22):8478–8485. https://doi.org/10.1021/es801383r

GSI_Environmental_Inc (2018) Matrix Diffusion Toolkit®. GSI Environmental Inc. http://www.gsi-net.com/en/software/free-software/matrix-diffusion-toolkit.html. Accessed April 19, 2018

Hazen TC (2010a) Biostimulation. In: Timmis KN (ed) Handbook of hydrocarbon microbiology: microbial interactions with hydrocarbons, oils, fats and related hydrophobic substrates and products. Springer, Berlin

Hazen TC (2010b) Cometabolic bioremediation. In: Timmis KN (ed) Handbook of hydrocarbon microbiology: microbial interactions with hydrocarbons, oils, fats and related hydrophobic substrates and products. Springer, Berlin

Hazen TC (2010c) In situ groundwater bioremediation. In: Timmis KN (ed) Handbook of hydrocarbon microbiology: microbial interactions with hydrocarbons, oils, fats and related hydrophobic substrates and products. Springer, Berlin

Hazen TC, Sayler GS (2016) Environmental systems microbiology of contaminated environments. In: Yates M, Nakatsu C, Miller R, Pillai S (eds) Manual of environmental microbiology, 4th edn. ASM Press, Washington, DC, pp 5.1.6–1–5.1.6–10. https://doi.org/10.1128/9781555818821.ch5.1.6

Hazen TC (2018) In situ: groundwater bioremediation. In: Consequences of microbial interaction with hydrocarbons, oils and lipids: biodegradation and bioremediation. Handbook of hydrocarbon and lipid microbiology series. Springer, Cham, pp 1–18. DOI: https://doi.org/10.1007/978-3-319-44535-9_11-1

Hazen TC, Lombard KH, Looney BB et al (1994) Summary of in situ bioremediation demonstration (methane biostimulation) via horizontal wells at the Savannah river site integrated demonstration project. In situ remediation: scientific basis for current and future technologies, pts 1 and 2. Battelle Press, Columbus

Hazen TC, Tien A, Worsztynowicz A et al. (2003) Biopiles for remediation of petroleum-contaminated soils: a polish case study. In: Sasek V, Glaser J, Baveye P (eds) Proceedings of the NATO advanced research workshop on the utilization of bioremediation to reduce soil contamination: problems and solutions, Prague, Czech Republic, June 14, 2000. NATA Science Series IV: Earth and Environmental Sciences. Kluwer Academic Publishers, pp 229–246

Hazen TC, Stahl DA, Hazen TC et al (2006) Using the stress response to monitor process control: pathways to more effective bioremediation. Curr Opin Biotechnol 17(3):285–290. https://doi.org/10.1016/j.copbio.2006.03.004

Hazen TC, Dubinsky EA, DeSantis TZ et al (2010) Deep-sea oil plume enriches indigenous oil-degrading bacteria. Science 330(6001):204–208. https://doi.org/10.1126/science.1195979

Hazen TC, Rocha AM, Techtmann SM (2013) Advances in monitoring environmental microbes. Curr Opin Biotechnol 24(3):526–533. https://doi.org/10.1016/j.copbio.2012.10.020

