InfoBiology by printed arrays of microorganism colonies for timed and on-demand release of messages

Manuel A. Palacios[a], Elena Benito-Peña[a], Mael Manesse[a], Aaron D. Mazzeo[b], Christopher N. LaFratta[a,c], George M. Whitesides[b], and David R. Walt[a,1]

[a]Department of Chemistry, Tufts University, 62 Talbot Avenue, Medford, MA 02155; [b]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138; and [c]Chemistry Department, Bard College, P.O. Box 5000, Annandale-on-Hudson, NY 12504

Edited by Chad A. Mirkin, Northwestern University, Evanston, IL, and approved August 16, 2011 (received for review June 15, 2011)

This paper presents a proof-of-principle method, called InfoBiology, to write and encode data using arrays of genetically engineered strains of *Escherichia coli* with fluorescent proteins (FPs) as phenotypic markers. In InfoBiology, we encode, send, and release information using living organisms as carriers of data. Genetically engineered systems offer exquisite control of both genotype and phenotype. Living systems also offer the possibility for timed release of information as phenotypic features can take hours or days to develop. We use growth media and chemically induced gene expression as cipher keys or "biociphers" to develop encoded messages. The messages, called Steganography by Printed Arrays of Microbes (SPAM), consist of a matrix of spots generated by seven strains of *E. coli*, with each strain expressing a different FP. The coding scheme for these arrays relies on strings of paired, septenary digits, where each pair represents an alphanumeric character. In addition, the photophysical properties of the FPs offer another method for ciphering messages. Unique combinations of excited and emitted wavelengths generate distinct fluorescent patterns from the Steganography by Printed Arrays of Microbes (SPAM). This paper shows a new form of steganography based on information from engineered living systems. The combination of bio- and "photociphers" along with controlled timed-release exemplify the capabilities of InfoBiology, which could enable biometrics, communication through compromised channels, easy-to-read barcoding of biological products, or provide a deterrent to counterfeiting.

biotechnology | infochemistry | information technology

The intrinsic high information content and information flow in biological systems has the potential to be used to translate nonbiological genetically encoded information into an easily read phenotypic signal. In this context, genetically engineered systems are of particular utility because they enable exquisite control of both genotype and phenotype (1). Here we describe the use of living organisms as the carriers of encoded messages. Phenotypic features have previously been used as cipher keys for the identification of individuals. Biometric ciphers, such as fingerprint, iris, and retinal scans, are examples of ways in which the unique phenotypic characteristics of individuals can be used to control access to facilities or data (2). Although biometrics have found their way into "real-world" applications, biometric ciphers only function as cipher keys and do not play a role in the storage, transmission, or encoding of data. Examples of information embedded in biological systems include the insertion of synthetic data-encoding DNA (nonprotein coding) for trademark and watermarking purpose (3, 4) and for long-term information storage (5–8). Although such systems seem convenient for high-density applications of data storage, decoding high-density information from nonprotein coding DNA requires sophisticated sequencing capabilities for data readout.

We have previously employed chemical methods for encoding, storing, and sending information using a method dubbed "Infochemistry" (9–11). In this paper, we develop a unique way of transmitting information called InfoBiology that uses living organisms for these functions. This work constitutes an initial step to combine biochemical signals with information theory to produce an alphanumeric message (Fig. 1*A*).

Results and Discussion

We use cytosolic expression of fluorescent proteins (FPs) in *Escherichia coli* as a phenotypic marker to encode messages. The levels and timing of expression of proteins can be controlled by several biological inputs (or biociphers); e.g., bacterial strain, type of vector (high- or low-copy origin of replication), growth medium, promoter site, and maturation time of fluorescent proteins. For our proof-of-principle experiments, we prepared different strains of *E. coli* that were engineered to express high-copy numbers of seven different FPs: GFPuv, AmCyan, ZsGreen, ZsYellow, mOrange, tdTomato, and mCherry under control of the bacteriophage-T7 promoter (*SI Text*). This series of FP encoding vectors contain the ampicillin-resistant gene as a selective marker. We used two different host strains of *E. coli* to express the FPs. First, we transformed BL21(DE3)pLysE *E. coli* cells with the series of FP encoding vectors mentioned above. BL21(DE3) pLysE cells contain the gene encoding for T7 RNA polymerase under the control of the *lacUV5* promoter, allowing expression of T7 RNA polymerase to be induced by isopropyl β-D-1-thiogalactopyranoside (IPTG). The BL21 strains yield an "on-demand" system as these strains require induction to develop a potential message properly. Next, TOP10 *E. coli* cells were transformed with the same FP encoding vectors to generate an "induction free" system. TOP10 cells do not contain the gene encoding for T7 RNA polymerase and therefore are not sensitive to IPTG induction. TOP10 cells are, however, engineered to allow stable replication of high-copy number plasmids. The high concentration of the plasmid allows for a high background "leaky" expression of the FPs, which can easily be detected by fluorescence imaging after 48 h of incubation under ambient conditions (Fig. S1).

