Mining Environmental Plasmids for Synthetic Biology Parts and Devices

ESTEBAN MARTÍNEZ-GARCÍA,1 ILARIA BENEDETTI,1 ANGELES HUESO,1 and VÍCTOR DE LORENZO1

1Systems and Synthetic Biology Programs, Centro Nacional de Biotecnología, CNB-CSIC, 3, Darwin Street, 28049 Madrid, Spain

ABSTRACT The scientific and technical ambition of contemporary synthetic biology is the engineering of biological objects with a degree of predictability comparable to those made through electric and industrial manufacturing. To this end, biological parts with given specifications are sequence-edited, standardized, and combined into devices, which are assembled into complete systems. This goal, however, faces the customary context dependency of biological ingredients and their amenability to mutation. Biological orthogonality (i.e., the ability to run a function in a fashion minimally influenced by the host) is thus a desirable trait in any deeply engineered construct. Promiscuous conjugative plasmids found in environmental bacteria have evolved precisely to autonomously deploy their encoded activities in a variety of hosts, and thus they become excellent sources of basic building blocks for genetic and metabolic circuits. In this article we review a number of such reusable functions that originated in environmental plasmids and keep their properties and functional parameters in a variety of hosts. The properties encoded in the corresponding sequences include inter alia origins of replication, DNA transfer machineries, toxin-antitoxin systems, antibiotic selection markers, site-specific recombinases, effector-dependent transcriptional regulators (with their cognate promoters), and metabolic genes and operons. Several of these sequences have been standardized as BioBricks and/or as components of the SEVA (Standard European Vector Architecture) collection. Such formatting facilitates their physical composability, which is aimed at designing and deploying complex genetic constructs with new-to-nature properties.

INTRODUCTION

In its most widespread meaning, synthetic biology has been defined as the engineering of biology (1, 2). There are two sides to this definition. The first deals with fundamental science, as synthesis is the counterpart of analysis, the second leg of any rigorous research endeavor. Synthesizing something by rational assembly of its individual components is, in fact, the ultimate proof of understanding, as the celebrated statement of R. Feynman posthumously declared (“What I cannot create, I do not understand.”). On the other hand, the definition above implicitly announces that biological objects can indeed be engineered for a practical purpose. While in the field of molecular biology the term “genetic engineering” is normally used as a metaphor or an analogy, synthetic biology adopts “engineering” as an authentic conceptual and technical frame for repurposing existing biological entities and for creating new-to-nature properties. The underlying idea is that any biological system can be regarded as a combination of individual functional elements that are comparable to those found in man-made devices. These can be described as wholes of a limited number of components that can possibly be combined...
in novel configurations to modify existing properties or to create new ones—and so can their biological counterparts. Paramount in this concept is the identification and rigorous description of biological parts (the shortest DNA sequences encoding unique, stand-alone, unambiguous biological functions) and devices (an assembly of parts that runs a specified action with a definite input and output governed by a fixed transfer function). Although parts and devices are the basis of any extant biological system, the angle of synthetic biology is that at least some of them can be excised from their natural context, reformatted to meet some compositional standards, and rewired with a different genetic connectivity to create synthetic systems that bring about novel phenotypic traits in a host organism (often called “the chassis” in the synthetic biology jargon) (3). For this to happen, it is essential that parts and devices maintain the functions and the parameters that they possess in their natural context once they are placed somewhere else.

Alas, any biologist knows this is not likely to be the case (4). Unlike components used in industrial manufacturing, the behavior of biological constituents is extremely context-dependent—let alone that they can mutate and evolve erratically. Furthermore, their combinations may create emergent properties that are often unpredictable, thereby hindering any rigorous engineering based on them. Not surprisingly, many contemporary synthetic biology efforts try to tackle this challenge. One of the avenues to this end is the pursuit of “orthogonality” of the engineered constructs in respect to their host. This concept, imported from mathematics and computer science, indicates zero or a minimal influence of the engineered devices and systems on their biological carrier and vice versa. One approach to orthogonality involves eliminating as many as possible of the undesired connections that a part or device may have in its natural context or to set new information codes that only the designed construct can decipher (5). But one can also ask whether nature has in some cases favored the evolution of functional modules with basic functions that are to an extent autonomous in respect to the biological container where they reside.

This is where plasmids and other mobile genetic elements come into play in the synthetic biology scenario, as they are able to travel through different types of microorganisms and express their genetic complement in diverse genomic and metabolic backgrounds. Phages, in particular, are programmed to take over the entire expression machinery of the host for its own sake, often by means of alternative RNA polymerases (RNAPs) that recognize promoter sequences altogether different from those of the host. Not surprisingly, the RNAP of bacteriophage T7 (T7 pol) is one of the most used biological parts in synthetic biology: the pair T7pol/P_{T7} (T7 RNAP and the cognate promoter of gene 10 of phage T7) suffices to trigger transcription of downstream DNA sequences whether in vivo or in vitro, with the only condition being that nucleoside triphosphates are available. This happens regardless of the biological carrier, from bacteria to animal cells, thus providing a reasonably orthogonal tool for manipulation gene expression at the user’s will. Also, repressors found in transposons are excellent assets for biological engineering. Repression is a less demanding mechanism of transcriptional regulation in prokaryotes (6) and is thus far easier to enter in a heterologous context. The tetracycline-dependent TetR repressor of the tetA gene of Tn10 and its target DNA sequence has been employed and refactored in many ways to this end (7), and it is also one of the most popular parts for construction of artificial genetic circuits. Nevertheless, synthetic biology is not only about heterologous, orthogonal transcription initiation; it requires many other engineer-able functions. These are not found in either phages or transposons but, as discussed below, on broad host range (BHR) conjugative or mobilizable plasmids.

