Published in final edited form as:
Biotechnol J. 2010 December ; 5(12): 1277–1296. doi:10.1002/biot.201000181.

Prokaryotic Gene Clusters: A Rich Toolbox for Synthetic Biology

Michael Fischbach[1] and **Christopher A. Voigt**[2]

Michael Fischbach: fischbach@fischbachgroup.org; Christopher A. Voigt: cavoigt@gmail.com

[1]Department of Bioengineering and Therapeutic Sciences, University of California – San Francisco, MC 2530, Room 308C, 1700 4th Street, (415) 514-9435

[2]Department of Pharmaceutical Chemistry, University of California – San Francisco, MC 2540, Room 408C, 1700 4th Street, San Francisco, CA 94158, (415) 502-7050

Abstract

Bacteria construct elaborate nanostructures, obtain nutrients and energy from diverse sources, synthesize complex molecules, and implement signal processing to react to their environment. These complex phenotypes require the coordinated action of multiple genes, which are often encoded in a contiguous region of the genome, referred to as a gene cluster. Gene clusters sometimes contain all of the genes necessary and sufficient for a particular function. As an evolutionary mechanism, gene clusters facilitate the horizontal transfer of the complete function between species. Here, we review recent work on a number of clusters whose functions are relevant to biotechnology. Engineering these clusters has been hindered by their regulatory complexity, the need to balance the expression of many genes, and a lack of tools to design and manipulate DNA at this scale. Advances in synthetic biology will enable the large-scale bottom-up engineering of the clusters to optimize their functions, wake up cryptic clusters, or to transfer them between organisms. Understanding and manipulating gene clusters will move towards an era of genome engineering, where multiple functions can be “mixed-and-matched” to create a designer organism.

Keywords

Systems Biology; Genetic Parts; Devices; Refactoring; Biotechnology

Introduction

Gene clusters are the genetic building blocks of bacteria and archaea. Prokaryotic genomes are highly organized and the genes associated with a particular function often occur near each other [1]. Occasionally, all of the genes that are necessary for a discrete function form a cluster in the genome. These clusters encode functions that affect all aspects of the life style of bacteria, including nutrient scavenging, energy production, chemical synthesis, and environmental sensing. Large protein scaffolds, nano-machines, and cytoplasmic organelles are also encoded within clusters. These functions could play a central role for many applications in biotechnology; however, their complexity makes them difficult to engineer. Here, we survey the wide range of cellular functions that are known to be encoded in these genetically compact units with an eye on their potential ability to be transferable modules in multiple host species for engineering applications.

The organization of genes into clusters may facilitate the transfer of complete functions during evolution [1, 2]. All of the gene clusters presented in this review have some evidence for horizontal transfer, including phylogenetic trees disparate from ribosomal RNA, differing G+C content, and the presence of flanking transposon/integron genes [3, 4]. Phage genomes and conjugative plasmids also contain bacterial gene clusters, implying that a

mobile element can confer a fitness advantage on its host by adding a novel function. For example, the photosynthetic apparatus [5] and type IV pili [6] have been observed in phage genomes. Because gene clusters appeared and were shaped by interspecies transfer, it is intriguing that they could be fodder for genome building, where they provide a convenient unit of DNA that could be utilized to introduce a novel function into a synthetic organism. To date, such transfers are sometimes successful and sometimes fail for unknown reasons [7, 8]. Potential problems include that the cluster may rely on regulatory interactions that are not present in the new host, the genes do not express or express at the wrong ratios, or there are auxiliary interactions with or dependencies on the host [1].

Within gene clusters, there can also be sub-gene clusters that evolve separately. This modular organization of clusters within clusters enables rapid diversification and can replicate a useful function in multiple contexts. Two examples of such sub-clusters are microcompartments that can sequester toxic metabolic intermediates (Section I.B) and the stressosome that can integrate signals and control different signaling mechanisms (Section V.A). Subclusters also occur within metabolic pathways, where particular conversion (e.g., the modification of a sugar moiety) can occur in different contexts. Examples of such sub-clusters are present in the erythromycin pathway (Section III.C). The useful functions encoded by these sub-clusters have propagated into different metabolic and signaling pathways.

As the number of complete sequenced genomes grows, it has become clear that many gene clusters are “cryptic;” in other words, there are no known conditions under which the genes are expressed [9]. Homology analysis can be useful (albeit inexact) in predicting the general classes of molecules produces. For example, there may be many novel antibiotics and other pharmaceuticals that are encoded by such clusters. Sometimes, it is possible to “wake up” a cluster by engineering its regulatory circuitry [10]. This could either be through the deletion of a repressor or the addition of an inducible system. However, many clusters remain intransigent to these approaches. As the sequence databases grow, it is going to be increasingly tempting to be able to access the functions encoded therein.

Genetic engineering is moving towards the era of the genome. Automated DNA synthesis has continued to advance, where the size of routine orders has increased to >50,000 bp, with declining cost and turnaround time [11]. Recently, the entire wild-type genome of *Mycoplasma* was synthesized and transferred into a new cell, producing a living organism [12]. However, no design was implemented in this tour-de-force project; essentially, a natural genome was replicated. In an attempt to improve our design capacity, synthetic biology has been contributing a growing toolbox of genetic parts (*e.g.*, ribosome binding sites, promoters, terminators) and devices (*e.g.*, genetic circuits, sensors) that enable programmable control over transcription and translation [13]. In addition, methods have been developed to rapidly assemble these parts into intermediate 10kb fragments [14], which can then be further assembled to multi-100kb pieces [15]. As synthetic biology moves towards genome design, gene clusters are an appropriate intermediate stepping-stone. On one hand, they are themselves composed of genetic parts and devices. On the other, they could be hierarchically combined to add functions to a genome. Indeed, this type of construction occurs frequently in nature, where large plasmids have been discovered that contain multiple gene clusters (for example, the pRSB107 plasmid combines 9 antibiotic resistance gene clusters and one that scavenges iron) [16].

This review is organized to describe and compare clusters that encode a variety of functions from different species. The clusters are loosely organized into five classes according to the type of function they encode: structural, scavenging, synthesis, energy, and sensing. Each example focuses on a well studied instance of that cluster. Many variations of each cluster

exist. We also describe the current and potential applications in biotechnology for each of the gene clusters.

I. Nano-machines, Organelles, and Large Protein Structures

I.A. A Molecular Hypodermic Needle: The Type III Secretion System

Salmonella typhimurium (sprB to invH): The type III secretion system (T3SS) is a molecular machine that exports proteins from the cytoplasm to the extracellular environment (Figure 1) [17]. Many Gram negative pathogens have such systems, where it forms a syringe-like structure [18] that injects proteins into animal or plant cells in order to hijack a variety of host processes. The length of the needle varies depending on the function, from 60 nm in *Yersinia* to 2 μm in *Pseudomonas syringae*, where it needs to penetrate a thick plant cell wall. Effector proteins that are exported have a N-terminal secretion tag and a chaperone binding domain that direct it to the needle. The proteins are actively unfolded before being transported through a 2 nm pore. All of the genes that are required to form the needle, as well as chaperones and effectors are encoded within a single gene cluster (Figure 2). Additional effector proteins can be scattered throughout the genome. A complex regulatory network encoded within the cluster integrates environmental signals and controls the dynamics of gene expression [19]. Pathogens often have multiple clusters encoding needles that are responsible for different phases of an infection. The flagellum, involved in propulsion when bacteria swim, is a distantly related T3SS, but its genes are usually not organized into a single cluster.

Salmonella and other pathogenic bacteria have been explored for therapeutic uses as a mechanism to deliver vaccines and to block the growth of tumors. As a live vaccine, *Salmonella* has been used to deliver heterologous antigens from diverse sources to the immune system, including through the integration of peptides into effectors that are secreted from the T3SS encoded in *Salmonella* pathogenicity island 1 (SPI-1) [20]. *Salmonella* naturally localizes to tumors after infection and it has been explored as an anticancer therapeutic, and has gone through clinical trials. The growth suppression of tumors has been shown to be dependent on the T3SS encoded within *Salmonella* pathogenicity island 2 (SPI-2) [21]. More generally, protein secretion is a critical tool in biotechnology. It is useful to ease purification, to avoid self-assembly and inclusion body formation, and for enzymes that act on substrates that cannot diffuse into the cell. The SPI-1 T3SS has been used to export a variety of proteins, including spider silk fibroins at a rate up to 1.8 mg/L-hr [22]. However, the flux is limited to a narrow range of environmental conditions and the maximum rate is only maintained for 4-6 hours.

I.B. Chemical Reactors within Bacteria: Microcompartments

Salmonella typhimurium (pduA to pduX): Many bacteria build geometrically regular polyhedral organelles with a diameter of 80-200 nm (Figure 1) [23, 24]. These act as microcompartments to encapsulate enzymes that participate in metabolic pathways where an intermediate is toxic or requires concentration. The canonical example is the carboxysome, which co-localizes carbonic anydrase, which increases the concentration of CO_2 with the CO_2-fixing enzyme RuBisCo. There are several examples where a pathway produces an intermediate molecule that causes toxicity when it is produced in the cytoplasm. Examples include pathways for the utilization of propanediol and ethanolamine, whose toxic intermediates are aldehyde and acetaldehyde, respectively. An example of the former is the *pdu* gene cluster in *Salmonella typhimurium*. The proteins that form the microcompartments produce a six-fold symmetry with a 2-3 Å pore that can potentially expand to 11-13 Å. The pore controls the transport of molecules in and out of the microcompartment and has a variable amino acid sequence. The core structural components of the microcompartment are conserved and appear in gene clusters with the enzymes that make up the metabolic pathway

(Figure 2). For example the ethanolamine utilization gene cluster has seventeen genes, four of which are the compartment shell proteins.

Microcompartments exist in natural pathways to solve many of the problems that emerge in metabolic engineering. Frequently, an intermediate metabolite is toxic and this is detrimental to the cell if the flux in and out are unbalanced, causing the intermediate to accumulate [25]. Further, some enzymes have intrinsically high K_m's requiring that the substrate be concentrated. Other functions require an anaerobic environment, and it has been shown that microcompartments can exclude oxygen molecules. Effectively harnessing microcompartments will require that targeting sequences be available to direct enzymes to the organelle [26] and protein engineering methods to alter the pores to allow desired substrates to enter [23]. Several microcompartments have been predicted to be involved in pathways of direct relevance to biotechnology, including in the production of ethanol in *Vibrio* [27] and butyrate in *Clostridium* [28].

I.C. Balloons in Bacteria: Gas Vesicles

Halobacterium salinarum (gvpA to gvpM): A variety of species of archaea and bacteria control their buoyancy by forming gas-filled balloons in their cytoplasm (Figure 1) [29, 30]. They function to maintain a desired depth in an aquatic environment, and their inflation and deflation are regulated by environmental signals that are used to identify the correct depth, including UV, light intensity, salinity, and oxygen. To have an effect on buoyancy, at least 10% of the cell volume needs to be composed of gas vesicles; thus, some bacteria have >10,000 per cell depending on their size. These vesicles are large, ranging in size from 33 to 250 nm in diameter, but they have very thin walls – about 2 nm – made entirely of protein. The protein that is the major component of the wall is GpvA, which forms a hydrophobic inner surface that blocks the formation of water droplets via surface tension. Gas enters the vesicles via diffusion and it is not stored there, rather it is in equilibrium with the surrounding concentration of dissolved gas [30]. The filling time is 0.4 μs. The vesicles can be filled by small and large gas molecules, including O_2, N_2, H_2, CO_2, CO, CH_4, Ar, and C_4F_8. All of the genes necessary for gas vesicle formation occur together in a gene cluster, including the core structural proteins, accessory proteins that control the size and shape, chaperones, seeding proteins, and regulators (Figure 2). Vesicles are produced when the gene cluster from *Bacillus megaterium* is transferred into *Escherichia coli* [31].