Infobiological data can be organized in different ways to convey a message. In the case of the previously reported infofuse (11), the spatially arrayed data along the infofuse resulted in a timed sequence of pulses of IR emission, which were then converted into a message. In the microorganism-based platform described here, we array the data in spatial domains to form a matrix of fluorescent colonies. To produce the Steganography by Printed Arrays of Microbes (SPAM), fluorescent bacterial strains

Author contributions: M.A.P., E.B.-P., M.M., A.D.M., C.N.L., G.M.W., and D.R.W. designed research; M.A.P., E.B.-P., and M.M. performed research; M.A.P. analyzed data; E.B.-P. and M.A.P. illustrated the paper; and M.A.P. and D.R.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: David.Walt@tufts.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1109554108/-/DCSupplemental.

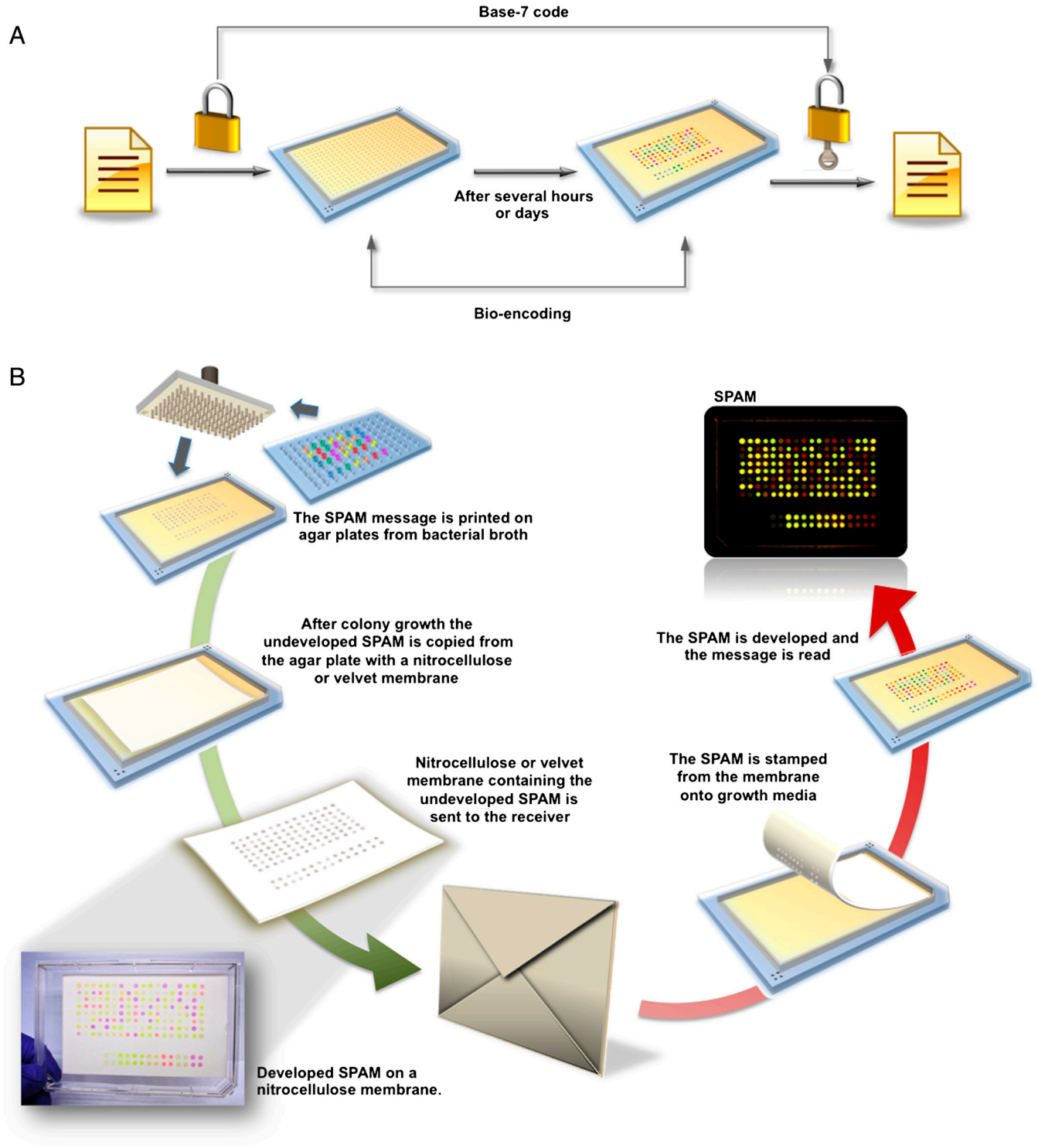

Fig. 1. (*A*) Schematic illustration of the information workflow in a bioencoding system. The sender encodes the message using a septenary code. The SPAM message is developed under predetermined growth conditions and read-out with a predetermined set of excitation/emission wavelengths, which constitute the biocipher and photocipher keys, respectively. Finally, the receiver compares the output with a predetermined code. (*B*) Scheme showing the preparation and read-out of a SPAM. The green arrow follows the sender's actions to prepare a SPAM, whereas the red arrow follows the receiver's actions to develop a SPAM. First, broth containing fluorescent bacteria is pipetted into a microtiter plate. Second, a multiblot pin replicator is used to transfer a small volume of the broth from each well onto a target plate containing the appropriate growth media. After the undeveloped SPAM is grown, it can be transferred to a nitrocellulose or velvet membrane for delivery. The receiver stamps the SPAM onto an appropriate growth medium, develops the signal, and reads the SPAM message. Note that the "undeveloped" SPAM does not have a clear color read-out because protein expression has not yet been induced. For illustration, an image of a nitrocellulose membrane containing a "developed" SPAM is shown at bottom left.

are first grown in selective broth media and are then transferred to a source microtiter plate. Subsequently, a multiblot pin replicator is used to transfer 0.1 μL of the bacterial broth onto target agar plates (Fig. 1*B*). Alternatively, after growing the colonies on agar, a nitrocellulose or velvet transfer membrane can be used to harvest the message from the agar plate. After copying the array

of colonies, the membrane containing the message can be used to regrow the message; e.g., in a different growth medium. Depending on the application or setting, the membranes have the potential to be used as a carrier to store and/or distribute messages that is more convenient than agar plates.

Given the photophysical properties of the FPs, the signal of each fluorescent protein can be different depending on which excitation light source and emission filters are used (Fig. S2). For our proof-of-principle example, we imaged the array with a filter combination (photo-cipher: $\lambda_{exc} = 470$ nm; $\lambda_{em} > 535$ nm) that gave the highest information density. Fig. S3 (*SI Text*) shows clustering of seven distinct microbial FP signals when plotting the green versus red channels of the color image of a SPAM. This information density allows for the data in the SPAMs to be encoded in a base-7 (septenary) encoding scheme (Fig. 2*A*). Each character is encoded by a pair of two septenary digits for a total of 49 (7^2) alphanumeric characters. Fig. 2*B* shows SPAM of 144 colonies encoding a message containing 70 characters: "this is a bioencoded message from the walt lab at tufts university 2011."