PROMISCUOUS PLASMIDS BEAR KEY HOST-INDEPENDENT BIOLOGICAL FUNCTIONS

BHR conjugative (or at least mobilizable) plasmids are objectively the most favorable source of natural orthogonal parts and devices, as long as they deploy their full or partial genetic complement in diverse biological carriers. The key component to this end is the vegetative origin of replication. It is remarkable that many environmental plasmids seem to have just such an oriV that either uses the host replication machinery or is accompanied by a cognate, plasmid-encoded replication protein(s) gene(s), with no other recognizable attribute encoded, in what would look like a perfect example of “selfish” DNA. One can consider such minimalist plasmids as the evolutionary solution to the engineering challenge of having an optimal DNA-carrying vector. Note that the bottlenecks that even the simplest plasmids must overcome for their propagation involve not only their replication proper (for which they have a suite of mechanisms; see, e.g., references 8, 9) but also signaling replication termination and resolution of replication intermediates. By default, this can happen through homologous recombination between DNA sequences, but in other instances, specific molecular mechanisms take
care of unknotting the otherwise intractable replicating molecules.

But just having autonomous replication is no guarantee of doing well unless it is accompanied by other complementary functions. Environmental plasmids often ensure their maintenance by having active counter-regulation systems that force bacteria to retain them or otherwise be killed. But in other cases, stable plasmid-host partnerships occur without any conspicuous selective advantage. In many instances, however, plasmids end up recruiting to their frame some traits that are clearly beneficial to the host. These last include a large suite of generally dispensable genes which, under certain circumstances, can endow the carrier with some recompense in return for the physiological load of being carried as an extrachromosomal element. These genes range from resistance to antibiotics (the best known plasmid-borne features) to novel metabolic capabilities (e.g., routes for biodegradation of environmental pollutants). As discussed below, such pathways are themselves a source of both blocks for metabolic engineering as well as the source of a wealth of conditional expression systems that can be co-opted in synthetic constructs. To assemble whatever circuit of choice, such functional DNA fragments are typically recruited to the plasmid frame through a suite of site-specific recombination (SSR) devices which either remain active in the extant plasmid structures or become vestigial or even erased after a period of stable coexistence with a given host.

In a further turn of the screw, some plasmids deliver to the host novel functions that change their lifestyle (e.g., the thermal window of optimal viability) and even switch the order of preference of carbon source consumption (10–12). Spreading these properties in a population depends on plasmids' ability to move between different hosts. By default, this can occur through natural transformation, but it usually happens through conjugation. To this end, some plasmids carry a complete set of genes for conjugal self-transmission, while others just have an origin of transfer (oriT) which can be recognized by other plasmids' machinery and thus bring about mobilization. All these components can be found in various configurations in extant plasmids accompanied by what we could call "genetic garbage," i.e., remnants of insertion sequences and nonfunctional pseudogenes for which no clear role may be assigned upfront (13).

It is not surprising that some synthetic biology efforts have been recently directed to re-create a sort of quintessential BHR plasmid frame, devoid of parasitic-looking elements. Along this line, Hansen (14) constructed an all-synthetic IncX1 plasmid based on the components just mentioned, although some of them are somewhat compressed in very close or overlapping DNA sequences: tra (conjugal transfer), mob1 (mobilization 1), rep/stb (replication initiation/toxin-antitoxin plasmid addiction), mob2 (mobilization 2), par (plasmid partitioning), and res (resistance marker). The synthetic plasmid kept all the properties shared by the natural members of the IncX1 group but, as long as it was artificially simplified, created a construct much easier to further engineer (15). Table 1 summarizes the key functions that environmental plasmids are typically endowed with and which, following some formatting and repurposing, become useful assets for biosystem engineering along the line of contemporary synthetic biology. Note that the list is by no means exhaustive, and the reader is directed to the ACLAME database (http://aclame.ulb.ac.be) for a thorough compilation of what genes and functions one can find in a large collection of plasmids and other mobile elements. In the following sections we discuss in more detail some of these parts and how they have been incorporated into the synthetic biology toolbox.

**NARROW HOST RANGE ORIGINS OF REPLICATION**

Plasmid vectors are the basis of recombinant DNA technology and the workhorses of any genetic engineering or synthetic biology project. The one key function that enables their existence as extrachromosomal elements is their origin of replication. Most of those employed for gene cloning are intended to work in *Escherichia coli* and thus have a narrow host range. The most used to this end are those based on the vegetative origin of replication (oriV) of plasmid CoIE1/pMB1, which is the type of default origin of replication found in a large number of plasmid vectors (16). The main advantage is its small size and its mechanism of replication, which is entirely enabled by host functions (i.e., *E. coli*). This makes the DNA sequences needed for autonomous propagation of plasmids with a CoIE1 origin compressed in a very small fragment that can be easily moved according to the user's needs. Most CoIE1-derived vectors bear variants of this oriV with mutations that cause a very high copy number. While this eases DNA extraction and preparation from host cells, it may become detrimental if the plasmid carries cloned genes—and more so if they are toxic. A second type of narrow host range vectors are those based on the p15A replicon (16). These have a lower copy number than their CoIE1 counterparts (but still high) and are compatible with them, thereby...
TABLE 1 Principal biological parts found in BHR environmental plasmids