Recombinant proteins can be inserted to an accessory protein (GvpC) that participates in the formation the shell of the gas vesicle. Antibodies have been produced and purified in this way, where the buoyancy of the vesicle aids purification and delivery [32]. Gas vesicles isolated from cyanobacteria have been shown to improve oxygen transport in a mammalian cell fermentation [33]. Very interestingly, vesicles have been discovered in terrestrial bacteria that have been studied for their diverse secondary metabolisms, including antibiotic production (such as *Streptomyces avermitilis*), although their role in these organisms has yet to be determined [34]. Perhaps the most interesting aspect of blowing up balloons in bacteria is that it is fun [35].

I.D. Metal Nanoparticles: Magnetosomes

Magnetospirillum magneticum AMB-1 (mamC to mamF): Biomineralization is a process by which bacteria build intricate three-dimensional nanostructures from dissolved metal. One of the most dramatic structures is the magnetosome, which is used by bacteria that can orient their swimming to align with the geomagnetic field (magnetotaxis) (Figure 1) [36]. Spherical nanocrystals ($d \sim$ 30-50 nm) of iron oxide magnetite (Fe_3O_4) are contained within lipid organelles [37]. A string of these crystals are held in a chain orientation by proteins that resemble those involved in the cytoskeleton. All of the genes that encode the proteins

involved in the biosynthesis, organization, and regulation of the magnetosome are encoded in a conserved Magnetosome Island (MAI) region of the genome that spans 80-150 kb (Figure 2). This region has been noted to be unstable, where it is lost frequently in lab culture. MAIs are phylogenetically widespread and there is diversity in the size, shape, and mineral composition of the crystals [37].

The process by which magnetosome crystals are grown produces remarkably uniform and highly ordered structures [36]. The bacteria also have the ability to concentrate low environmental abundances of metals and form crystals under mild conditions compared to chemical routes [38]. Further, their synthesis in a lipid membrane makes them easily dispersed in an aqueous environment. Thus, magnetosome-producing cells have the potential to be harnessed for the industrial production of magnetic nanoparticles, with diverse applications in medicine, imaging, and commercial uses in magnetic tape and ink [38, 39]. Beyond Fe-based magnetosomes, many bacteria are able to build structures of diverse size (1-6000 nm), complexity (spherical, triangular, octahedral, decahedral, cubic, plates), and out of a variety of materials (gold, silver, cadmium, palladium, selenium, titanium, lanthanum, zinc, uranium, lead) [40].

II. Scavenging for Nutrients

II.A. Molecular Chainsaws: The Cellulosome

Clostridium cellulolyticum (cipC to cel5N): Cellulose is a polymer of sugar molecules and is an abundant component of plant cell walls. It is the most abundant polymer in biomass and many organisms have the ability to degrade this material to obtain carbon and energy. They do this by secreting cellulases and other enzymes that are able to break down complex cell walls to release simpler sugars that can diffuse into the cell. In clostridia and rumen organisms, these enzymes frequently cluster to form a large cellulosome that protrudes from the cell surface (Figure 1) [41]. The cellulosome consists of a scaffolding protein (scaffoldin) that contains a series of cohesin domains. These domains bind to dockarin domains at the N-termini of the cellulases that are involved in the assembly of the cellulosome. The scaffoldin is tethered to the cell surface (often to the S layer) and it and many of the cellulases contain cellulose binding domains so that they bind to the cellulose in the plant cell wall [42]. The advantage of having a cellulosome has been postulated to be that there is a higher likelihood that the released sugars will be consumed by the organism [41]. There is much diversity in the size and composition of cellulosomes and additional enzyme activities are commonly present, including hemicellulases and pectins to aid the decomposition of the plant cell wall. The cellulosomes can be very large, with up to 200 enzymes, and can be up to 16MDa [42, 43]. Within species, the cohesin-dockarin interactions are not specific and individual enzymes are not discriminated at each position in the scaffoldin. Remarkably, the enzyme composition of the cellulosome reflects the substrate on which the cells were grown; for example, on grass clippings, pectins are expressed, and on sewage soils consisting of insect biomass, chitinases are expressed [42]. The genetic regulation that controls this adaptation is unknown.

The scaffolding protein and many of the enzymes are often organized into a gene cluster (Figure 2). The cluster in *C. cellolyticum* is a model system with 12 genes encoded in a 26kb region, and the pattern of phylogenetic distribution implies horizontal transfer [42]. CipC is the scaffoldin and there are 8 cellulases, 1 hemicellulase (Man5K), and 1 pectinase (Cc-Rgl11Y) present in the cluster. The breakdown of crystalline cellulose is complex and requires synergistic activities between multiple enzymes. Some enzymes cut the polymer at random locations (exoglucanase), whereas others start from either free reducing or non-reducing ends and progressively cleave the polymer (endoglucanase). The most prevalent enzymes of this cellulosome represent one from each category: Cell9E and Cell48F [42].

Cell9E randomly cuts the cellulose strands. In contrast, Cell48F forms a long hydrophobic tunnel, through which cellulose strands are threaded as it progressively cuts and releases simpler sugars. *C. cellulolyticum* contains 62 enzymes that contain dockarin domains and the enzyme composition has been shown to vary based on the substrate on which bacteria grow [44].

Biofuels and renewable chemicals require a source of carbon. Currently, this most frequently is obtained in the form of sugar from crops such as corn and sugarcane. Particularly for fuels, this poses sustainability problems as agricultural crops would be diverted from food. In contrast, the amount of carbon that could be extracted from biomass is on the same scale as the fuels industry (and microbes have been estimated to naturally release the equivalent of 640 billion barrels of crude per year!) [41]. The problem is in the efficient liberation of carbon from cellulosic feedstocks, and this is currently a very active area of research. Significant effort has been put into the engineering of the cellulosome for this purpose [45]. One approach is to metabolically engineer natural cellulosic microbes to produce valuable products. A problem with this approach is that organisms containing cellulosomes are often not adapted for high carbon fluxes [42]. A variety of metabolic engineering approaches have been used to increase the catabolism of these organisms. Another approach is to move the cellulosome into a noncellulolytic organism. This has been achieved for moving a minimal cellulosome to *C. acetobutylicum* [46] and the yeast *S. cerevisiae* [47]. Finally, the adhesin-dockarin domains have been harnessed as modules that control protein-protein interections for a variety of applications outside of bioenergy, including protein purification and display [48].

II.B. Eating Oil: Alkane Degradation Pathways

Pseudomonas putida GPO1 (alkB to alkS): Numerous marine and terrestrial bacteria have the ability to utilize hydrocarbons as a carbon and energy source [49]. When oil leaks into seawater, this leads to the growth of a bloom of bacteria that are obligate consumers of alkanes [50]. Many of the genes involved in the utilization of hydrocarbons occur together in a gene cluster [51]. Petroleum is a chemically-diverse substance and there are a range of enzymes and related pathways that break down different classes of molecules [51, 52]. The gene cluster in *P. putida* is one of the most well studied systems and is able to degrade medium-length alkanes (Figure 2) [51]. The metabolic pathway begins with an alkane hydroxylase (AlkB – a membrane-associated non-heme diiron monooxygenase), which converts the alkane to an alcohol (Figure 3) [52]. Often, strains contain multiple alkane hydroxylases in order to broaden the range of substrates that can be consumed [49]. Electrons are delivered to AlkB by two rubredoxins (AlkF and AlkG). The alcohol is converted to acyl-CoA in three steps (AlkHJK), at which point it can enter metabolism. Two additional proteins, AlkL and AlkN, putatively encode an importer and chemotaxis sensory protein, respectively. AlkS acts as an alkane sensor and upregulates gene expression. The *alk* gene cluster occurs in many phylogenetically distinct bacteria [49]. It has a lower G+C content than the genome and is flanked by transposon genes, which indicate frequent horizontal transfer.

Petroleum degrading organisms have been proposed to be used in a wide variety of industrial applications. This includes a variety of potential roles in environmental clean-up, from biosensing and site evaluation to environmental dispersal, fermenter-based waste treatment, refinery waste treatment, and tanker ballast cleaning [49]. Organisms and related pathways have been identified that can break down nearly all of the components of petroleum, including benzene, ethylbenzene, trimethylbenzene, toluene, ethyltoluene, xylene, naphthalene, methylnapthalene, phenanthrene, C_6-C_8 alkanes, C_{14}-C_{20} alkanes, branched alkanes, and cymene [49]. In addition, alkane-degrading organisms could be used

as biocatalysts to add value to petroleum products [49]. For example, *Alcanivorax* has been engineered to direct the carbon flux from alkanes to the production and export of the bioplastic precursor PHA [53]. A particularly interesting use is for microbial enhanced oil recovery (MEOR), where bacteria are introduced into oil wells to facilitate secondary recovery [49]. The injection of oil-degrading organisms can increase recovery by reducing viscosity or secreting surfactants. MEOR has been tested worldwide, including the U.S., and has led to increases of 15-23% for oil wells in Japan and China [49]. Finally, the alkane-sensing transcription factor (AlkS) and the AlkB promoter have been transferred into *E. coli* in order to construct a genetic biosensor [54].

II.C. Fertilizer Factories: Nitrogen Fixation

K. pneumoniae (nifJ to nifQ): The availability of nitrogen limits the growth of many organisms [55]. In agriculture, fixed nitrogen is a critical component of fertilizer and its availability has been linked to the growth of the human population. The primary source of nitrogen is from the atmosphere in the form of N_2. Converting this into a form that can enter metabolism – such as ammonia (NH_3) – is a difficult chemical reaction. The Haber-Bosch process can chemically convert N_2 to NH_3 using high temperatures and pressures using an iron catalyst. In contrast, biological nitrogen fixation uses a complex enzyme (nitrogenase) to perform this reaction (Figure 3). Remarkably, the current flux of fixed nitrogen from synthetic chemical and natural biological processes is about equal [55].

Only prokaryotes and some archaea have the ability to fix nitrogen [56]. Often, all of the genes necessary for nitrogen fixation are encoded in a gene cluster. One of the simplest and most well studied clusters is from *Klebsiella pneumoniae*, which consists of 20 genes encoded in 23 kb (Figure 2) [57]. These genes encode all of the necessary components for nitrogen fixation, including the nitrogenase, a metabolic pathway for the synthesis of metal co-factors, e^- transport, and a regulatory network. Nitrogenase consists of two core proteins (NifH and the NifDK complex) that participate in a reaction cycle [55]. The reaction itself is very energy and redox intensive with the balance, $N_2 + 8e^- + 16ATP + 8H^+ \rightarrow 2NH^3 + 16ADP + 16Pi + H_2$.

Each reaction cycle consists of the transfer of 1 e^- and the consumption of 2 ATP (the energy of which is used to greatly accelerate e^- transfer). It is marked by a transient interaction between NifH, which receives an e^- from a variety of sources, and NifDK, which contains the reaction center where N_2 binds and fixation occurs. The cycle of binding, electron transfer, and dissociation needs to be repeated 8 times to fix a single N_2 molecule. Nitrogenase is slow ($k_{cat} = 5\ s^{-1}$) and is thought to be limited by the dissociation step [55]. Three co-factors form the core of the e- transfer and catalysis: [Fe_4-S_4] in NifH, the P-cluster [Fe_8-S_7] in NifDK, and FeMo-co [Mo-Fe_7-S_9-X] (Figure 4) where the reaction occurs [57]. The enzymes involved in the synthesis of these co-factors and chaperones for their incorporation to form mature nitrogenase make up the majority of the cluster (Figure 3). It has been proposed that these proteins all form a macromolecular “biosynthetic factory” centered on the NifEN proteins [57]. NifF and NifJ are a flavodoxins that feed electrons to NifH, with pyruvate as one source [58]. Nitrogenase is extremely oxygen sensitive and expensive for the cells to make and run [59]. A simple regulatory cascade is formed by the activator NifA and the anti-activator NifL, which integrate signals to ensure that the genes are only expressed in the absence of oxygen and fixed nitrogen [56].