As mentioned, TOP10 and BL21 cells were transformed to produce systems for delayed-release and an on-demand delivery of messages, respectively. Fig. 2*C* shows the timeline of an array of TOP10 fluorescent colonies growing at room temperature. Even though background leakage of the FP's expression is evident after 18 h, the signal intensity remains very low; at this time, fluorescent identification of the FPs is difficult. After 48 h, the signal is strong enough for the message to be decoded. Fig. 2*D* shows that in the case of BL21, the fluorescent bacterial strains also show some basal level of FP expression, but it is still difficult to identify the FP's emission signal correctly. After IPTG induction, overexpression of the FPs takes place, but the message does not fully develop until after approximately 8 h. This feature is due to the intrinsic clock associated with the fluorescent protein maturation time, which varies from protein to protein (12). Although there is delay for the on-demand system to deliver the message, the receiver can trigger the development of the message in a controlled manner using IPTG induction. It is worth noting that there are FP mutants capable of changing their emission properties over time (13, 14). These mutants would add an inherent security measure by self-deleting the message as it develops; similar to the way the *Mission Impossible* recording self-destructed.

One of the most useful features of this data storage/encoding system is the possibility of using selection markers as cipher keys to develop the correct message. Selection markers are genes that are introduced to an organism to provide a method of artificial selection when the organism is grown in a particular medium.