<table>
<thead>
<tr>
<th>Functional segment</th>
<th>Description</th>
<th>Synthetic biology application</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin of replication</td>
<td>Narrow</td>
<td>Physical assembly, gene cloning in E. coli</td>
<td>16–24</td>
</tr>
<tr>
<td></td>
<td>Broad</td>
<td>Analysis and engineering of diverse Gram-negative bacteria</td>
<td>25–47</td>
</tr>
<tr>
<td>Selection marker</td>
<td>Antibiotic resistance</td>
<td>Plasmid retention, biomarker</td>
<td>49–53</td>
</tr>
<tr>
<td></td>
<td>Autotrophic compensation</td>
<td>Self-selecting plasmid maintenance</td>
<td>54–56</td>
</tr>
<tr>
<td>Toxin/antitoxin systems</td>
<td>Plasmid countersegregation</td>
<td>Marker-free vectors, biosafety circuits, conditional viability</td>
<td>62–73</td>
</tr>
<tr>
<td>Conjugative transfer</td>
<td>Origin of transfer</td>
<td>Plasmid transmission between E. coli and diverse recipients</td>
<td>89–90</td>
</tr>
<tr>
<td></td>
<td>Transfer and mobilization</td>
<td>Suicide delivery of transposon and transposon-vectors</td>
<td>91–93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Biological computing and execution of logic (Boolean) programs</td>
<td>94</td>
</tr>
<tr>
<td>Regulated promoters</td>
<td>Substrate-induced transcription of metabolic operons</td>
<td>Regulatory devices turned on by chemical effectors</td>
<td>98–113</td>
</tr>
<tr>
<td>Cis-acting antisense RNAs</td>
<td>Formation of asRNA/target RNA duplexes that are degraded by RNase III</td>
<td>Gene expression silencing</td>
<td>116–123</td>
</tr>
<tr>
<td>Metabolic genes</td>
<td>Catabolism of unusual carbon sources</td>
<td>Models for metabolic engineering and implant-chassis retroactivity</td>
<td>130, 131</td>
</tr>
<tr>
<td>Site-specific recombination</td>
<td>Resolution of replication intermediates</td>
<td>Excision of undesired DNA segments in genomes of interest</td>
<td>74–80</td>
</tr>
<tr>
<td>systems</td>
<td></td>
<td>Engineering of operative systems, memory, and logic gates</td>
<td>82, 84</td>
</tr>
<tr>
<td></td>
<td>Recruitment of functional sequences</td>
<td>(Combinatorial) assembly of functional DNA in a gene cluster</td>
<td>85, 96</td>
</tr>
</tbody>
</table>

enabling expression of cloned genes from two coexisting vectors. Another oriV quite popular in plasmids for E. coli is SC101, whose advantage is its low copy number. The downside in this case is its lower yield when extracted from host bacteria and the need for an extra plasmid-encoded replication protein, which increases the size of the DNA required for its autonomous propagation.

Finally, one narrow host range replication origin that has been heavily refactored for different purposes since its description (17) is that of plasmid R6K. The property that makes this system so attractive is that the plasmidborne replication protein II (encoded by the so-called pir gene) can also actuate in trans on the origin of replication (oriVR6K). This allows the physical separation as well as the mutual dependence of the two components (pir gene and oriV) for plasmid functioning (Fig. 1). By artificially placing the pir gene in the chromosome and the oriV in a plasmid, one not only ensures a narrow-range replication system, but fulfills the complete requirement of a strain-specific host for vector propagation (18). This feature has been exploited in a large number of synthetic biology circuits aimed at suicide delivery of transposon vectors (19–21) (see below) and engineering of biosafety circuits by designing mutual host-plasmid addiction (22). Not surprisingly, the small collection of narrow host range origins of replication just mentioned have been incorporated into the various repositories of biological parts available to synthetic biologists (23, 24) (http://parts.igem.org).

BHR ORIGINS OF REPLICATION

As long as synthetic biology projects go beyond model E. coli strains and toy circuits and enter deployment of deeply engineered traits for industrial, medical, and environmental settings, new hosts (i.e., chassis) different from E. coli are required, and BHR plasmid vectors become mandatory. BHR origins of replication have been known since the onset of the recombinant DNA era (23, 26), but their generally more complex replication mechanism (and therefore their larger size) have limited their use in vector development. Such origins afford autonomous propagation of the replicon on a variety of bacterial carriers. Most BHRs of this kind have been developed for Gram-negative microorganisms, although some of them could also be useful in Gram-positive bacteria (see below). One widely used BHR origin is that of plasmid RSF1010. Although its replication mechanism is somewhat intricate and the corresponding functional segment contains intermingled replication and transfer functions (27, 28), RSF1010-based plasmids are among the most host-promiscuous vectors. The segment of the natural plasmid that suffices for autonomous replication has been recently streamlined (29) and formatted as a separate module for vector assembly (23). A second origin very frequently found in BHR vectors is that of the RK2/RP4 plasmid (30). Despite the (comparatively) small size of the cognate functional segment, the RK2 origin also counts among the most promiscuous. A large number of vectors—before and after
the onset of synthetic biology—based on RK2 have been instrumental in analyzing and constructing complex phenotypes in bacteria such as *Pseudomonas, Rhizobium, Agrobacterium*, and a suite of Gram-negative species (31–33).

But other origins of replication are increasingly used for BHR vector development as well. For instance, pBBR1, a plasmid first isolated from *Bordetella bronchiseptica* (34), was adopted as the basis of a considerable number of genetic tools (24, 35–37). One of the advantages is that BBR1-derived plasmids generally have a higher copy number than their RK2-based and RSF1010-based counterparts (with all the pluses and minuses discussed above). By the same token, the origin of environmental IncW plasmid pSa (38), which was adopted in the late 1980s for constructing BHR vectors for *Agrobacterium* (39) and *Xanthomonas* (40), has been recently reshaped and formatted as a building block for generating standardized vectors for *Ralstonia* (41). In contrast, genus-specific (e.g., *Pseudomonas*) origins of replication from, e.g., *Pseudomonas savastanoi* (42) which were proposed at some point for generating a suite of new BHR vectors (43) never made it to mainstream use. One way of broadening the range of such otherwise narrow-range plasmids is to join origins that function alternatively in two or more hosts in the same DNA segment, as has been done with the originally *P. aeruginosa*-only plasmid pRO1614 (44, 45) and the *E. coli*/*Pseudomonas putida/Bacillus* shuttle named pEBP (46). However, despite some incursions into other functional origins of replication, vectors based on RSF1010, RK2, BBR1, and pSa have been the pillars of the genetic engineering of Gram-negative organisms other than *E. coli* for a long time, and now they remain the basis of standardized modules for assembly of synthetic devices for the same type of organisms (23, 24) (http://parts.igem.org). Alas, the same standardization effort has been more limited thus far in the Gram-positive realm, although some exciting developments are under way at the time of writing this article (47, 48).