Since the earliest tools in genetic engineering were developed, it has been a dream of biotechnology to create cereal crops that can fix their own nitrogen. The complexity of the nitrogen fixation pathway and a lack of efficient tools for modifying non-model plants have hindered progress in this area [60]. In contrast, the complete gene cluster was functionally transferred from *Klebsiella* to *E. coli* relatively early in 1972 [7]. The choloroplast may be a

potentially good target for the maturation and function of nitrogenase because: 1. it is where ammonia assimilation occurs, 2. ATP is generated there, 3. there is evidence that the ancillary proteins for Fe-S formation exist, and 4. the genetic context is similar to a prokaryote, including the ability to transcribe operons [61]. Individual genes from the pathway have been transferred to the Tobacco genome with a chloroplast-targeting peptide and to the plastid in the algae *Chlamydomonas reinhardtii* [60]. Neither of these efforts yielded appreciable expression. Besides the difficulty of expressing *Klebsiella*-encoded genes in these contexts, there are regulatory issues around the oxygen sensitivity of nitrogenase and, thus, its inconsistency with the photosynthetic processes in the chloroplast. One way to overcome this would be to place the nitrogenase under the control of light or oxygen sensitive transcription factors [60], or to express oxygen-protective factors, including the "Shethna" protein from *Azotobacter vinelandii* [62] and some nitrogenases are intrinsically less sensitive to oxygen [61]. Several applications of nitrogenase have been proposed that do not involve moving the system into a plant, including for the use of N_2 as a cheap source of nitrogen during fermentation [63], cyanide detoxification [64], the use of bacteria as biofertilizer [65], and for the industrial production of ammonia [66].

II.C. Bioremediation: Polychlorinated biphenyl degradation

Burkholderia xenovorans LB400 (orf0 to bphD): Some bacteria can use harmful organic pollutants as their sole source of carbon and energy [67]. For example, *Burkholderia xenovorans* LB400 can subsist on polychlorinated biphenyls (PCBs), which are used industrially as, *inter alia*, fire retardants and plasticizers (Figure 4) [68]. This capability has made *B. xenovorans* and other PCB metabolizing bacteria key elements of bioremediation strategies for chemical spills. Highly chlorinated PCBs are reductively dehalogenated by organisms such as *Dehalococcoides*, which can use PCBs as a terminal electron acceptor for anaerobic respiration [68]. These lower chlorinated PCBs are the substrate for the *B. xenovorans* degradation pathway, which consists of a series of enzyme-mediated oxidations culminating in the cleavage of one of the linked aromatic rings by the ring-opening dioxygenase BphC. The cleaved ring is converted to two equivalents of acetate in a three-step pathway, while the uncleaved ring is released as benzoic acid and then further processed to catechol by the protein products of the *benABCD* gene cluster (Figure 2) [68].

Several strategies are being employed to increase the number of PCBs that can be degraded microbially, including directed evolution of a ring-cleaving dioxygenase [69] and functional screening of metagenomic libraries from activated sludge [70]. Future efforts may attempt to introduce PCB degradation gene clusters into bacterial strains that synthesize compounds of industrial value, which would allow these strains to consume feedstocks that would otherwise require expensive and environmentally unfriendly disposal.

III. Biosynthesis of Chemicals

III.A. Bioplastic biosynthesis: Poly(3-hydroxybutryrate)

Ralstonia eutropha H16 (phbA to phbC): Many bacteria synthesize poly(3-hydroxybutyrate) (PHB) and other poly(hydroxyalkanoates) (PHAs) as a means of storing carbon and energy intracellularly (Figure 4). The biosynthetic pathway for PHB, exemplified by the *phb* gene cluster in *Ralstonia eutropha* (Figure 2) [71], consists of three steps: PhbA catalyzes a Claisen condensation to convert two molecules of acetyl-CoA to acetoacetyl-CoA, PhbB reduces acetoacetyl-CoA to 3-hydroxybutryl-CoA, and PhbC polymerizes 3-hydroxybutryl-CoA with release of CoA to form PHB [72]. PHB is hydrophobic and accumulates in cytoplasmic granules.

PHB and other PHAs are versatile bioplastics; biodegradable forms of a diverse set of products ranging from plastic bottles to golf tees are produced commercially from

bacterially synthesized PHAs [72]. Efforts to metabolically engineer the synthesis of bioplastics are proceeding along two tracks. First, the genes for the production of PHB and other PHAs have been introduced into plants in order to realize the benefits of using CO_2 as a carbon source rather than fermentation feedstocks [73]. However, these efforts have been only modestly successful; to date, the best PHA production titer seen in plants is only ∼10% of dry weight. Second, a variety of engineering efforts including genetic engineering and the provision of unnatural substrate derivatives in the fermentation broth have led to the optimization of PHA yields in native and engineered hosts and the production of novel PHA derivatives [74].

III.B. Nonribosomal peptide biosynthesis: Echinomycin

Streptomyces lasaliensis (ecm1 *to* ecm18): Nonribosomal peptides (NRPs) are a class of peptidic small molecules that includes the antibiotic vancomycin and the immunosuppressant cyclosporine [75]. The gene cluster for echinomycin (Figure 4), a DNA-damaging NRP from the quinoxaline class, is typical in encoding four categories of gene products (Figure 2): 1) Genes for miniature, self-contained metabolic pathways that provide unusual monomers. Eight *ecm*-encoded enzymes convert tryptophan into quinoxaline-2-carboxylic acid (QC), an unusual monomer that enables echinomycin to intercalate between DNA base pairs. 2) Genes for an assembly-line-like enzyme known as an NRP synthetase (NRPS) that link monomers (typically amino acids) into a peptide and then release it from covalent linkage to the assembly line, often with concomitant macrocyclization. The *ecm* gene cluster encodes two NRPS enzymes, Ecm6 (2608 amino acids) and Ecm7 (3135 amino acids), that convert QC, serine, alanine, cysteine, and valine into a cyclic, dimeric decapeptidolactone. 3) Genes for chemical ‘tailoring’ after release from the NRPS. Two *ecm*-encoded enzymes oxidatively fuse the two cysteine sidechains into a thioacetal. 4) Genes that encode regulatory and resistance functions. Transporters are also commonly found in NRP gene clusters [76].

There are two ways in which synthetic biology is being used in the area of NRPS engineering. First, efforts are being made to express NRPS gene clusters in heterologous hosts, either in their native form [77] or as refactored gene clusters in *E. coli* [76]. Expression in a heterologous host can serve three purposes: making the encoded NRP accessible for structure elucidation or biological characterization, particularly useful if the native host is unknown or unculturable; making the genes easier to manipulate, which is useful if the native host is not amenable to genetics; and improving the production titer of its small molecule product, which is helpful if the gene cluster is repressed by an external regulatory system in the native host. The production of NRP derivatives has been engineered by replacing portions of NRPS genes with variants from other gene clusters that lead to the incorporation of alternative amino acid building blocks. This technique has been used most extensively to generate derivatives of the NRP antibiotic daptomycin [78].

III.C. Polyketide biosynthesis: Erythromycin

Saccharopolyspora erythraea *NRRL 2338 (SACE_0712 to* eryCI): Polyketides (PKs) are a class of acetate- and propionate-derived small molecules that includes the immunosuppressant FK506, the antibiotic tetracycline, the cholesterol-lowering agent lovastatin, and a number of rapamycin analogues made by genetic engineering are in clinical trials [75]. The biosynthetic pathways for PKs and fatty acids are similar in their chemical logic and use related enzymes: both involve the polymerization of acetate- or propionate-derived monomers by a series of Claisen condensations followed by reduction of the resulting ß-ketothioester [75]. The gene cluster for erythromycin (Figure 4), an antibacterial PK from the macrolide class, encodes the following classes of gene products (Figure 2): 1) Three large PK synthase (PKS) enzymes – DEBS 1 (3545 amino acids), DEBS 2 (3567

amino acids), and DEBS 3 (3171 amino acids) – that convert seven equivalents of the propionate-derived monomer methylmalonyl-CoA into the intermediate 6-deoxyerythronolide B (6-DEB). 2) Two P450s that hydroxylate the nascent scaffold. 3) Twelve enzymes that synthesize the unusual sugars desosamine and mycarose from glucose and attach them to the nascent scaffold. Without these sugars, erythromycin does not have appreciable antibiotic activity. 4) An erythromycin resistance gene that modifies the 50S subunit of the ribosome to prevent erythromycin from binding [79].

Many PKSs have been expressed in heterologous hosts such as *E. coli*, including the PKSs for erythromycin and the anticancer agent epothilone [80]. Another notable heterologous host is a variant of the *Streptomyces fradiae* strain used for the industrial production of the antibiotic tylosin; having gone through many rounds of classical strain improvement, the metabolism of this strain is well suited to the production of PKs. A variant of the strain was created in which the tylosin gene cluster was replaced by the erythromycin PKS yielding a high titer of the non-native product [81]. The PKS genes have been mutated or replaced with variants from other gene clusters to generate PK derivatives [82], or to create custom PKSs that synthesize small PK fragments by assembling portions of several PKS genes [83].

III.D. Terpenoid biosynthesis: Lycopene

Rhodobacter capsulatus (crtE *to* crtY): Terpenoids are a class of molecules that include the anticancer agent taxol, the antibiotic pleuromutilin, and the carotenoid pigments. While terpenoids are more common among plants than bacteria [84], carotenoids are produced by a range of bacteria. Lycopene and other carotenoids are generally used in one of two ways: to harvest light (either for energy or photoprotection) or as antioxidants (Figure 4). As with other terpenoids, the first step in the biosynthetic pathway for lycopene is the CrtE-catalyzed polymerization of the C_5 monomer isopentenyl pyrophosphate (IPP) or its Δ^2 isomer dimethallyl pyrophosphate (DMAPP), in this case to the C_{20} polymer geranylgeranyl diphosphate (GGDP). CrtB then dimerizes two equivalents of GGDP in a tail-to-tail fashion, resulting in the formation of the linear C_{40} polymer phytoene. CrtI catalyzes four successive desaturations to yield lycopene. Alternative products such as beta-carotene are formed by the action of CrtY, which cyclizes the termini of the linear polymer [85]. All of the genes in this pathway occur together in a cluster (Figure 2).

The colored nature of carotenoids has enabled their pathways to be engineered by genetic screens with colony color phenotypes. For example, a library of shuffled phytoene desaturases was screened in an *E. coli* strain harboring the *crt* gene cluster, resulting in the identification of desaturase clones that enabled the production of two lycopene variants, 3,4,3′,4′-tetradehydrolycopene and torulene [86]. Much synthetic biological work has been done by Keasling and coworkers on the production of plant terpenes (e.g., artemisinin) in the microbial hosts *S. cerevisiae* [87] and *E. coli* [88]. This effort has involved two key challenges. First, since biosynthetic genes are not physically clustered in plant genomes, identifying the genes involved in terpenoid biosynthesis has been difficult, although the ongoing projects to sequence the genomes of hundreds of plants should enable bioinformatic efforts to identify biosynthetic genes. Second, the metabolism of *S. cerevisiae* and *E. coli* has been optimized for the production of terpenoids by increasing the flux of carbon toward IPP and DMAPP; in *E. coli* this was accomplished by supplementing the endogenous IPP biosynthetic pathway with the one from *S. cerevisiae*.