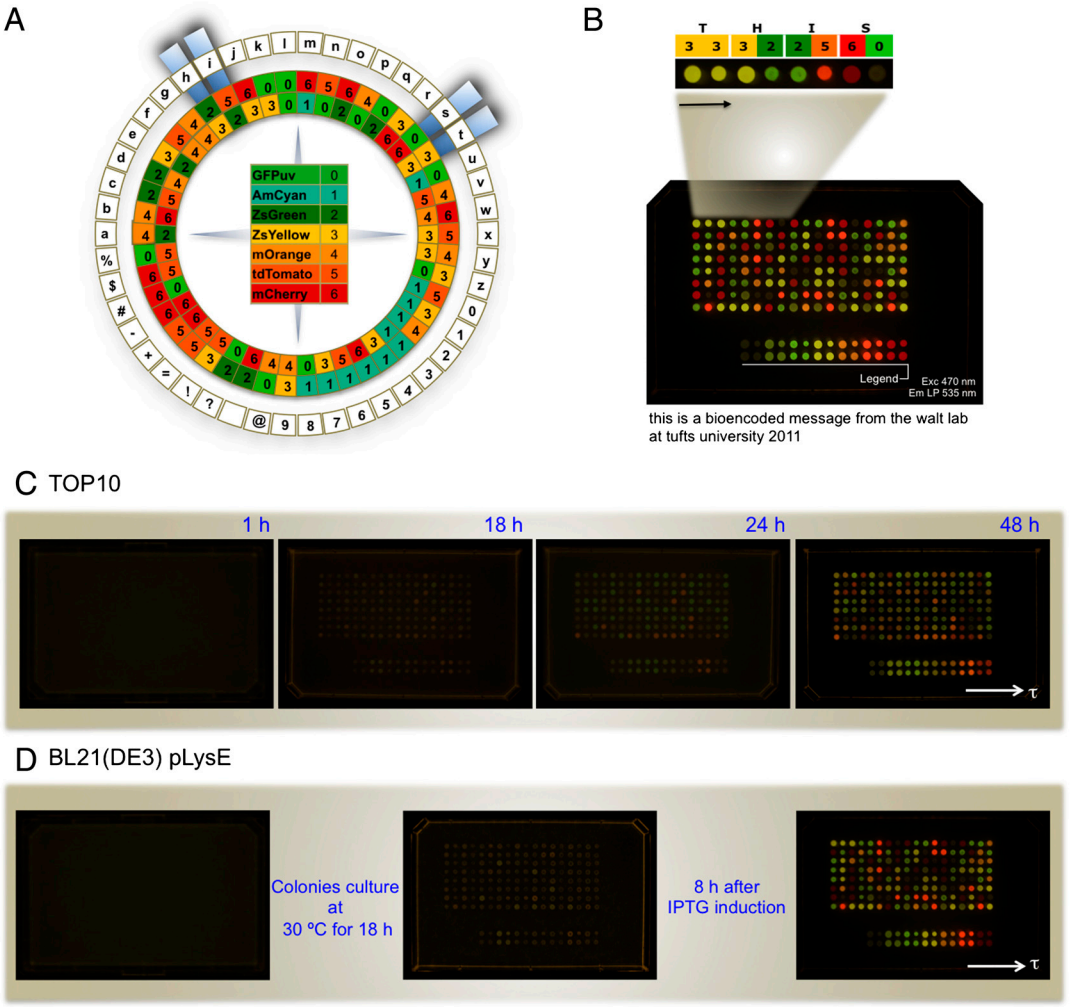

Fig. 2. (*A*) Septenary alphanumeric code. (*B*) Fluorescence image of a SPAM consisting of 144 colonies, which encodes a message containing 72 characters. The message is read from left to right along lines that read top to bottom. The message reads "this is a bioencoded message from the walt lab at tufts university 2011." (*C*) Fluorescence images of TOP10 fluorescent strains showing an array of colonies developed in approximately two days at room temperature. (*D*) Fluorescence images of BL21(DE3)pLysE fluorescent strains showing that after growth of colonies, the FP expression can be induced by IPTG. Maturation of FPs takes approximately 8 h.

Selection markers are commonly used as an indicator for the success of DNA transfection. One commonly used selection marker in bacteria confers antibiotic resistance to the cell. To demonstrate the possibility of using antibiotic resistance as a cipher key, all seven FP genes were cloned into kanamycin-resistant expression vectors containing a T7 promoter. Fig. 3 shows a message encoded with multiple FPs and resistance genes, which leads to a different message depending on the growth medium employed. When ampicillin is used as the cipher, the SPAM **A** message correctly reads, "this is a bioencoded message from the walt lab at tufts university 2011." When kanamycin is used as the cipher, the SPAM **B** message reads, "you have used the wrong cipher and the message is gibberish." The last SPAM **C** does not produce a message because the combination of two FPs results in color emissions that do not correlate with the previously described septenary alphabet.

The apparent low information-density of the SPAMs is one of its major drawbacks. The most straightforward way to increase the information density would be by printing smaller features (15, 16). Furthermore, the different layers of information given by selective markers (Fig. 3) can also be used to increase the information density in the array; i.e., the information density could be multiplied n-times, with n being the number of antibiotic resistant genes available. Additionally, cell lines could be designed to be resistant to a combination of antibiotics adding more dimensionality into the biological domain. Furthermore, each colony in a SPAM is a higher-order structure formed by millions of individual cells that are engaged in quorum sensing; i.e., microbial consortium could