**SELECTION MARKERS**

Antibiotic resistance genes have been systematically the favorite resource for enabling selection of recombinant plasmids. By adding the cognate agent to the selection medium, one can not only pick out transformants after DNA ligation or assembly but also ensure retention of a given plasmid if no endogenous countersegregation system is in operation. A variety of resistance genes for ampicillin (*bla*, Ap) tetracycline (*tet*, Tc), chloramphenicol (*cat*, Cm), kanamycin (*kan*, Km), streptomycin/spectinomycin (*Sm/Sp*), and gentamycin (*Gm*) have been retrieved repeatedly from their original natural context (often a clinical isolate) and reused in a large variety of vectors (49, 50). More recently, their corresponding sequences have been formatted as modules for assembling all types of synthetic biology constructs (23, 24) (http://parts.igem.org). But the ease of use of these markers has two major drawbacks for applications involving industrial or environmental settings. First, the products of these genes act on antibiotics that are in clinical use (or are similar to them), and engineering bacterial biomass which carries resistance genes may thus raise safety concerns (51). Second, antibiotics themselves are expensive, and using them as a bulk additive to any large-scale growth medium might be economically prohibitive.

This has prompted the quest for alternative markers or strategies to ensure stable plasmid maintenance without having to use typical antibiotics. One such strategy is the adoption of resistance to herbicides (e.g., bialaphos) (52) or heavy metals (e.g., tellurite) (53) as markers. Another useful approach is placing in the plasmid the sequence of an essential gene that complements...
a deletion of the same gene excised from the chromosome of the host strain. In this way, any bacterium that loses the plasmid is eliminated. The genes infA (encoding translation initiation factor 1) (54), asd (aspartate-beta-semialdehyde dehydrogenase) (55), dapA (dihydrodipopicolinate synthase) (22), and thyA (thymidylate synthase) (56) have been used to this end, and many others are surely eligible for the same purpose. This type of marker eliminates the need for drug resistance markers in the vector and is thus appealing whenever stable maintenance of engineered traits is required in the absence of external selection. An alternative approach to ensure recombinant plasmid retention without antibiotics relies on the refactoring of toxin-antitoxin systems (see below).

In addition, users can consider the definite insertion in the chromosome of the DNA segment of choice by using one of the many mini-transposon vectors based on either Tn5 (20, 21) or Tn7 (57, 58) that have been designed to that end in recent years. Some of these transposon vectors incorporate in their design the possibility of removing the antibiotic markers employed for selection of insertion events once transposition has occurred (59, 60).

**TOXIN-ANTITOXIN SYSTEMS**

Although many naturally occurring plasmids, in particular the smallest, seem to be maintained on the sheer basis of their high numbers, others with lower copy doses have developed sophisticated molecular mechanisms to coordinate chromosomal and plasmid replication and/or to penalize cells that may accidentally lose the extra replicon. These so-called plasmid addiction systems (61) seem to be a special case of a more general method for protecting the integrity of the genomic complement under environmental and nutritional stress conditions (62–64). These devices systematically involve a DNA segment that encodes a toxin and its antidote. Toxin-antitoxin modules usually are composed of two cognate proteins with different expression levels and/or half lifetimes (e.g., the parD locus encoding Kis/Kid proteins in plasmid R1) (65). Under normal conditions, the antidote inhibits the toxin, whereas plasmid loss makes proteolytic degradation of the antidote faster, titrating the remedy out and killing the host cells. Alternatively, the antidote of the toxin-antitoxin pair is an RNA that prevents translation of a toxic gene in such a way that cells that lose the plasmid are actively eliminated by the system. The archetypal example of such a scenario is that of the parB system of R1, which encodes two small genes, bok and sok (66). The expression of the first (a toxin) is regulated by the complementary bok antisense RNA to the bok mRNA. Postsegregation killing in single cells is brought about by the differential decay of the bok and sok RNA. The modularity of the system stabilizes any plasmid that is inserted with a parB-encoding DNA segment (67). This feature has made the parB system a remarkable tool for ensuring plasmid maintenance in individual bacteria without any external selection (68).

A different type of toxin-antitoxin system is that of colicins (69, 70), large protein poisons that are released into the medium to kill related bacteria which may have lost the colicin locus (i.e., the DNA segment encoding the structural gene for the toxin and a colicin immunity gene) or similar strains that have a receptor for the colicin in question. Although the biological role of colicins is still debated (71), they provide a powerful tool to engineer population-wide circuits for genetic and biological containment of recombinant DNA (72, 73). In reality, one of the most sophisticated containment host-plasmid mutual systems designed thus far combines conditional plasmid replication (i.e., initiators for the R6K or ColE2-P9 origins are provided in trans by a specified host), rich-media-compatible auxotrophies (dapA and thyA; see above), and toxin-antitoxin pairs (22). Such constructs ensure a potentially perfect genetic firewall to engineered genetic devices without significantly affecting (at least in the short run) growth parameters.