III.E. Oligosaccharide biosynthesis: Xanthan

Xanthomonas campestris *pv. campestris* (gumB *to* gumM): Every year, 10,000-20,000 tons of xanthan are produced for use in foods (e.g., to control the crystallization of ice cream and to emulsify salad dressings) and in industry (e.g., to modulate the viscosity of explosives

and laundry detergents) [89]. Xanthan, an oligosaccharide produced by the plant pathogen *Xanthomonas campestris*, is composed of a cellulose backbone, on alternating sugars of which a mannose-ß-1,4-glucuronate-ß-1,2-mannose trisaccharide is appended (Figure 4). A portion of the terminal mannoses have pyruvate linked as a ketal to the 4′- and 6′-hydroxyls, and some of the internal mannoses are acetylated on the 6′-hydroxyl. Owing to the glucuronate units and pyruvoyl substituents, xanthan is an acidic polymer. Xanthan biosynthesis involves the action of five glycosyltransferases (GumDMHKI), and the growing chain is anchored on undecaprenyl pyrophosphate, similarly to peptidoglycan biosynthesis (Figure 2). Three tailoring enzymes (GumFGL) add the aforementioned pyruvoyl and acetyl substitutents, and GumBCE are required for xanthan export [89].

Future synthetic biology efforts are likely to proceed along two tracks. First, while substrate to xanthan conversion rates of 60-70% have been achieved [89], *X. campestris* could be engineered to grow on cheaper feedstocks or to make the separation of the cells from the xanthan less costly; alternatively, the *gum* gene cluster could be moved to an alternative host. Second, changes to the structure of xanthan have important effects on its rheological properties. Efforts to use genetic engineering to alter the structure (and therefore the rheological properties) of xanthan – or of other microbial exopolysaccharides such as alginate or gellan [90] – have the potential to create new polymers with, e.g., altered viscosity and shear stability.

III.F. Indolocarbazole biosynthesis: Staurosporine

Streptomyces sp. *TP-A0274* (staR *to* staMB): Indolocarbazoles are natural products formed by the oxidative fusion of primary metabolic monomers [84]. Staurosporine, an indolocarbazole, is a promiscuous, nanomolar inhibitor of serine/threonine protein kinases that binds in an ATP-competitive manner to these enzymes (Figure 4) [91, 92]. The staurosporine gene cluster encodes three categories of gene products (Figure 2): 1) Four oxidoreductases (two P450s and two flavoenzymes) that catalyze a net 10-electron oxidation to fuse two molecules of tryptophan into the indolocarbazole aglycone [93]. 2) Nine enzymes to synthesize and attach an unusual hexose to the indolocarbazole scaffold at the indole nitrogens. 3) A transcriptional activator that regulates the expression of the gene cluster. Other naturally occurring indolocarbazoles differ in the oxidation state of the indolocarbazole scaffold, the derivatization of the indole ring by chlorination, and the sugar substituent appended to the indolocarbazole aglycone.

More than 50 unnatural indolocarbazole derivatives have been made by assembling artificial gene clusters in a non-native host [94]. These molecules harbor chemical modifications that would be difficult to introduce by semisynthetic derivatization of naturally occurring indolocarbazoles or by total synthesis. The majority of these efforts have used genes from the gene clusters for indolocarbazoles as the building blocks for the artificial gene clusters. Future efforts to explore the activities of completely unrelated enzymes (e.g., ring-opening dioxygenases) may enable the modification of portions of the indolocarbazole scaffold – such as the external six-membered rings – that would be difficult to access using synthetic organic chemistry or enzymes from indolocarbazole gene clusters.

IV. Energy Generation and Transfer

IV.A. Solar Powerpacks: Photosynthetic Light Harvesting

Rhodobacter sphaeroides (pufH *to* pufX– *also include puc genes*): Sunlight is converted into power by the light harvesting system of anaerobic photosynthetic bacteria (Figure 1) [95]. Light energy is captured by two light harvesting complexes (LH1 and LH2) and is funneled to a reaction center (RC) [87]. The RC uses the energy to produce a transmembrane charge separation mediated by the reduction of a quinone. This ultimately

causes a proton to move from the cytoplasm to periplasm, which powers the production of ATP via the protonmotive force. The *R. sphaeroides* membrane has spherical invaginations that increase the surface area and number of RCs. Each RC occupies a hole in a ring formed by LH1. Light harvesting complex 2 also forms rings that surround the RC:LH1 complex in the membrane. The LH complexes use carotenoids and bacteriochlorophyll (Figure 4) to absorb green and near-infrared light, respectively [87]. The photosynthetic genes are frequently found in a single cluster in purple bacteria [96, 97]. In *R. sphaeroides*, the photosynthesis gene cluster is 40.7 kb long and contains all of the necessary genes for the formation of the reaction center/light harvesting complex 1 (*puf* genes), and light harvesting complex 2 (*puc* genes) (Figure 2). Two biosynthetic pathways make up the bulk of the cluster, where bacteriochlorophyll is produced from heme in a 16 gene pathway (*bch* genes) and the carotenoid sphaeroidine is produced from isopentenyl pyrophosphate in a 7 gene pathway (*crt* genes). The gene cluster is regulated by oxygen concentration, as well as the light intensity and color [87].

The *Rhodobacter* light harvesting system has been model system for studying photosynthesis. It is relatively simple, there is only one photosystem, and the organism is genetically tractable. This has enabled detailed quantum mechanical measurements to be made on light absorption and electron transfer [87], which may enable the design of next-generation “biologically inspired” photovoltaic cells. Going one step further, the light harvesting complex can be functionally reconstituted *in vitro* and this has led to the construction of various hybrid systems, where the electrons are shuttled to inorganic materials [98]. A particularly interesting approach is the development of a self-assembled monolayer, where the reaction center is tethered to a metal (Pt, Hg or Au) surface by an organic molecule that ends with a quinone. Multi-layer films have constructed and shown to efficiently capture electrons. *Rhodobacter* has also been harnessed for producing H_2 from light for use in fuel cells or microbial fuel cells [99].

IV.B. Nanowires: Conductive Surface Pili

Geobacter sulfurreducens (pilB *to* pilA): Metal-reducing bacteria are able to discharge electrons to solid surfaces through “wires” formed by pili that extrude from the cell surface (Figure 1) [100]. This gives these bacteria the required terminal electron acceptor for oxidative phosphorylation in the absence of other dissolved acceptors (oxygen, nitrate, sulfate, etc) [101]. *Geobacter* is able to form pili that attach to Fe(III) oxide surfaces. The genes that form the pilus are encoded in a gene cluster, with PilA being the pilin subunit that is homologous with the Type IV pili from other organisms (Figure 2) [100]. The pili have a diameter of 50 nm and can extend up to 20 μm from the cell surface [102]. High electron transfer rates of 10^{11}/s through the pili have been observed [101]. The nanowires can also connect multiple bacteria, implying that a community of cells can be wired for rapid electron transfer (Figure 1) [76]. Beyond *Geobacter*, many other species have been shown to produce conductive pili in response to electron-acceptor limitation [76].

Microbial fuel cells have emerged as a potential source of alternative energy [76]. A microbial fuel cell involves bacteria that are sequestered such that the only mechanism of electron transfer during respiration is to a graphite or gold anodes [103]. The closest applications for microbial fuel cells are for long-term sensors deployed in the ocean and in wastewater treatment. Ocean sensors make use of the natural generation of electrical currents in the sea floor for power [102]. Electricity can also be recovered from wastewater treament as a byproduct of the breakdown of biomass. To deliver electricity to the anode, either a dissolved electron carrier or pili nanowires are required. *Geobacter* is one of the most efficient electron donors where it forms 50 μm thick biofilms on the surface [102]. Pili are critical for electron transfer through the biofilm [103].

V. Environmental Sensing and Signal Processing

V.A. Physical Integrated Circuits: The Stressosome

Bacillus subtilis (rsbR *to* rsbX): One goal of synthetic biology is to build genetic circuits that can integrate information from environmental sensors or produce a dynamic response. To date, connecting circuits to form a program involves building a cascade at the level of transcription or translation. Each layer of the cascade requires about 20 minutes to complete [104]. More complex operations that require multiple layers can be particularly slow in the propagation of the signal. It would be much faster if the circuitry could be built as a molecular machine, where signals are directly received and signal integration occurs due to conformational changes or signal propagation (*e.g.*, via a phosphorelay).

Bacteria have such a machine, known as the stressosome [105]. In *B. subtilis*, it is in the gene cluster that contains many regulatory factors that converge on the anti-sigma factor σ_B, which controls the general stress response. This pathway integrates energy and environmental stresses through a complex parter-switching mechanism (Figure 5) involving anti-sigma factors, anti-anti-sigma factors, kinases, and phosphatases, most of which occur together in a gene cluster (Figure 2). Three of these proteins (RsbRST) form the structure of the stressosome, which is as large as a ribosome (1.8 MDa, 300Å diameter) and appears as a spiked ball, with a core and protrusions (Figure 1) [106]. The protrusions are composed of RsbR as well as four RsbR paralogs [105]. The N-terminal domains of these proteins are variable, leading to the hypothesis that they act like sensors to receive diverse signals, including small molecules, protein-protein interactions, and even light [106, 107]. These signals are integrated by the core of the stressosome through conformational changes or other biochemical mechanisms. The RbsT protein interacts with the stressosome and transmits the signal to the σ_B pathway. The stressosome is induced by environmental stress and the release of RsbT is highly cooperative (a Hill factor of $n = 8$). About 20 stressosomes are present in a single *B. subtilis* cell and they are closely associated with nucleoids and exhibit little diffusion. The σ_B gene cluster contains a number of other regulatory proteins that participate in partner-swapping and kinase/phosphatase interactions and there are internal transcriptional positive and negative feedback loops (Figure 5). Together, this regulation ensures that the response is: 1. transient with a 30 minute pulse of activity, 2. fast, 3. graded and 4. the magnitude matches the degree of stress [107, 108].

The stressosome and the σ_B stress response pathway have several applications in biotechnology. The core proteins of the stressosome (RsbRST) are present in many species, including Gram negatives, and are associated with a variety of regulatory mechanisms, including aerotaxis, two-component sensors, and the biosynthesis of signaling molecules [105, 107]. It may be that this structure is a common mechanism by which signals are integrated and understanding how to “reprogram” this structure would potentially enable much faster signal integration than transcriptional circuitry. In *B. subtilis*, there is evidence that five signals are integrated, but given the size of the structure, many more may be possible. The complete σ_B gene cluster is much less distributed than the stressosome. Several industrially-relevant strains contain σ_B and related gene clusters, including *Bacillus* and *Streptomyces* [107]. The general stress response mediated by σ_B is involved with a number of stresses that are relevant to biotechnology, including response to shifts in salt concentration, pH, ethanol, ATP, cell wall stress, and UV light [106]. These types of stresses are common as the result of product accumulation, shifts in growth phase, and occur in different microenvironments in a bioreactor. Understanding how to rapidly integrate these signals would enable the construction of cell controllers that could regulate metabolic flux based on the changing environmental conditions of a fermentation.

V.B. An Immune System against Phage: CRISPR Arrays

Escherichia coli (cas3 *to* *CRISPR4*): Many bacteria and archea contain an "immune system" that recognizes and intercepts foreign DNA based on previous exposure [109, 110]. This improves resistance against phage and the conjugative transfer of plasmid (Figure 5). This function is encoded by a clustered, regularly spaced short palindromic repeat (CRISPR) region that occurs next to a gene cluster (Figure 2). Each repeat spacer in the CRISPR region represents a DNA sequence of a phage or plasmid to which the bacteria have been exposed. The region is actively reprogrammed to respond to new challenges, which lead to the extension of the CRISPR region. Each repeat consists of a ∼31 bp region of the targeted DNA and up to 374 repeats in a single cluster [110]. Several genes are encoded in the cluster which form the Cas complex, which performs the tasks for the insertion of new spacer repeats and for the destruction of foreign DNA. The whole Cas-CRISPR cluster has undergone frequent horizontal transfer and some organisms have multiple clusters. An extreme example is *Methanocaldoccus jannaschii*, which contains 18 complete clusters with a total of 1188 repeat elements [110].