be used to tune the expression of a fluorescent protein or possibly any other phenotype (17–19). Consequently, expanding the information density in the biological domain is limited by the number of endogenous or exogenous genes that could be engineered into an organism. In the future, another approach to increase the information density could be to encode multidimensional phenotypes; i.e., using combinations of FP chimeras. Cell lines expressing chimera FPs will generate SPAMs that can display different emission patterns that depend on the excitation wavelength used to read the message. Thus, the information density will increase in both the biological and physical domains.

This work demonstrates the use of biological systems to store and deliver information and, is a unique example of using phenotypic characteristics of living organisms to carry and deliver an alphanumeric message. Any distinguishable phenotype could potentially be used as a signaling mechanism, as long as the expression is reliable. For this proof-of-principle, we chose to engineer laboratory strains of *E. coli*, because they are relatively straightforward and safe to handle. However, the development and viability of these microorganisms are very sensitive to environmental conditions. Future work will include extending the platform to more robust microorganisms, such as yeast (20) or endospore forming Gram-positive bacteria such as *Bacillus subtilis* (21, 22). Using yeast will open possibilities for other types of selective markers, such as auxotrophy and/or hormonal signaling for gender selection, because yeast can reproduce asexually. Sexual reproduction could be used to add yet another level of complexity to the information system. For example, a pair of binding proteins fluorescently labeled as a fluorescence resonance energy transfer (FRET) pair could be separately cloned into *a* and α strains. The mating product of these two strains will yield the FRET emission, adding a very simple optical logic gate to the system. Our labs have also begun exploring the concept of using multicellular organisms, such as plants, that could offer a longer timed-release clock and could add other useful phenotypic features as read-out signals. Finally, the large number of adjustable parameters (FPs, promoters, media, excitation wavelength, release time, etc.) makes our infobiological system a strong platform from which to explore the field of InfoChemistry.