**SITE-SPECIFIC RECOMBINATION**

While homologous recombination typically requires a considerable degree of DNA identity through a good length of the sequences in question, SSR occurs between shorter segments sharing limited similarity. This phenomenon, which is widespread through all biological kingdoms, is brought about by the action of the so-called site-specific recombinases on short DNA sites which they manage to recognize, cleave, exchange between DNA strands, and rejoin. This basic device, which by default just needs a recombinase and a suite of target sites, is a key component of a large collection of biological phenomena in both eukaryotic and prokaryotic organisms. One of these functions is the resolution of replication intermediates in plasmids with an active partitioning mechanism. An SSR which has been repeatedly reshaped and repurposed for genetic engineering is that of the multimer resolution system (mrs) of BHR plasmid RP4. This system is one of the functions encoded within the par region, which determines the major RP4 stabilization system. The product of the parA gene is a site-specific resolvase that catalyzes recA-independent
intramolecular recombination between two directly oriented resolution (res) sites. If these sites flank a supercoiled DNA segment, the result of the process is the excision of the intervening sequence. Transient expression of para has been exploited to delete specific DNA segments (e.g., antibiotic resistance genes) delivered with transposons to the chromosome of target bacteria (e.g., E. coli or Pseudomonas) (74). This permits the retention of heterologous DNA segments devoid of any phenotypic marker to the genome of the target bacterium, an important consideration for microorganisms destined for environmental release (75). Further plasmid-based site-specific recombination systems have been extensively employed for the same or similar purposes, e.g., the Cre-lox resolvase of the phage/plasmid P1 (76, 77). But others are available and well suited as tools for deep refactoring of both prokaryotic and eukaryotic systems. Among them is the β-recomytic system of pSM19035 (78), a virtually orthogonal SSR that originates in the complex replication mechanism of this plasmid of Gram-positive bacteria (79, 80).

Integrons are a second type of useful SSR system often found in environmental plasmids and other mobile elements (81). They are DNA elements with an ability to capture genes with a mechanism that typically involves an integrase gene (int) and a nearby recombination site (att). The recruitable DNA is located on gene cassettes in which the functional sequences are adjacent to a matching att recombination site. The essence of the integron is that such DNA sequences must occur in covalently closed circular DNA but cannot be replicated as such; they can only be replicated when recombination occurs between the two att sites, thereby entering the cassette into the cognate target. As the att sites are maintained after the integration, the resulting molecular product can again act as a landing pad for further integration rounds. Furthermore, this is a reversible process, so that DNA sequences can be further excised and integrated in a very dynamic fashion. Although the activity of integrons is one of the most dreaded mechanisms of acquisition of antibiotic resistance in clinical settings (81), they provide an extraordinary platform for creating genetic diversity in synthetic biology constructs (82). In one case, a tryptophan production operon was reconstructed by separately delivering individual genes as recombination cassettes to a synthetic integron platform in E. coli. Integrase-mediated recombination generated a large number of genetic combinations in vivo with a large number of combinations of thereby arrayed genes, some with increasing production of the amino acid by 11-fold compared to the natural gene order of the same pathway (83).

In a further twist of their natural properties, SSR systems can not only be used as secondary tools for construction of synthetic biology devices but also can be engineered as actuators able to record information and implement a gene expression program. For instance, by flanking transcriptional terminators or promoters with SSR targets one can engineer complex circuits based on the control of the flow of RNAP along DNA. Integrase-mediated inversion or deletion of such transcription checkpoints allows construction of AND, NAND, OR, XOR, NOR, and XNOR logic gates that use common signals and enable design of sequential logic operations (84). Similarly, it is possible to assemble genetic circuits that use recombinases for running Boolean logic functions by means of DNA-encoded memory of events (85). One appealing possibility in this context is the engineering of circuits in which stable gene expression (output) is the result of transient environmental or chemical inputs (86). If one combines SSR-based gates with suitable reporter systems (e.g., the green fluorescent protein gene [GFP]) and PCR, it becomes possible to interrogate the state of the synthetic devices for past events.

**CONJUGATION AND TRANSFER FUNCTIONS**

Plasmids are the main agents of horizontal gene transfer in nature. This can occur through natural transformation (sometimes electro-transformation due to atmospheric lighting) (87, 88), but usually plasmid transit from one bacterium to another occurs through conjugation. This phenomenon includes different mechanisms of DNA delivery that require a physical contact between a donor and a recipient, often facilitated by specific pili which bridge and bring together mating partners. Complete conjugation systems share a basic layout composed of four constituents (89). The first is the so-called relaxase, the key protein in conjugation, which recognizes the second component: a short DNA sequence that acts as the oriT. Having a relaxase and an oriT in cis or in trans generally suffices for a plasmid to be conjugally transmissible (Fig. 1). The relaxase catalyzes (i) a site-specific cleavage of the oriT sequence in the donor, thereby producing a DNA strand that will be transferred, and (ii) the eventual religation of DNA once it reaches the recipient bacteria and transfer is completed, thereby reconstituting the whole plasmid in the new host. Depending on the plasmid in question, the relaxase may work in concert with other proteins that form a complex called relaxosome. To complete the conjugation process, one needs two additional functions. One is the...
formation of a mating channel through which DNA can pass from one cell to the other. This structure consists of a type IV secretion system (T4SS), which transports the relaxase attached to the DNA to be transferred. The second is the so-called type IV coupling protein (T4CP), which makes the essential connection of the relaxosome/DNA complex and the transport channel. In addition, some conjugative plasmids encode structural and regulatory genes for formation of pilus on the cell surface, which eases and tightens the physical connection between the donor and the recipient.

Typical self-transmissible environmental plasmids (e.g., RK2/RP4 or R388) encode the entire complement (oriT/tra/mob genes) for all functions necessary to bring about the complete transfer process. Other plasmids may carry only the relaxosomal components (i.e., oriT, the relaxase gene, and auxiliary proteins). In this case (e.g., RSF1010) the DNA can initiate its own transfer but has to be complemented by functions of other conjugative plasmids for finishing the task. Finally, other plasmids may just bear DNA sequences that can be recognized by relaxases encoded by other replicons and used as oriT for lateral transfer.