CRISPR operates as an immune system by incorporating foreign DNA as a new repeat spacer and then recognizing this sequence in foreign DNA and destroying it (Figure 5) [109]. The Cas complex cleaves foreign DNA and integrates it into the CRISPR region. The spacers are then transcribed together and processed into individual CRISPR RNAs (crRNAs), which then associate with the Cas complex. The Cas-crRNA complex then recognizes the sequences in foreign DNA based on the crRNA sequence and the DNA is directed for degradation.

Bacteriophage are relevant in biotechnology as they are notorious for disrupting fermentations involving bacteria [111, 112]. Traditionally, this is done through a process of "phage immunization" where resistant bacteria are identified through serial dilution of surviving cells [111]. In one industrial example, dairy starter cultures of *S. thermophilus* were isolated and are now in use where phage resistance is conferred by CRISPR [110]. Bacteriophage have also been harnessed in biotechnology as agents for self-organization in the construction of materials [113]and to be used as antibiotics [114, 115]. For the latter, one of the issues that arises is rapid resistance that arises in the bacteria. In a study where bacteriophage were used to treat tooth decay, resistance in *Steptococcus mutans* occurred due to CRISPR immunity [116]. The use of CRISPR elements has also been proposed as a mechanism to block the transfer of plasmids that confer antibiotic resistance and the horizontal transfer of pathogenicity islands that confer virulence [109, 110].

V.C. Smelling Bacteria: Quorum Sensing Receiver Clusters

Photorhabdus luminescens (yhfS *to* rafZ): A nemotode (*Heterorhabditis bacteriophora*) implements biological warfare on its insect targets [117]. It has developed a symbiotic relationship with the bacterium *P. luminescens* to attack and digest a wide range of insects. The genome of the bacterium is filled with more toxins and virulence factors than any known organism [118]. After the nemotode invades, the bacteria are released into the bloodstream where they kill and breakdown the host. *P. luminescens* is filled with gene clusters that encode multiple type I pili for adhesion, a type III secretion system, many toxin and virulence factor pathways, and 22 clusters that encode PKSs and NRPSs. These produce a variety of small molecules, including antibiotics that kill other bacteria that may compete for the dead insect's nutritional resources.

Interestingly, the genome contains clusters of homologues to LuxR homologues – 32 divided into two major clusters (Figure 2) [118]. In *Vibrio fischeri*, LuxR and LuxI participate in a quorum sensing circuit that enables bacteria to communicate [119]. LuxI is

an enzyme that produces a small molecule (AI-1) that freely diffuses through the membrane. AI-1 accumulates and, when a threshold is crossed, it activates the response regulator LuxR. Many bacteria contain multiple orthogonal pairs of LuxI/LuxR homologues, which forms a sort of “language” by which cells can communicate (Figure 5). Oddly, *P. luminescens* has many LuxR homologues, but no corresponding LuxI homologues [118]. It has been postulated that the cluster of LuxR proteins may be there to sense many of the bacteria that would be competing for the nutrients available in the diseased insect. In addition, some of these sensors may have evolved to respond to host hormones, such as insect juvenile hormone, to determine the identity and developmental state of the insect [120].

It is useful to be able to program communication between cells for a variety of applications in biotechnology [121]. Quorum sensing provides a language by which this can be achieved, where each chemical signal represents a channel for communication [119]. The LuxI enzyme is a “sender device” that produces the signal and LuxR is a “receiver device” that responds to the signal [122]. When these devices are separated between cells, this enables cells to communicate. This has been used to program cells to form patterns, including bull's eyes (Figure 5) [122] and to implement an edge detection algorithm [123]. More direct applications in biotechnology have been proposed for quorum sensing, including controlling the density of a population of bacteria in a fermentor and killing cancer cells once an invading population crosses a threshold density [121, 124].

Conclusions

In this review, we have attempted to capture the diversity of functions that are encoded in gene clusters. All of these have many potential applications in biotechnology. Achieving this potential will require methods that enable the reliable re-engineering of clusters. To date, this has been challenging because of the size of clusters, the number of genetic parts that are involved, and the complexity of the genetic regulation. Recently, progress in genetic engineering has increased the scale of projects that are achievable. Synthetic biology has emerged as a field and libraries have been populated with genetic “parts” that can carefully control transcription and translation and “devices” that encode regulatory sensors and circuits. Methods have been developed for the rapid assembly of these parts on the scale of gene clusters, and methods exist to ultimately combine multiple clusters into a host. Together, these advances will enable the bottom-up assembly of synthetic gene clusters to simplify and optimize their function. Beyond functions that are naturally organized in clusters, it will be interesting if the design principles from studying these systems could be applied to other functions – such as the flagellum – to create similarly transferable units. This may move genetic engineering to an era of genome design where simplified gene clusters are combined in order to pull together functions from many diverse organisms in order to build a synthetic one.

References

1. Fischbach MA, Walsh CT, Clardy J. The evolution of gene collectives: How natural selection drives chemical innovation. Proc Natl Acad Sci U S A. 2008; 105:4601–4608. [PubMed: 18216259]
2. Lawrence JG, Roth JR. Selfish operons: Horizontal transfer may drive the evolution of gene clusters. Genetics. 1996; 143:1843–1860. [PubMed: 8844169]
3. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000; 405:299–304. [PubMed: 10830951]
4. Karlin S. Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. Trends in Microbiology. 2001; 9:335–343. [PubMed: 11435108]

5. Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chrisholm SW. Transfer of photosynthesis genes to and from *Prochlorococcus* virus. Proc Natl Acad Sci U S A. 2004; 101:11013–11018. [PubMed: 15256601]
6. Karaolis DKR, Somara S, Maneval DR Jr, Johnson JA, Kaper JB. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. Nature. 1999; 399:375–379. [PubMed: 10360577]
7. Dixon R, Postgate JR. Genetic transfer of nitrogen fixation from *Klebsiella pneunomia* to *Escherichia coli*. Nature. 1972; 237:102–103. [PubMed: 4555442]
8. Hansen-Wester I, Chakravortty D, Hensel M. Functional transfer of *Salmonella* pathogenicity island 2 to *Salmonella bongori* and *Escherichia coli*. Infection and Immunity. 2004; 72:2879–2888. [PubMed: 15102800]
9. Challis GL. Mining mircobial genomes for new natural products and biosynthetic pathways. Microbiology. 2008; 154:1555–1569. [PubMed: 18524911]
10. Scherlach K, Hertweck C. Triggering cryptic natural product biosynthesis in microorganisms. Organic & Biomolecular Chemistry. 2009; 7:1753–1760. [PubMed: 19590766]
11. Czar MJ, Anderson JC, Bader JS, Peccoud J. Gene synthesis demystified. Trends in Biotechnology. 2008; 27:63–73. [PubMed: 19111926]
12. Gibson DG, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science. 2010; 329:52–56. [PubMed: 20488990]
13. Voigt CA. Genetic parts to program bacteria. Curr Opin Biotechnol. 2006; 17:548–557. [PubMed: 16978856]
14. Gibson DG, Young L, Chuang RY, Venter JC, Hutchinson CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods. 2009; 6:343–345. [PubMed: 19363495]
15. Benders GA, et al. Cloning whole bacterial genomes in yeast. Nucleic Acids Research. 2010; 38:2558–2569. [PubMed: 20211840]
16. Szczepanowski R, Braun S, Riedel V, Schneiker S, Krahn I, Puhler A, Schulter A. Microbiology. 2005; 151:1095–1111. [PubMed: 15817778]
17. Cornelis GR. The type III secretion injectisome. Nature Reviews Microbiology. 2006; 4:811–825.
18. Schraidt O, Lefebre MD, Brunner MJ, Schmied WH, Schmidt A, Radics J, Mechtler K, Galan JE, Marlovits TC. Topology and organization of the *Salmonella typhimurium* type III secretion needle complex components. PLOS Pathogens. 2010; 6:e1000824. [PubMed: 20368966]
19. Temme K, Salis H, Tullman-Ercek D, Levskaya A, Hong SH, Voigt CA. Induction and relaxation dynamics of the regulatory network controlling the type III secretion system encoded within *Salmonella* pathogeneity island 1. J Mol Biol. 2008; 377:47–61. [PubMed: 18242639]
20. Galen JE, Pasetti MF, Tennant S, Ruiz-Olvera P, Sztein MB, Levine MM. *Salmonella enterica serovar Typhi* live vector vaccines finally come of age. Immunology and Cell Biology. 2009; 87:400–412. [PubMed: 19417771]
21. Pawelek JM, Sodi S, Chakraborty AK, Platt JT, Miller S, Holden DW, Hensel M, Low KB. *Salmonella* pathogenicity island-2 and anticancer activity in mice. Cancer Gene Therapy. 2002; 9:813–818. [PubMed: 12224021]
22. Widmaier DM, Tullman-Ercek D, Mirsky EA, Hill R, Govindarajan S, Minshull J, Voigt CA. Engineering the *Salmonella* type III secretion system to export spider silk monomers. Molecular Systems Biology. 2009; 5:309. [PubMed: 19756048]
23. Yeats TO, Crowley CS, Tanaka S. Bacterial microcompartment organelles: Protein shell structure and evolution. Annu Rev Biophys. 2010; 39:185–205. [PubMed: 20192762]
24. Kerfeld CA, Heinhorst S, Cannon GC. Bacterial Microcompartments. Annu Rev Microbiol. 2010; 64
25. Mukhopadhyay A, Redding AM, Rutherford BJ, Keasling JD. Importance of systems biology in engineering microbes for biofuel production. Current Opinion Biotechnology. 2008; 19:228–234.
26. Fan C, Cheng S, Liu Y, Escobar CM, Crowley CS, Jefferson RE, Yeates TO, Bobik TA. Short N-terminal sequences package proteins into bacterial microcompartments. Proc Natl Acad Sci USA. 2010; 107:7509–7514. [PubMed: 20308536]