Materials and Methods

Experimental details on molecular cloning are included in *SI Text*. All bacterial strains and plasmids are listed in Tables S1 and S2. The *E. coli* strains were routinely grown in Luria–Bertani agar (LBA) (EMD Chemicals) at 37 °C for 48 h and Terrific Broth (TB) (Tecknova) at 37 °C with aeration and vigorous shaking (approximately 300 rpm) for 4–6 h. For genomic DNA extraction, the strains were grown overnight and were maintained as frozen stocks at −80 °C in TB containing 50 $\mu g\,L^{-1}$ of correspondent antibiotic plus 20% (v/v) glycerol (Sigma–Aldrich). Antibiotics (Sigma–Aldrich) were added to the appropriate media at 50 $\mu g\,mL^{-1}$ of both ampicillin and kanamycin. The plasmids pGFPuv, pAmCyan, pZsGreen, pZsYellow, pmOrange, ptdTomato, and pmCherry (Clontech Laboratories) were used as source DNA for cloning and expression. The plasmids pYes3/CT (Invitrogen) and pQE-T7-2 (Qiagen) were used as host vectors for the construction of cloned *E. coli* strains containing the gene templates of the selected recombinant FPs (Table S2).

Preparation of SPAMs. To produce the colony matrices, single cell colonies of fluorescent bacterial strains were grown in their corresponding selection media, washed with 1× PBS to eliminate any residual antibiotic, resuspended in TB, and transferred to a source microtiter plate. A 96-pin Multi-Blot™ replicator (V&P Scientific Colony Copier™ VP 409) was used to inoculate the arrays of colonies on LB agar casted on an Omni Tray (Nunc), with the appropriate antibiotics. The SPAMs were then incubated accordingly. For the IPTG induction experiments, the arrays of colonies were prepared following the same procedure. After 18 h of incubation, a 10 mM solution of IPTG (Sigma) was sprayed onto the colony arrays and the plates were placed back in the incubator. SPAM replication was carried out with cotton velvet (Cora

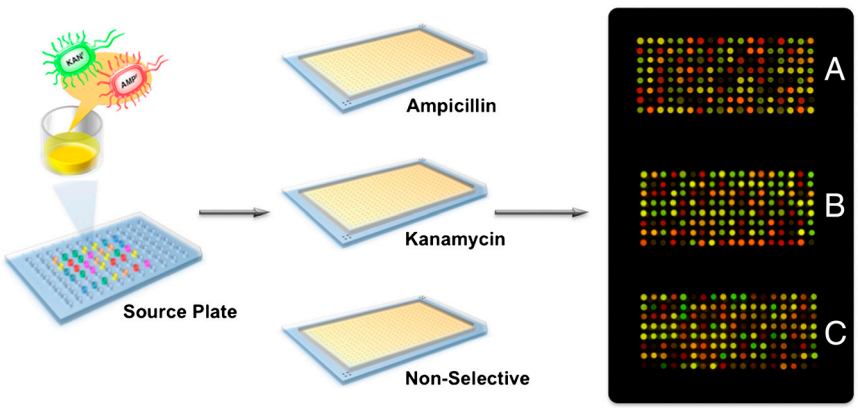

Fig. 3. Scheme of SPAM messages developed from three different growth media. Selective markers are used as cipher keys with three possible outcomes. Using ampicillin as a cipher key gives SPAM **A**, which reads "this is a bioencoded message from the walt lab @ tufts university 2010." Using kanamycin as a cipher key gives SPAM **B**, which reads "you have used the wrong cipher and the message is gibberish." The last matrix **C** does not produce a message because the combination of two FPs results in colors that do not correlate with the septenary alphabet.

Styles Needles N Blocks) or nitrocellulose membranes (GE Water and Process Technology).

Image Acquisition. Preliminary studies (Fig. S3) determined that the combination of $\lambda_{exc} = 470$ nm and $\lambda_{em} > 535$ nm shows seven discernible signal from all seven FPs. We used a Safe Imager™ 2.0 Blue Light Transilluminator (Invitrogen) equipped with an array of blue light-emitting diodes (LEDs) (approximately 470 nm) and an amber filter unit, for which the cut-off is shown in Fig. S2, *Right* (InvFilter). The images of the SPAM were acquired using a digital single-lens reflex (DSLR) color camera (Nikon D7000 equipped with a Nikkor lens 18–200 mm, F/3.5–5.6) or alternatively the camera of a smartphone (Apple iPhone 4). Fig. S4 shows the comparison between the images acquired with both detection systems.