Some components of the conjugation and mobilization machineries just mentioned have been repurposed for diverse genetic engineering and synthetic biology applications. The most salient involves the inclusion of the RK2/RP4 oriT in a variety of plasmid vectors for making them mobilizable to new hosts by means of the transfer functions of the same plasmid provided in trans (Fig. 1). Note that this conjugative system is the most promiscuous known, as it is able not only to move plasmids among bacteria, but also to deliver them to yeasts and mammalian cells (90). Such a broad capability has been formatted in various ways. The first methods to mobilize oriT-containing plasmids from a donor host (typically E. coli) to other Gram-negative recipient bacteria required the action of a third mating partner which transiently provided such tra functions to the donor but could not replicate in the final recipient (91). The products encoded by the helper plasmid could thereby escort the transfer of the oriT construct without itself being inherited in the destination strain (Fig. 2). In a subsequent advance, specialized strains were designed such that the tra functions of RP4 were integrated in the chromosome of a donor E. coli strain where the oriT-plasmids could be placed and poised for conjugation with available target bacteria (92). This allowed users to get the desired plasmid transfer by setting a simple two-partner mating (rather than the tri-parental mating above). A side benefit of such a scheme was the possibility to combine plasmid mobilization with conditional replication of the same plasmid (possible in the donor bacterium, not possible in the recipient) as a way to deliver transposons engineered in the same circular DNA frame (18). Alas, the genetic technology available at the time obliged the donor bacteria to carry the Mu phage also (as a carrier of the RP4 tra functions), thus creating a chance of artifactual insertions and misleading phenotypes. Only recently has this problem been completely solved by engineering an E. coli Mu-free donor strain that stably harbors all of the RP4 conjugative functions and sustains replication of Π-dependent suicide vectors (93) (see above).

However, conjugation has been exploited in synthetic biology not only for the sake of using its functional parts, but also as a phenomenon that can be utilized for biological computing and digital cell-to-cell communication (94). One can see DNA transfer as a scenario where individual bacterial types perform specific sub-tasks, the results of which are then communicated to other cell types for further processing. Specific cell-cell conjugation allows direct transfer of genetic information in a digital fashion; i.e., cells in a population can only have or not have a given plasmid. Mixing different strains in a single population allows a multicellular pool to execute the Boolean XOR function. Models predict this approach to be more powerful than other biological computation approaches using, e.g., quorum sensing. A conjugation-wired computing system could thus become a realistic methodology to implement new-to-nature properties in biological systems (94).

To finish this section we need to refer to the so-called integrative and conjugative elements (ICEs), which combine interesting properties of transfer systems and site-specific recombination in the same DNA segment. ICEs are mobile genetic elements that by default stay in the host chromosome but which can occasionally be activated to adopt a plasmid-like form that can be transferred between cells by conjugation (95). ICEs can encode genes for antibiotic resistance, metabolic functions, virulence factors, and many other advantageous traits. The organization of ICEs varies enormously, but inter alia they encode an integrase that plays a role in different steps of the element’s life cycle. One interesting ICE specimen is the so-called ICEelec, an element found in Pseudomonas knackmussii B13 that provides the host with the capacity to metabolize chlorocatechols (95). The integrase activity of this ICE has been recently used for developing a genetic tool able to stably implant large DNA fragments into one or more specific sites of a target chromosome (96). The genetic kit to this end involves only a cloning
vector and a mini-transposable element with which large DNA segments engineered in *E. coli* can be tagged with the integrase gene. The system has been instrumental in introducing cosmids and bacterial artificial chromosome DNAs from *E. coli* into *P. putida* in a site-specific manner. These developments illustrate how fundamental knowledge of the different steps and components of horizontal gene transfer (HGT) become a treasure trove of building blocks for biosystems engineering.

**EXPRESSION SYSTEMS**

Environmental plasmids found in bacteria that inhabit niches with a history of pollution by industrial chemicals frequently carry in their DNA one or more gene clusters that expand the range of compounds that can be used as carbon sources (97). It is often the case that transcription of such genes responds to either the substrates of the corresponding pathways or to some of their metabolic intermediates. The bottom line is that in the very competitive environments where these bacteria thrive, expression of extra genes must be tightly controlled for being triggered only when and where they are needed. This scenario is thus an excellent ground in which to mine pairs of transcriptional factors (TFs)/cognate chemical inducers which can be reformatted for creating new conditional expression systems (98). This approach is helping to overcome the dearth of alternative inducible devices for synthetic biology circuits (Fig. 3), which is thus far limited to those typically used in *E. coli*: LacI/Plac (and its many variants), the thermo-inducible cI-repressed λ phage promoters, AraC/P_BAD, TetR/Ptet (24), and more recently, the RhaR/Prha system (99), either by themselves or in combination with the T7 RNAP.

A number of alternative effector-triggered systems from various origins have been assembled for utilization in a variety of hosts including (but not limited to) *E. coli*. The basic expression node includes a low-activity promoter that transcribes the effector-responsive TF gene and the target cognate promoter that is activated by the TF/inducer complex. Promoters regulated by TFs of this kind endowed with various degrees of standardization include AlkS/PalkB (activated by n-octane) (100, 101), NahR/Psal (activated by salicylate) (102), CprK/Pcpr
(activated by o-chlorophenylacetic acid) (103), ChnR/
PchmB (activated by cyclohexanone) (104), and MekR/
PmekA (activated by methylethyl ketone) (105). The benzoate-inducible pair XylS/Pm, which stems from the toluene/m-xylene degradation plasmid pWW0 of P. putida, has been the basis of a large number of recombinant expression systems both in plasmids (106–108) and in transposon vectors (60, 102). It is noteworthy that the first rationally engineered feed-forward loop described in the literature was built on the basis of genetically rewiring two salicylate-responsive regulators for forming a transcriptional amplification cascade (109, 110). Another regulator of the pWW0 plasmid, the m-xylene-responsive TF called XylR, has also been adopted alone or in combination with the T7 RNAP for engineering whole-cell biosensors of aromatic compounds (111–113).