27. Wackett LP, Frias JA, Seffernick JL, Sukovich DJ, Cameron SM. Genomic and biochemical studies demonstrating the absence of an alkane-producing phenotype in *Vibrio furnissii* M1. Appl Environ Microb. 2007; 73:7192–7198.
28. Seedorf H, Fricke WF, Veith B, Bruggemann H, Liesegang H, et al. The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. Proc Natl Acad Sci USA. 2008; 105:2128–2133. [PubMed: 18218779]
29. Pfeifer, F. Gas Vesicles of Archea and Bacteria. In: Shively, JM., editor. Complex Intracellular Structures in Prokaryotes. Springer-Verlag; Berlin: 2006. p. 115-140.
30. Walsby AE. Gas Vesicles. Microbiological Reviews. 1994; 58:94–144. [PubMed: 8177173]
31. Li N, Cannon M. Gas vesicle genes identified in Bacillus megaterium and functional expression in *Escherichia coli*. J Bacteriol. 1998; 180:2450–2458. [PubMed: 9573198]
32. Sremac M, Stuart ES. Recombinant gas vesicles from *Halobacterium* sp. displaying SIV peptides demonstrate biotechnology potential as a pathogen peptide delivery vehicle. BMC Biotechnology. 2008; 8:9. [PubMed: 18237432]
33. Sundararajan A, Ju L. Glutaraldehyde treatment of proteinaceous gas vesicles from cyanobacterium *Anabaena flosaquae*. Biotechnol Prog. 2000; 16:1124–1128. [PubMed: 11101344]
34. van Keulen G, Hopwood DA, Dijkhuizen L, Sawers RG. Gas vesicles in actinomyces: old buoys in novel habitats? Trends in Microbiology. 2005; 13:350–354. [PubMed: 15993071]
35. Endy D, Deese I, Wadey C. Adventures in Synthetic Biology. Nature. 2005; 438:449–453. [PubMed: 16306983]
36. Jogler C, Schuler D. Genomics, genetics, and cell biology of magnetosome formation. Annu Rev Microbiol. 2009; 63:501–521. [PubMed: 19575557]
37. Komeili A. Molecular mechanisms of magnetosome formation. Annu Rev Biochem. 2007; 76:351–366. [PubMed: 17371202]
38. Schuler D, Frankel RB. Bacterial magnetosomes: microbiology, biomineralization, and biotechnological applications. Appl Microbiol Biotechnol. 1999; 52:464–473. [PubMed: 10570793]
39. Matsunaga T, Suzuki T, Tanaka M, Arakaki A. Molecular analysis of magnetotactic bacteria and development of functional bacterial magnetic particles for nano-biotechnology. Trends in Biotechnology. 2007; 25:182–188. [PubMed: 17306901]
40. Korbekandi H, Iravani S, Abbasi S. Production of nanoparticles using organisms. Critical Reviews in Biotechnology. 2009; 29:279–306. [PubMed: 19929319]
41. Fontes MGA, Gilbert HJ. Cellulosomes: Highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu Rev Biochem. 2010; 79:655–681. [PubMed: 20373916]
42. Desvaux M. The cellulosome of *Clostridium cellulyticum*. Enzyme and Microbial Technology. 2005; 37:373–385.
43. Ding SY, Rincon MT, Lamed R, Martin JC, McCrae SI, et al. Cellusomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. J Bacteriol. 2001; 183:1945–1953. [PubMed: 11222592]
44. Blouzard JC, Coutinho PM, Fierobe HP, Henrissat B, Lignon S, Tardiff C, Pages S, de Philip P. Modulation of cellulosome composition in *Clostridium cellulyticum*: Adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses. Proteomics. 2010; 10:541–544. [PubMed: 20013800]
45. Bayer EA, Morag E, Lamed R. The cellulosome - a treasure-trove for biotechnology. Trends in Biotechnology. 1994; 12:379–385. [PubMed: 7765191]
46. Sabathe F, Soucaille P. Characterization of the CipA scaffolding protein and in vivo production of a minicellulosome in *Clostridium acetobutylicum*. J Bacteriol. 2003; 185:1092–1096. [PubMed: 12533485]
47. Lilly M, Fierobe HP, van Zyl WV, Volschenk H. Heterologous expression of a *Clostridium* minicellulosome in *Saccharomyces cerevisiae*. FEMS Yeast Res. 2009; 9:1236–1249. [PubMed: 19744245]
48. Nordon RE, Craig SJ, Foong FC. Molecular engineering of the cellulosome complex for affinity and bioenergy applications. Biotechnol Lett. 2009; 31:465–476. [PubMed: 19116695]

49. van Hamme JD, Singh A, Ward OP. Recent advances in petroleum microbiology. Microbiology and Molecular Biology Reviews. 2003; 67:503–549. [PubMed: 14665675]
50. Yakimov MM, Timmis KN, Golyshin PN. Obligate oil-degrading marine bacteria. Current Opinion Biotechnology. 2007; 18:257–266.
51. van Beilen JB, Panke S, Lucchini S, Franchini AG, Rothlisberger M, Wilholt B. Analysis of *Pseudomonas putida* alkane-degradation gene clusters and flanking insertion sequences: evolution and regulation of the alk genes. Microbiology. 2001; 147:1621–1630. [PubMed: 11390693]
52. van Beilen JB, Funhoff EG. Alkane hydroxylases involved in microbial alkane degradation. Appl Microbiol Biotechnol. 2007; 74:13–21. [PubMed: 17216462]
53. Sabirova JS, Ferrer M, Lunsdorf H, Wray V, Kalscheuer R, Steinbuchel A, Timmis KN, Golyshin PN. Mutation in a “tesB-like” hydroxyacyl-coenzyme A-specific thioesterase gene causes hyperproduction of extracellular polyhydroxyalkanoates by *Alcanivorax borkumensis* SK2. J Bacteriol. 2006; 188:8452–8459. [PubMed: 16997960]
54. Sticher P, Jasper MCM, Stemmler K, Harms H, Zehnder AJB, van der Meer JR. Development and characterization of a whole cell bioluminescent sensor for bioavailable middle-chain alkanes in contaminated groundwater samples. Appl Environ Microb. 1997; 63:4053–4060.
55. Igarashi RY, Seefeldt LC. Nitrogen fixation: The mechanism of the Mo-dependent nitrogenase. Critical Reviews in Biochemistry and Molecular Biology. 2003; 38:351–384. [PubMed: 14551236]
56. Dixon R, Kahn D. Genetic regulation of biological nitrogen fixation. Nature Reviews Microbiology. 2004; 2:621–631.
57. Rubio LM, Ludden PW. Biosynthesis of the iron-molybdenum cofactor of nitrogenase. Annu Rev Microbiol. 2008; 62:93–111. [PubMed: 18429691]
58. Rubio LM, Ludden PW. Maturation of nitrogenase: a biochemical puzzle. J Bacteriol. 2005; 187:405–414. [PubMed: 15629911]
59. Fischer HM. Genetic regulation of nitrogen fixation in rhizobia. Microbiological Reviews. 1994; 58:352–386. [PubMed: 7968919]
60. Dixon R, Cheng Q, Shen GF, Day A, Dowson-Day M. Nif gene transfer and expression in chloroplasts: prospects and problems. Plant and Soil. 1997; 194:193–203.
61. Cheng Q. Perspectives in biological nitrogen fixation. J Integrative Plant Biology. 2008; 50:786–798.
62. Moshiri F, Kim J, Fu C, Maier R. The FeSII protein of *Azotobacter vinelandii* is not essential for aerobic nitrogen fixation but confers significant protection to oxygen mediated inactivation of nitrogenase in vitro and in vivo. Mol Microb. 1994; 14:104–114.
63. Chen JS, Toth J, Kasap M. Nitrogen-fixation genes and nitrogenase activity in *Clostridium acetobutylicum* and *clostridium beijerinckii*. J Industrial Microbiolgy & Biotechnology. 2001; 27:281–286.
64. Gupta N, Balomajumder C, Agarwal VK. Enzymatic mechanism and biochemistry for cyanide degradation: A review. J Hazardous Materials. 2010; 176:1–13.
65. Bhattacharjee RB, Singh A, Mukhopadhyay SN. Use of nitrogen-fixing bacteria as biofertiliser for non-legumes: prospects and challenges. Appl Microbiol Biotechnol. 2008; 80:199–209. [PubMed: 18600321]
66. Brouers M, Hall DO. Ammonia and hydrogen production by immobilized cyanobacteria. J Bacteriol. 1986; 3:307–321.
67. Pieper DH. Aerobic degradation of polychlorinated biphenyls. Appl Microbiol Biotechnol. 2005; 67:170–191. [PubMed: 15614564]
68. Pieper DH, Seeger M. Bacterial metabolism of polychlorinated biphenyls. J Mol Microbiol Biotechnol. 2008; 15:121–138. [PubMed: 18685266]
69. Fortin PD, MacPherson I, Neau DB, Bolin JT, Eltis LD. Directed evolution of a ring-cleaving dioxygenase for polychlorinated biphenyl degradation. J Biol Chem. 2005; 280:42307–42314. [PubMed: 16227200]
70. Suenaga H, Ohnuki T, Miyazaki K. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. Environ Microbiol. 2007; 9:2289–2297. [PubMed: 17686025]

71. Pohlmann A, Fricke WF, Reinecke F, Kusian B, et al. Genome sequence of the bioplastic-producing “Knallgas” bacterium *Ralstonia eutropha* H16. Nat Biotechnol. 2006; 24:1257–1262. [PubMed: 16964242]
72. Madison LL, Huisman GW. Metabolic engineering of poly(3-hydroxyalkanoates): from DNA to plastic. Microbiol Mol Biol Rev. 1999; 63:21–53. [PubMed: 10066830]
73. Slater S, Mitsky TA, Houmiel KL, Hao M, et al. Metabolic engineering of Arabidopsis and Brassica for poly(3-hydroxybutyrate-co-3-hydroxyvalerate) copolymer production. Nat Biotechnol. 1999; 17:1011–1016. [PubMed: 10504704]
74. Aldor IS, Keasling JD. Process design for microbial plastic factories: metabolic engineering of polyhydroxyalkanoates. Curr Opin Biotechnol. 2003; 14:475–483. [PubMed: 14580576]
75. Fischbach MA, Walsh CT. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. Chem Rev. 2006; 106:3468–3496. [PubMed: 16895337]
76. Gorby YA, Yanina S, McLean JS, Rosso KM, Moyles D, Dohnalkova A, Beveridge TJ, Chang IS, Kim BH, Kim KS, Culley DE, Reed SB, Romine MF, Saffarini DA, Hill EA, Shi L, Elias DA, Kennedy DW, Pinchuk G, Watanabe K, Ishii S, Logan B, Nealson KH, Fredrickson JK. Electrically conductive bacterial nanowires produced by *Shewanella oneidensis* strain MR-1 and other microorganisms. Proc Natl Acad Sci USA. 2006; 103:11358–11363. [PubMed: 16849424]
77. Penn J, Li X, Whiting A, Latif M, et al. Heterologous production of daptomycin in *Streptomyces lividans*. J Ind Microbiol Biotechnol. 2006; 33:121–128. [PubMed: 16261359]
78. Baltz RH. Daptomycin: mechanisms of action and resistance, and biosynthetic engineering. Curr Opin Chem Biol. 2009; 13:144–151. [PubMed: 19303806]
79. Staunton J, Weissman KJ. Polyketide biosynthesis: a millennium review. Nat Prod Rep. 2001; 18:380–416. [PubMed: 11548049]
80. Fujii I. Heterologous expression systems for polyketide synthases. Nat Prod Rep. 2009; 26:155–169. [PubMed: 19177221]
81. Rodriguez E, Hu Z, Ou S, Volchegursky Y, et al. Rapid engineering of polyketide overproduction by gene transfer to industrially optimized strains. J Ind Microbiol Biotechnol. 2003; 30:480–488. [PubMed: 12698320]
82. McDaniel R, Thamchaipenet A, Gustafsson C, Fu H, et al. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel “unnatural” natural products. Proc Natl Acad Sci U S A. 1999; 96:1846–1851. [PubMed: 10051557]
83. Menzella HG, Reid R, Carney JR, Chandran SS, et al. Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. Nat Biotechnol. 2005; 23:1171–1176. [PubMed: 16116420]
84. Walsh CT, Fischbach MA. Natural products version 2.0: connecting genes to molecules. J Am Chem Soc. 2010; 132:2469–2493. [PubMed: 20121095]
85. Umeno D, Tobias AV, Arnold FH. Diversifying carotenoid biosynthetic pathways by directed evolution. Microbiol Mol Biol Rev. 2005; 69:51–78. [PubMed: 15755953]
86. Schmidt-Dannert C, Umeno D, Arnold FH. Molecular breeding of carotenoid biosynthetic pathways. Nat Biotechnol. 2000; 18:750–753. [PubMed: 10888843]
87. Cogdell RJ, Gall A, Kohler J. The architecture and function of the light-harvesting apparatus of purple bacteria: from single molecules to in vivo membranes. Quarterly Reviews of Biophysics. 2006; 39:227–324. [PubMed: 17038210]
88. Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. Nat Biotechnol. 2003; 21:796–802. [PubMed: 12778056]
89. Becker A, Katzen F, Puhler A, Ielpi L. Xanthan gum biosynthesis and application: a biochemical/genetic perspective. Appl Microbiol Biotechnol. 1998; 50:145–152. [PubMed: 9763683]
90. Sa-Correia I, Fialho AM, Videira P, Moreira LM, et al. Gellan gum biosynthesis in *Sphingomonas paucimobilis* ATCC 31461: genes, enzymes and exopolysaccharide production engineering. J Ind Microbiol Biotechnol. 2002; 29:170–176. [PubMed: 12355314]
91. Sanchez C, Mendez C, Salas JA. Indolocarbazole natural products: occurrence, biosynthesis, and biological activity. Nat Prod Rep. 2006; 23:1007–1045. [PubMed: 17119643]

92. Nakano H, Omura S. Chemical biology of natural indolocarbazole products: 30 years since the discovery of staurosporine. J Antibiot (Tokyo). 2009; 62:17–26. [PubMed: 19132059]

93. Howard-Jones AR, Walsh CT. Staurosporine and rebeccamycin aglycones are assembled by the oxidative action of StaP, StaC, and RebC on chromopyrrolic acid. J Am Chem Soc. 2006; 128:12289–12298. [PubMed: 16967980]

94. Salas JA, Mendez C. Indolocarbazole antitumour compounds by combinatorial biosynthesis. Curr Opin Chem Biol. 2009; 13:152–160. [PubMed: 19251468]

95. Jones MR. The petite purple powerpack. Biochemical Society Transactions. 2009; 37:400–407. [PubMed: 19290870]

96. Naylor GW, Addlesee HA, Gibson LCD, Hunter CN. The photosynthesis gene cluster of *Rhodobacter sphaeroides*. Photosynthesis Research. 1999; 62:121–139.