ACKNOWLEDGMENTS. We thank Dr. Lorena B. Harris and Dr. Kristina H. Schmidt and her lab at University of South Florida for the assistance with the cloning and for providing some vectors for this project. Also, we thank Aaron Phillips and Stephanie M. Schubert for insightful discussions during the preparation of this manuscript. This work was supported by Defense Advanced Research Projects Agency Award W911NF-07-1-0647 under the Chemical Communications program. E.B.P. also acknowledges support from the Spanish Foundation for Science and Technology (FECYT).

1. Glick BR, Pasternak JJ, Patten CL (2010) *Molecular Biotechnology: Principles and Applications of Recombinant DNA* (ASM, Washington, DC), 4th Ed,, pp 1–1000.
2. Xi K, Hu J (2010) Bio-cryptography. *Handbook of Information and Communication Security*, eds P Stavroulakis and M Stamp (Springer, New York), pp 129–157.
3. Arita M, Ohashi Y (2004) Secret signatures inside genomic DNA. *Biotechnol Prog* 20(5):1605–1607.
4. Gibson DG, et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329(5987):52–56.
5. Bancroft C, Bowler T, Bloom B, Clelland C (2001) Long-term storage of information in DNA. *Science* 293(5536):1763c–1765.
6. Clelland CT, Risca V, Bancroft C (1999) Hiding messages in DNA microdots. *Nature* 399(6736):533–534.
7. Smith GC, Fiddes CC, Hawkins JP, Cox JPL (2003) Some possible codes for encrypting data in DNA. *Biotechnol Lett* 25(14):1125–1130.
8. Yachie N, Sekiyama K, Sugahara J, Ohashi Y, Tomita M (2007) Alignment-based approach for durable data storage into living organisms. *Biotechnol Prog* 23(2):501–505.
9. Kim C, Thomas SW, III, Whitesides GM (2010) Long-duration transmission of information with infofuses. *Angew Chem Int Ed Engl* 49(27):4571–4575.
10. Hashimoto M, et al. (2009) Infochemistry: Encoding information as optical pulses using droplets in a microfluidic device. *J Am Chem Soc* 131(34):12420–12429.
11. Thomas SW, et al. (2009) Infochemistry and infofuses for the chemical storage and transmission of coded information. *Proc Natl Acad Sci USA* 106(23):9147–9150.
12. Shaner NC, Steinbach PA, Tsien RY (2005) A guide to choosing fluorescent proteins. *Nat Methods* 2(12):905–909.
13. Subach FV, et al. (2009) Monomeric fluorescent timers that change color from blue to red report on cellular trafficking. *Nat Chem Biol* 5(2):118–126.
14. Terskikh A, et al. (2000) "Fluorescent timer": Protein that changes color with time. *Science* 290(5496):1585–1588.
15. Xu T, et al. (2004) Construction of high-density bacterial colony arrays and patterns by the ink-jet method. *Biotechnol Bioeng* 85(1):29–33.
16. Merrin J, Leibler S, Chuang JS (2007) Printing multistrain bacterial patterns with a piezoelectric inkjet printer. *PLoS ONE* 2(7):e663.
17. Brenner K, Karig DK, Weiss R, Arnold FH (2007) Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *Proc Natl Acad Sci USA* 104(44):17300–17304.
18. Tamsir A, Tabor JJ, Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical. *Nature* 469(7329):212–215.
19. Regot S, et al. (2011) Distributed biological computation with multicellular engineered networks. *Nature* 469(7329):207–211.
20. Kitano H (2004) Biological robustness. *Nat Rev Genet* 5(11):826–837.
21. Romero D, Aguilar C, Losick R, Kolter R (2010) Amyloid fibers provide structural integrity to Bacillus subtilis biofilms. *Proc Natl Acad Sci USA* 107(5):2230–2234.
22. Traag BA, et al. (2010) Do mycobacteria produce endospores? *Proc Natl Acad Sci USA* 107(2):878–881.