The list of chemically inducible TFs that can be exploited to the same or similar ends is very large (114, 115) and likely to be expanded in the near future. An added bonus of these expression systems is that owing to being encoded by promiscuous plasmids, they are likely to be operative in a large variety of hosts. Furthermore, when the devices are placed in a host without the corresponding metabolic genes, then the biochemical network of the recipient is usually blind to the effector, and the induction method has minimal or no impact (i.e., is orthogonal) on the host physiology.

Apart from transcriptional regulators, plasmids often also encode antisense RNAs (asRNAs) involved in various functions that can be recruited for engineering post-transcriptional circuits and expression systems based on RNA (116–118). To this end, a large number of cis-encoded asRNAs can be found in plasmids, phages, and transposons (119). Some of these asRNAs have been repurposed for engineering gene silencing (120, 121) and artificially halting bacteriophage infections (122, 123).

**REPOSITORIES OF PLASMID-RELATED PARTS AND DEVICES: THE SEVA COLLECTION**

One of the key synthetic biology tenets is the adoption of rules for the physical (and, wherever possible, functional) composition of biological parts to create devices with clear boundaries and defined input-output transfer functions (124). The advantages, downsides, and challenges of this view have been extensively discussed elsewhere (4) and will not be addressed here. Different synthetic biology communities have adopted different strategies for building increasingly complex genetic constructs on the basis of their own repositories of biological components. Table 2 shows a nonexhaustive list of web resources where some of the most popular materials available for synthetic biology purposes are compiled. Note that each collection has a different degree of curation and reliability and also a very diverse retrieve policy—from complete open access to pay-per-request to intricate material transfer agreement (MTA) procedures. One of the most accessed is the Registry of Standard Biological Parts (http://parts.igem.org). This collection, founded in 2003 at MIT, has accumulated a very large number of parts and devices that have been employed in educational synthetic biology projects.

![Chemical / environmental inducer](image1.png)

**FIGURE 3** The basic organization of synthetic biology constructs. Genetic devices are usually composed of a regulatory node that includes a transcriptional factor (an activator or a repressor) responsive to a physicochemical signal (e.g., a chemical inducer) that triggers its ability to stimulate transcription. $R$, gene encoding the regulatory protein; $P_R$, promoter of the regulatory $R$ gene; $T$, transcriptional terminator; UTR, untranslated regions of mRNA. The output of the regulatory node is a given level of PoPS (polymerase per second), which represents the count of RNAP molecules that pass through the promoter DNA each second. PoPS is then wired to a gene of interest (GOI), which is punctuated by 5′-UTR and 3′-UTR regions. doi:10.1128/microbiolspec.PLAS-0033-2014.f3
(iGEM: http://igem.org). Although the collection is mostly used and produced by undergraduate students, and quality control is not optimal, the large number of items and their inclusion as standardized parts has made the registry a very valuable resource for many synthetic biology endeavors (125–128).

The Standard European Vector Architecture database (SEVA-DB) deserves a separate comment (Fig. 4). This platform is both a web resource and a compendium of standardized and modular plasmid vectors. This collection originated in our efforts to expand the realm of synthetic biology toward Gram-negative bacteria different from E. coli but endowed with a biotechnological or environmental value. The idea and design behind this plasmid vector compilation is simple but has strict rules. This repository is composed of vectors that have up to four different modules decorating a fixed plasmid backbone. The rules of the SEVA compositional standard are as follows: (i) the DNA elements that comprise the plasmids have to be devoid of certain restriction sites (for specific details see reference 23; http://seva.cnb.csic.es), and (ii) the different modules have to be flanked by other sets of rare restriction sequences. In brief, the plasmid backbone is common to all constructs and includes an origin of transfer (oriT) and two transcriptional terminators (T1 and T0) to keep the adjacent DNA segments transcriptionally insulated. A typical plasmid from the collection is composed of an antibiotic resistance marker module, an origin of replication segment, and a cargo part. Optimally, vectors may also carry a genetic gadget to endow plasmids with additional properties. The antibiotic module comprises the six most common antibiotic resistance markers, and the origin of replication can be one out of nine choices with different functionalities in terms of host range (broad and narrow) and copy number (low, medium, and high copy), leaving their optimal combination to the user’s choice. The cargo endows functionalities to the plasmid, e.g., by means of a polylinker, different promoter-probe elements, or different expression systems. Since modules are punctuated by uncommon restriction sites, their exchange for creation of different plasmid combinations is straightforward. The collection is updated in real time and is increasingly adopting the synthetic biology open language (SBOL) for describing all vectors of the SEVA repository in a language that can be interfaced