He Z, Zhang P, Wu L et al (2018) Microbial functional genes predict groundwater contamination and ecosystem functioning. MBio 9:e02435–e02417. https://doi.org/10.1128/mBio.02435-17
Holmes DE, Nevin KP, Lovley DR (2004) In situ expression of nifD in Geobacteraceae in subsurface sediments. Appl Environ Microbiol 70(12):7251–7259. https://doi.org/10.1128/aem.70.12.7251-7259.2004
Jiao YQ, D'Haeseleer P, Dill BD et al (2011) Identification of biofilm matrix-associated proteins from an acid mine drainage microbial community. Appl Environ Microbiol 77(15):5230–5237. https://doi.org/10.1128/aem.03005-10
Kay E, Lesk VI, Tamaddoni-Nezhad A et al (2010) Systems analysis of bacterial glycomes. Biochem Soc Trans 38:1290–1293. https://doi.org/10.1042/bst0381290
Kircher M, Kelso J (2010) High-throughput DNA sequencing – concepts and limitations. BioEssays 32(6):524–536. https://doi.org/10.1002/bies.200900181
Konikow LF, Bredeheoft JD (1978) Computer model of two dimensional solute transport and dispersion in ground water. techniques of water resources investigations of the U.S. Geological Survey: Washington, DC
Liu J, Techtmann SM, Woo HL et al (2017) Rapid response of eastern mediterranean deep sea microbial communities to oil. Sci Reports 7:11. https://doi.org/10.1038/s41598-017-05958-x
Lu ZM, Deng Y, Van Nostrand JD et al (2012) Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. ISME J 6(2):451–460. https://doi.org/10.1038/ismej.2011.91
Madsen EL (2006) The use of stable isotope probing techniques in bioreactor and field studies on bioremediation. Curr Opin Biotechnol 17(1):92–97
Pombo SA, Schroth MH, Pelz O, Zeyer J (2002) Tracing microbial activity in a contaminated aquifer at the field scale using C-13-labeling of bacterial fatty acids. Geochim Cosmochim Acta 66(15A):A610–A610
Rifai HS, Bedient PB, Borden RC et al. (1987) BIOPLUME II computer model of two-dimensional contaminant transport under the influence of oxygen limited biodegradation in groundwater user's manual version 1.0 Houston
Rifai HS, Bedient PB, Wilson JT et al (1988) Biodegradation modeling at an aviation fuel spill. ASCE J Environ Eng 114:1007–1029
Smith MB, Rocha AM, Smillie CS et al (2015) Natural bacterial communities serve as quantitative geochemical biosensors. MBio 6(3):13. https://doi.org/10.1128/mBio.00326-15
Tang YJ, Chakraborty R, Martin HG et al (2007) Flux analysis of central metabolic pathways in *Geobacter metallireducens* during reduction of soluble Fe(III)-nitrilotriacetic acid. Appl Environ Microbiol 73(12):3859–3864. https://doi.org/10.1128/aem.02986-06
Thomas T, Gilbert J, Meyer F (2012) Metagenomics – a guide from sampling to data analysis. Microb Inf Exp 2(1):3–3. https://doi.org/10.1186/2042-5783-2-3
Travis BJ, Rosenberg ND (1997) Modeling in situ bioremediation of TCE at Savannah River: effects of product toxicity and microbial interactions on TCE degradation. Environ Sci Technol 31(11):3093–3102
Trexler R, Solomon C, Brislawn CJ et al (2014) Assessing impacts of unconventional natural gas extraction on microbial communities in headwater stream ecosystems in Northwestern Pennsylvania. Front Microbiol 5:522. https://doi.org/10.3389/fmicb.2014.00522
USEPA (2018a) BIOCHLOR®. United States environmental protection agency. https://www.epa.gov/water-research/biochlor-natural-attenuation-decision-support-system. Accessed April 19, 2018
USEPA (2018b) BIOSCREEN®. United States environmental protection agency. http://www.epa.gov/water-research/bioscreen-natural-attenuation-decision-support-system. Accessed April 19, 2018
USEPA (2018c) REMChlor®. United States environmental protection agency. http://www.epa.gov/water-research/remediation-evaluation-model-chlorinated-solvents-remchlor. Accessed April 19, 2018
USEPA (2018d) REMFuel®. United States environmental protection agency. http://www.epa.gov/water-research/remediation-evaluation-model-fuel-hydrocarbons-remfuel. Accessed April 19, 2018

Woo HL, Hazen TC, Simmons BA et al (2014) Enzyme activities of aerobic lignocellulolytic bacteria isolated from wet tropical forest soils. Syst Appl Microbiol 37(1):60–67. https://doi.org/10.1016/j.syapm.2013.10.001

Yan J, Im J, Yang Y et al (2013) Guided cobalamin biosynthesis supports Dehalococcoides mccartyi reductive dechlorination activity. Phil Trans R Soc B Biol Sci 368(1616):10. https://doi.org/10.1098/rstb.2012.0320

Yao Q, Li Z, Song Y et al (2018) Community proteogenomics reveals the systemic impact of phosphorus availability on microbial functions in tropical soil. Nat Ecol Evol 2:1–11. https://doi.org/10.1038/s41559-017-0463-5

Zhang P, Van Nostrand JD, He Z et al (2015) A slow-release substrate stimulates groundwater microbial communities for long-term in situ Cr(VI) reduction. Environ Sci Technol 49 (21):12922–12931. https://doi.org/10.1021/acs.est.5b00024