97. Alberti, M.; Burke, DH.; Hearst, JE. Structure and sequence of the photosynthesis gene cluster. In: Blankenship, RE.; Madigan, MT.; Bauer, CE., editors. Anoxygenic Photosynthetic Bacteria. Kluwer Academics; Netherlands: 1995. p. 1083-1106.

98. Lu Y, Xu J, Liu B, Kong J. Photosynthetic reaction center functionalized nano-composite films: Effective strategies for probing and exploiting the photo-induced electron transfer of photosensitive membrane protein. Biosensors and Bioelectronics. 2007; 22:1173–1185. [PubMed: 16815004]

99. Logan BE, Regan JM. Electricity-producing bacterial communities in microbial fuel cells. Trends in Microbiology. 2006; 14:512–518. [PubMed: 17049240]

100. Reguera G, McCarthy KD, Mehta T, Nicoll JS, Tuoinen MT, Lovley D. Extracellular electron transfer via microbial nanowires. Nature. 2005; 435:1098–1101. [PubMed: 15973408]

101. El-Naggar MY, Gorby YA, Xia W, Nealson KH. The molecular density of states in bacterial nanowires. Biophysical Journal. 2008:L10–L12. [PubMed: 18441026]

102. Revil A, Mendonca CA, Atekwana EA, Kulessa B, Hubbard SS, Bohlen KJ. Understanding biogeobatteries: Where geophysics meets microbiology. J Geophysical Res. 2010; 115:1–22.

103. Richter H, McCarthy K, Nevin KP, Johnson JP, Rotello VM, Lovley DR. Electricity generation by geobacter sulfurreducens attached to gold electrodes. Langmuir. 2008; 24:4376–4379. [PubMed: 18303924]

104. Hooshangi S, Thiberge S, Weiss R. Ultasensitivity and noise propagation in a synthetic transcriptional cascade. Proc Natl Acad Sci U S A. 2005; 102:3581–3586. [PubMed: 15738412]

105. Marles-Wright J, Grant T, Delumeau O, van Duinen G, Firbank SJ, Lewis PJ, Murray JW, Newman JA, Quin MB, Race PR, Rohou A, Tichelaar W, van Heel M, Lewis RJ. Molecular architecture of the “stressosome,” a signal integration and transduction hub. Science. 2008; 322:92–96. [PubMed: 18832644]

106. Marles-Wright J, Lewis RJ. The stressosome: molecular architecture of a signaling hub. Biochemical Society Transactions. 2010; 38:928–933. [PubMed: 20658979]

107. Hecker M, Pane-Farre J, Volker U. SigB-dependent general stress response in *Bacillus subtilis* and related gram-positive bacteria. Annu Rev Microbiol. 2007; 61:215–236. [PubMed: 18035607]

108. Igoshin OA, Brody MS, Price CW, Savageau MA. Distinctive topologies of partner-switching signaling networks correlate with their physiological roles. J Mol Biol. 2007; 369:1333–1352. [PubMed: 17498739]

109. Horvath P, Barrangou R. CRISPR/Cas, the Immune System of Bacteria and Archea. Science. 2010; 327:167–170. [PubMed: 20056882]

110. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archea. Nature Reviews Genetics. 2010; 11:181–190.

111. Jones DT, Shirley M, Wu X, Keis S. Bacteriophage infections in the inudstrial acetone butanol (AB) fermentation process. J Mol Micrbiol Biotechnol. 2000; 2:21–26.

112. Los M, Czyz A, Sell E, Wegrezyn A, Neubauer P, Wegrzyn G. Beacteriophage contamination: is there a simple method to reduce its deleterious effects in laboratory cultures and biotechnological factories? J Appl Genet. 2004; 45:111–120. [PubMed: 14960775]

113. Flynn CE, Lee SW, Peelle BR, Belcher A. Viruses as vehicles for growth, organization and assembly of materials. Acta Materials. 2003; 51:5867–5880.

114. Sulakvelidze A, Alavidze Z, Morris JG. Bacteriophage Therapy. Antimicrob Agents Chemother. 2001; 45:649–659. [PubMed: 11181338]

115. Lu TK, Collins JJ. Dispersing biofilms with engineered enzymatic bacteriophage. Proc Natl Acad Sci U S A. 2007; 104:11197–11202. [PubMed: 17592147]

116. van der Ploeg JR. Analysis of CRISPR in *Streptococus* mutans suggests frequent occurrence of aquired immunity against infection by M102-like bacteriophages. Micrbiology. 2009; 155:1966–1976.

117. Clarke DJ. Photorhabdus: a model for the analysis of pathogenicity and mutualism. Cellular Microbiology. 2008; 10:2159–2167. [PubMed: 18647173]

118. Duchaud E, et al. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. Nature Biotech. 2003; 21:1307–1313.

119. Waters CM, Bassler BL. Quorum Sensing: Cell-to-cell communication in bacteria. Annu Rev Cell Dev Biol. 2005; 21:319–346. [PubMed: 16212498]

120. Wilkinson P, et al. Comparative genomics of the emerging human pathogen *Photorhabdus asymbiotica* with the insect pathogen *Photorhabdus luminescens*. BMC Genomics. 2009; 10:302. [PubMed: 19583835]

121. Brenner K, You L, Arnold FH. Engineering microbial consortia: a new frontier in synthetic biology. Cell. 2008; 26:483–489.

122. Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R. A synthetic multicellular system for programmed pattern formation. Nature. 2005; 434:1130–1134. [PubMed: 15858574]

123. Tabor JJ, Salis HM, Simpson ZB, Chevalier AA, Levskaya A, Marcotte EM, Voigt CA, Ellington AD. A synthetic genetic edge detection program. Cell. 2009; 137:1272–1281. [PubMed: 19563759]

124. Anderson JC, Clarke EJ, Arkin AP, Voigt CA. Environmentally controlled invasion of cancer cells by engineered bacteria. J Mol Biol. 2006; 355:619–627. [PubMed: 16330045]

125. Kubori T, Matsushima Y, Nakamura D, Uralil J, Lara-Tejaro M, Sukhan A, Galan JE, Aizawa SI. Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. Science. 1998; 280:602–605. [PubMed: 9554854]

126. Kubori TA, Sukhan S, Aizawa I, Galan JE. Molecular characterization and assembly of the needle complex of *Salmonella typhimurium* type III protein secretion system. Proc Natl Acad Sci U S A. 2000; 97:10225–10230. [PubMed: 10944190]

127. Scheuring S, Rigaud JL, Sturgis JN. Variable LH2 stoichiometry and core clustering in native membranes of *Rhodospirillum photometricum*. Embo J. 2004; 23:4127–4133. [PubMed: 15457213]

128. Lovley DR. Bug juice: harvesting electricity with microorganisms. *Nature Reviews*. Microbiology. 2006; 4:497–508. [PubMed: 16778836]

129. Ntarlagiannis D, Atekwana EA, Hill EA, Gorby Y. Microbial nanowires: Is the subsurface “hardwired”? Geophysical Res Lett. 2007; 34:L17305.

130. Straight PD, Fischbach MA, Walsh CT, Rudner DZ, Kolter R. A singular enzymatic megacomplex from *Bacillus subtilis*. Proc Natl Acad Sci U S A. 2007; 104:305–310. [PubMed: 17190806]

131. Noodleman, L.; Case, DA.; Han, WG.; Lovell, T.; Liu, T.; Thompson, MJ.; Torres, RA. Scientific Report. Scripps; 2004.

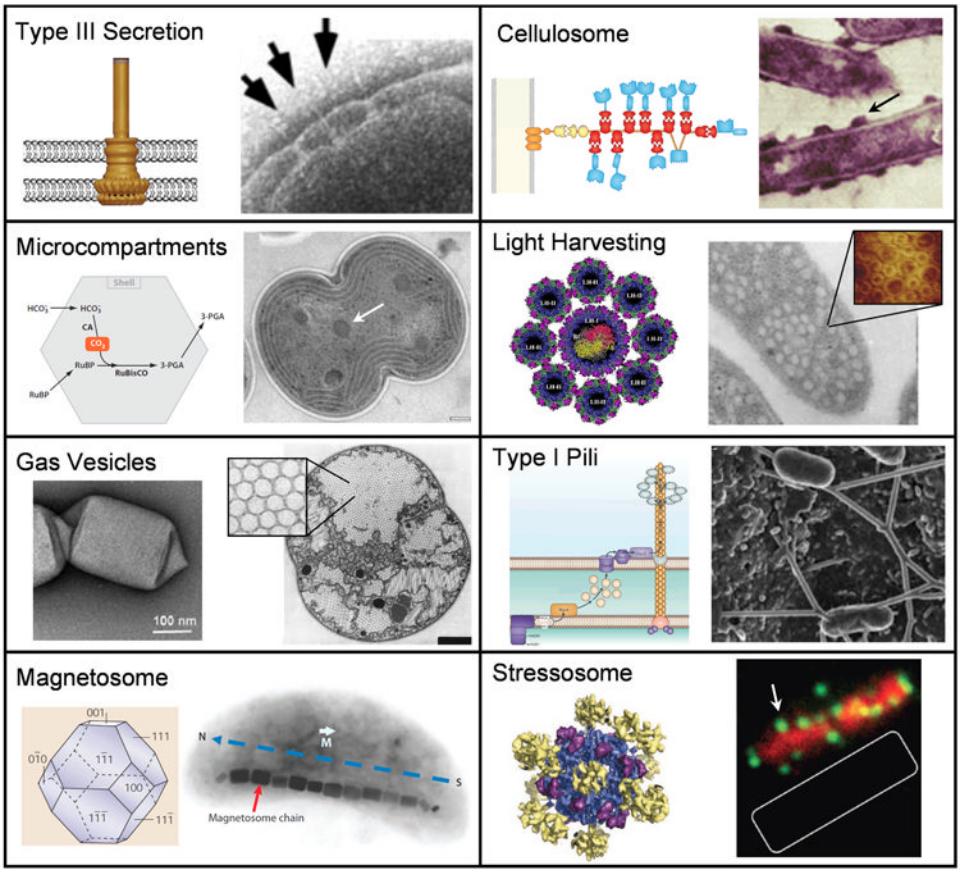

Figure 1. Gene clusters encode organelles and molecular machines

A schematic (left) and image (right) is shown for each system that appears in this review. (clockwise from top left) A reconstruction of the cryo-EM structure of the Salmonella type III secretion system is shown along with an image of multiple needle complexes spanning the inner and outer membranes [125, 126]. A schematic of the *C. thermocellum* cellulosome is shown with their surface localization [41]. The crystal structure of the R. sphaerodes Reaction center and LH1 (center) and LH2 (outer eight rings) are shown next to a high-resolution AFM of the photosynthetic membrane of Rsp. photometricum (inset) [127]. A cartoon of the type I pili from G. sulfurreducens (Fe^{3+} oxide is shown localized at the tip) [128] is compared to a SEM of the pilus-like appendages from S. oneidensis MR-1 [129]. The cryo-EM structure of the *B. subtilis* stressosome and their localization using RsbR-specific antibodies are shown [105]. An idealized {100}+{111} Fe_3O_4 crystal is shown with an image of a magnetosome chain [36]. An electron micrograph of a gas vesicle from *Hfx. Mediterranei* [29] and the packing of multiple vesicles in *Microcystis sp*. [30] are shown. Carboxysomes are shown from Synechocystis sp. PCC 6803 along with the pathway for carbon dioxide fixation [23]. All images reproduced with permission.