### TABLE 2 Databases, repositories, and suppliers of synthetic biology parts and devices

<table>
<thead>
<tr>
<th>Database/repository webpage</th>
<th>Description/application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public/open source/nonprofit</td>
<td>The largest collection of standardized synthetic biology parts and devices. Not a curated repository</td>
</tr>
<tr>
<td><a href="http://parts.igem.org">http://parts.igem.org</a></td>
<td>Large number of resources for the open source synthetic biology community</td>
</tr>
<tr>
<td><a href="http://biobricks.org">http://biobricks.org</a></td>
<td>Curated repository of standardized plasmid vectors for engineering non-E. coli bacteria</td>
</tr>
<tr>
<td><a href="http://seva.cnb.csic.es">http://seva.cnb.csic.es</a></td>
<td>The largest distributor of vectors, plasmids, and cloned genes to the academic community</td>
</tr>
<tr>
<td><a href="http://www.addgene.org">http://www.addgene.org</a></td>
<td>Cloning vectors for E. coli, preserved as purified DNA samples</td>
</tr>
<tr>
<td><a href="http://www.shigen.nig.ac.jp">http://www.shigen.nig.ac.jp</a></td>
<td>Strains, plasmids, libraries, and other biological materials contributed by Belgian researchers</td>
</tr>
<tr>
<td><a href="http://bccm.belspo.be">http://bccm.belspo.be</a></td>
<td>Curated collection of vectors (mostly plasmids) for biosystems engineering</td>
</tr>
<tr>
<td><a href="http://www.bios.uni-freiburg.de/toolbox">http://www.bios.uni-freiburg.de/toolbox</a></td>
<td>Compilation of databases of mobile genetic elements, plasmids, and vectors</td>
</tr>
<tr>
<td><a href="http://www.science.co.il">http://www.science.co.il</a></td>
<td>Repository for plasmid clones and collections with links to collections of individual researchers</td>
</tr>
<tr>
<td><a href="http://dnasu.org">http://dnasu.org</a></td>
<td>Software and platform for managing information about biological parts and host strains</td>
</tr>
<tr>
<td><a href="https://public-registry.jbei.org">https://public-registry.jbei.org</a></td>
<td>Biological materials and tools with emphasis on infectious diseases</td>
</tr>
<tr>
<td><a href="http://www.beiresources.org">http://www.beiresources.org</a></td>
<td>Commercial/private</td>
</tr>
<tr>
<td><a href="https://www.snapgene.com">https://www.snapgene.com</a></td>
<td>Software for simulation and guiding of DNA cloning in different types of vectors</td>
</tr>
<tr>
<td><a href="http://cms.plasmidfactory.com/en/">http://cms.plasmidfactory.com/en/</a></td>
<td>Plasmid vectors made à la carte for specific bioengineering projects</td>
</tr>
<tr>
<td><a href="http://www.lgcstandards-atcc.org">http://www.lgcstandards-atcc.org</a></td>
<td>ATCC-linked repository of biological materials including vectors and strains</td>
</tr>
<tr>
<td><a href="http://www.promega.es/products/vectors">http://www.promega.es/products/vectors</a></td>
<td>Tailored DNA clones, vectors, and bioinformatic services</td>
</tr>
<tr>
<td><a href="http://www.biocline.us">http://www.biocline.us</a></td>
<td>Large collection of vectors and other genetic tools for gene expression and reporter systems</td>
</tr>
<tr>
<td><a href="http://plasmid.med.harvard.edu/PLASMID/">http://plasmid.med.harvard.edu/PLASMID/</a></td>
<td>Commercial provider of molecular tools for optimization of gene expression</td>
</tr>
<tr>
<td><a href="https://www.neb.com">https://www.neb.com</a></td>
<td>Harvard-based provider of sequence-verified plasmid constructs</td>
</tr>
<tr>
<td><a href="http://www.lifetechologies.com">http://www.lifetechologies.com</a></td>
<td>Materials and general support for iGEM projects following standardized assembly methods</td>
</tr>
<tr>
<td><a href="http://www.genscript.com">http://www.genscript.com</a></td>
<td>Chemical synthesis, cloning, and sequence verification of genetic sequences à la carte</td>
</tr>
<tr>
<td><a href="http://www.labguru.com/features/molecular/plasmids">http://www.labguru.com/features/molecular/plasmids</a></td>
<td>Plasmid vector database along with tools for cloning projects and plasmid retrieval</td>
</tr>
</tbody>
</table>
with other platforms and understood by robotic DNA synthesis platforms (129). While the manifest destiny of genetic constructs is having them made through chemical synthesis (for which assembly vectors and material DNA repositories may not be necessary any longer), these collections of synthetic biology building blocks will still be required for some time, in particular for addressing basic biological questions and engineering simple phenotypes in non-\textit{E. coli} bacteria.

**OUTLOOK**

As discussed above, environmental plasmids offer synthetic biology a diverse palette of gene-encoded activities that once excised from its natural context, minimized, and standardized, become an extraordinary asset for a wide range of applications. But there is surely more that plasmid biology can deliver beyond providing building blocks for genetic constructs. First, as mentioned earlier, plasmids, especially those that are promiscuous, move through the microbial population, and their encoded genes are to be expressed in different types of hosts. What determines such ability? Both the plasmid-borne promoters involved, their regulation, and the proteins they encode must operate in a fashion that is relatively independent of the host, thereby creating a natural case of implant-chassis orthogonality that must be ruled by principles that deserve further research. Second, the acquisition of environmental plasmids encoding new metabolic capabilities is merely a natural case of biochemical engineering where the preexisting molecular network of the receiving cells has to cope with implantation of a new genetic and enzymatic system (97). Studies of such scenarios (e.g., the interplay of the pWW0-encoded system for toluene and \textit{m}-xylene degradation and the \textit{P. putida} host) reveal both the type of biochemical conflicts that such situations create and how bacteria have found a solution to them (130, 131).
More instances of this sort could be worth examining for the sake of handling the problem of implant-chassis retroactivity, one of the central challenges in biological engineering (132, 133).

Finally, once plasmids make it to a new host, their encoded proteins have to be placed somewhere in the tridimensional architecture of the receiving cells, where molecular crowding imposes considerable constraints. Are plasmid-encoded functions, cognate protein structures, and processes associated with such three-dimensional addresses? It will be important to classify plasmid-encoded proteins into those that distribute evenly versus those that go to a specific cell site and to see what the difference is, thereby understanding the principles that may rule such a positioning. In this sense, promiscuous environmental plasmids will continue enriching the material toolbox for engineering biological systems as well as providing new conceptual insights on how to program microorganisms to do (new) things that they normally do not do.

ACKNOWLEDGMENTS

The work in the authors’ laboratory was supported by the BIO program of the Spanish Ministry of Economy and Competitiveness, the ST-FLOW, ARISYS, and EVOPROG contracts of the EU, the ERANET-IB Program, and funding from the Autonomic Community of Madrid (PROMPT).

Conflicts of interest: We declare no conflicts.

REFERENCES