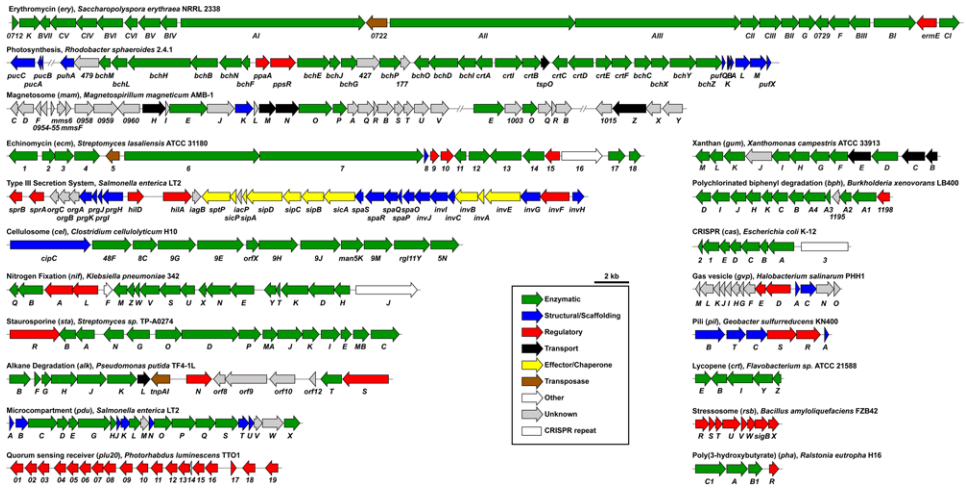

Figure 2. The gene clusters described in this review are compared

The colors of the genes loosely classify their functions. Many genes contain multiple functions. A classification of being “structural” includes genes genes that associate with a large complex and are necessary for function; for example ATPases in type I pili and type III secretion.

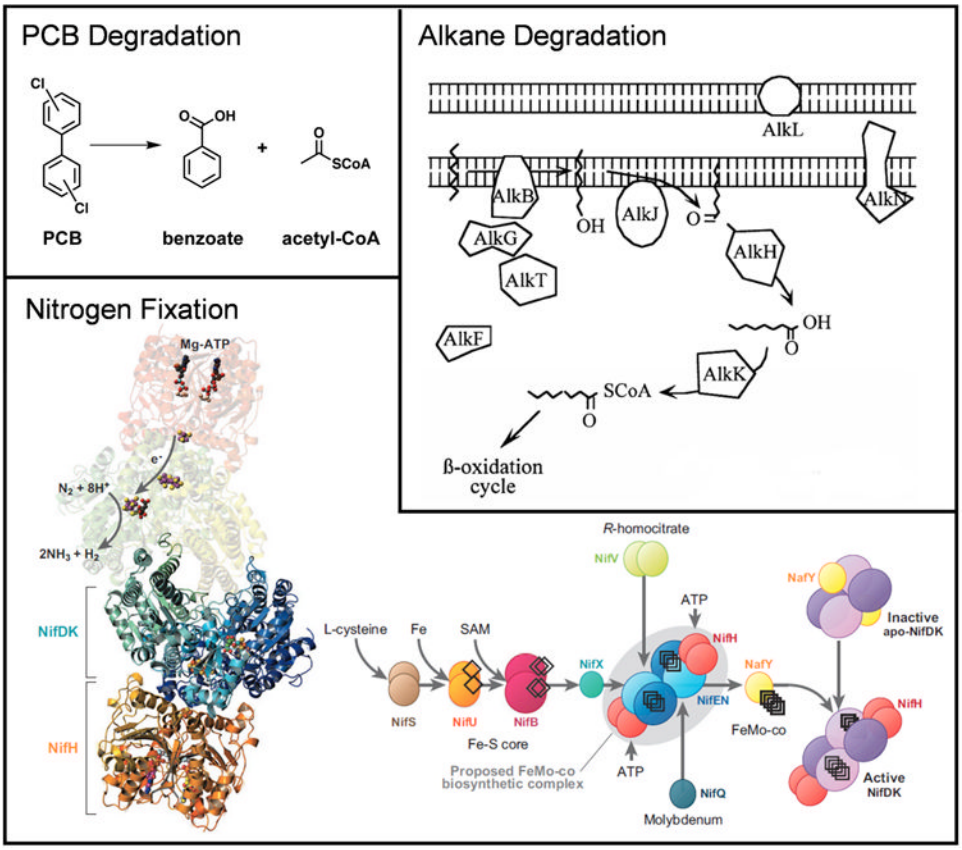

Figure 3. Utilization and breakdown pathways encoded in gene clusters are shown

The alkane degradation pathway from *P. putida* is adapted from Witholt and co-workers [51]. Nitrogenase is shown along with the pathway for the production of FeMoCo [57]. All images reproduced with permission.

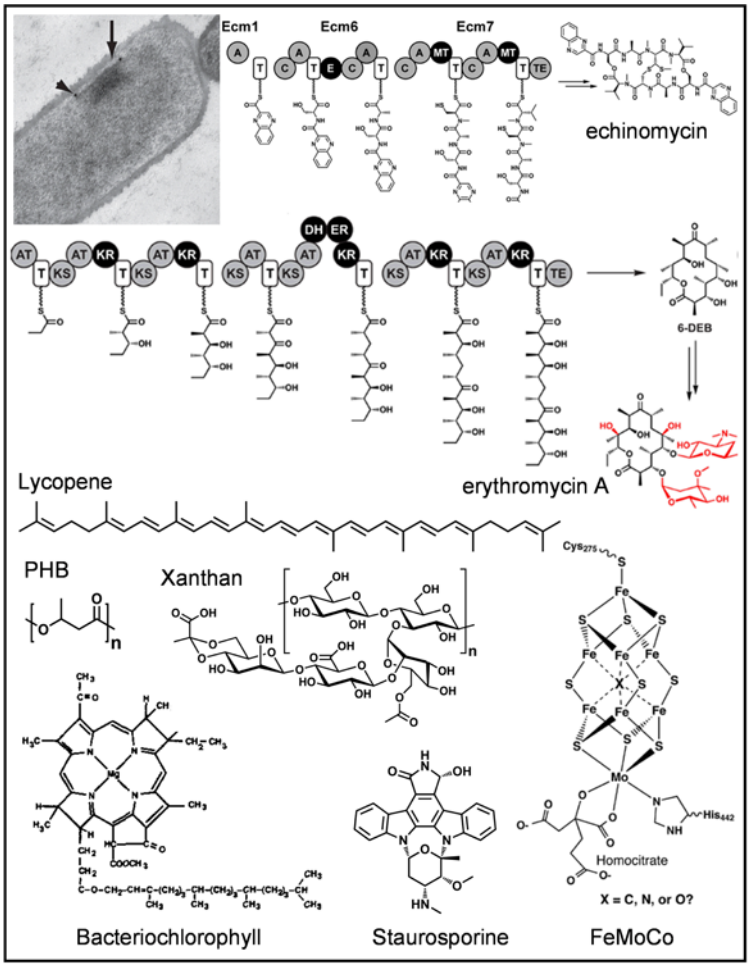

Figure 4. Chemical production pathways are often encoded within gene clusters

The image is of an organelle containting 10-100 associated 2.5 megadalton NRPS-PKS megacomplexes from *B. subtilis* [130]. The erythromycin pathway is shown from *Saccharopolyspora erythraea* NRRL 2338 and echinomycin pathway from *Streptomyces lasaliensis*. For erythromycin, chemical groups added by post-assembly-line tailoring enzymes (two P450s and two glycosyltransferases) are shown in red. Abbreviations: A = adenylation, T = thiolation, C = condensation, E = epimerization, MT = methyltransferase, TE = thioesterase, AT = acyltransferase, KS = ketosynthase, KR = ketoreductase, DH = dehydratase, ER = enoylreductase. Bacteriochlorophyll (bottom left) is incorporated into light harvesting complexes. FeMoCo (bottom right) is produced by a metabolic pathway (Figure 4) and incorporated into nitrogenase (image from [131]). All images reproduced with permission.

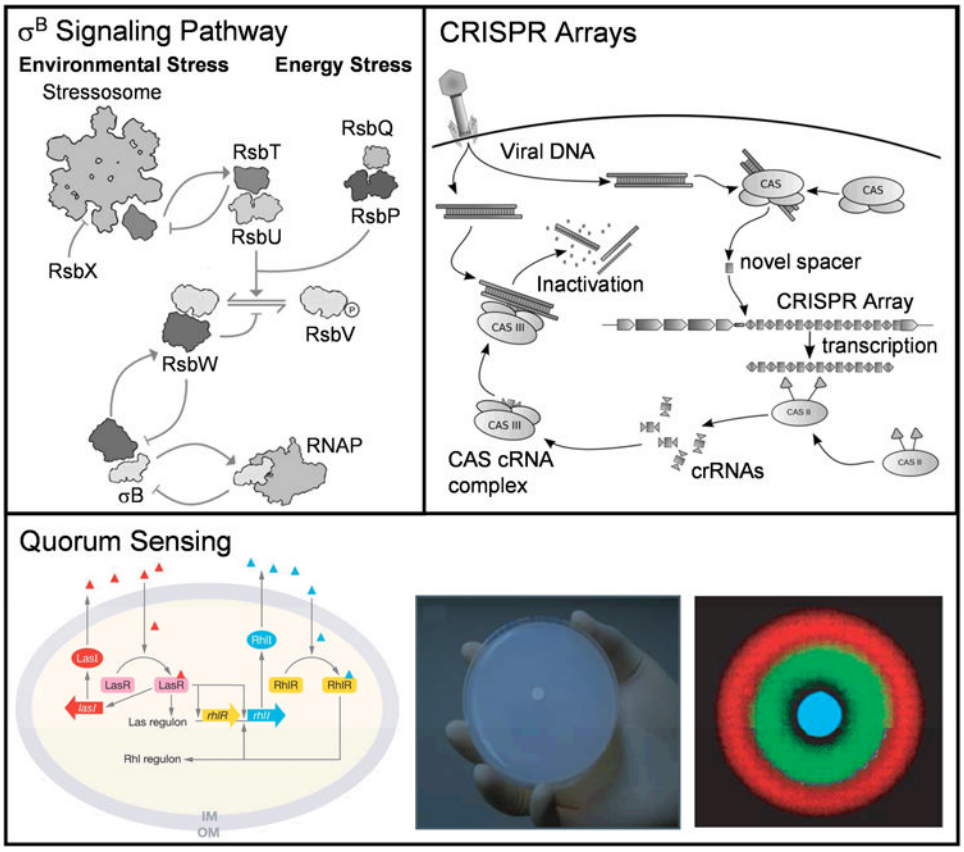

Figure 5. Complex regulatory pathways can be encoded by gene clusters

The signaling network formed by the σ_B gene cluster is shown [106]. Environmental stress is received by the stressosome, whereas energy stress is sensed by a different branch of the pathway. Quorum sensing pathways from Pseudomonas aeruginosa [119] are shown with their synthetic use to build pattern-forming programs in *E. coli* [122]. All images reproduced with permission. The CRISPR image adapted from one drawn by James Atmos.