GENOMIC AND METAGENOMIC APPROACHES TO NATURAL PRODUCT CHEMISTRY

A Dissertation

by

SCOTT EDWARD ANGELL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2008

Major Subject: Chemistry

GENOMIC AND METAGENOMIC APPROACHES TO NATURAL PRODUCT CHEMISTRY

A Dissertation

by

SCOTT EDWARD ANGELL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Coran M. H. Watanabe
Committee Members,	Eric E. Simanek
	Daniel A. Singleton
	Donald W. Pettigrew
Head of Department,	David H. Russell

May 2008

Major Subject: Chemistry

ABSTRACT

Genomic and Metagenomic Approaches

to Natural Product Chemistry. (May 2008)

Scott Edward Angell, B.S., Texas A&M University

Chair of Advisory Committee: Dr. Coran M. H. Watanabe

For many years, natural products have been a primary source of new molecules for the treatment of disease, and microorganisms have been a prolific source of these molecules. Recent studies have indicated, however, that many biosynthetic pathways are present in organisms for which no natural product can be associated, and only a small fraction of the microbial life present in the environment can be grown in culture. This indicates that if methods could be developed for the isolation of these pathways and production of their target molecules in heterologous hosts, great numbers of potentially valuable compounds might be discovered.

In these investigations, large insert libraries of two microorganisms were constructed, one a bacterial artificial chromosome (BAC) library, the other a fosmid library, and two large insert fosmid libraries were constructed with DNA isolated from marine environmental samples. A mathematical formula was derived to estimate probabilities of cloning intact biosynthetic pathways with large insert genomic libraries and tested with a computer simulation. This indicated that even large pathways could be cloned intact in large insert libraries, provided there was an adequate size difference between the target pathway and the library inserts, and there was a concomitant increase in the size of the library with the targeting of these larger pathways. In addition, an investigation into a mixed marine culture sample lead to the identification of an unusual relationship between two bacteria for which extended co-culture leads to the production of pyocyanin. However, no useful biosynthetic pathways were located within the genomic libraries.

It is concluded that significant improvements would be required to make this approach feasible for larger scale investigations. It is further concluded, on the basis of recent developments in the field, including a reduction in the cost of sequencing, improvements in

techniques of whole-genome shotgun sequencing, and the development of recombination based cloning, that the employment of mass sequencing efforts and sequence-driven, recombination-based cloning, might prove to be a more fruitful and efficient alternative to large-insert library construction for the isolation and expression of these pathways. A possible paradigm for the cloning of pathways on the basis of this technology is proposed.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Orthodox natural product isolation chemistry proceeds through the sequential steps of metabolite extraction, assay-guided fractionation, purification to homogeneity, and identification, usually through various spectroscopic methods (173). This approach has produced many beneficial compounds over the years and continues to be the method of choice to this day (136, 138). It has been expanded through advances in culturing techniques (3, 28, 40, 93), which have increased the variety of organisms available in culture and allowed greater access to metabolites for which production is often regulated by arcane signaling pathways, as well as through technological advances in spectroscopy and chromatography, which in some cases have even integrated separation and identification (39). These advances have improved throughput and sensitivity, reduced the necessary sample size for characterization, and expanded the scope of the natural world amenable to characterization.

The genetic revolution has made many new techniques and types of information available, in particular the ability to glean sequence information about the organisms and biosynthetic pathways present in a biological sample (199). The information revealed to date has changed the way we look at the natural world. 16S rRNA sequencing of environmental samples has shown that science has barely scratched the surface of the prokaryotic world, with orders of magnitude more forms of life present in a mundane tablespoon of dirt or a typical liter of seawater than can be isolated through traditional approaches (Table 1) (4, 93, 175). Genomic sequencing of even well characterized organisms has revealed unknown pathways for which no known natural products can be associated (70, 107, 199, 220). Our perspective is in a constant state of change as more and more sequence information becomes available.

This has not gone unnoticed among natural product chemists, and indeed many other scientists. As productive as traditional techniques have been, and as successful as improvements have been over the years, many researchers have turned their attention to the techniques of molecular biology in an attempt to exploit the unexplored chemistry of the life forms which have been inaccessible up until this point. This new area of study, called metagenomics (73), aims to isolate and study the DNA of organism cohorts which reside together in communities defined by

---

This dissertation follows the style of *Applied and Environmental Microbiology*.

TABLE 1. Culturability of bacteria from different habitats. From Amman, et al. (4)

Habitat	Culturability (%)
Marine, seawater	0.001 to 0.1
Freshwater	0.25
Mesotrophic lake	0.1 to 1
Unpolluted estuarine	0.1 to 3
Activated Sludge	1 to 15
Sediments	0.25
Soil	0.3

particular sets of environmental parameters. It is hoped that with the aid of this new information and with the development of new techniques, researchers will one day be able to readily access the genetic and biosynthetic capabilities of virtually any organism without the burden of having to actually observe and manipulate that organism at all. To some extent there has already been considerable success along these lines.

THE BIRTH OF METAGENOMICS

For quite some time it has been known that the "culturable" fraction of microorganisms was considerably smaller than the total number present in environmental samples, as deduced by comparisons of the cellular morphology of those organisms grown in culture with microscopic images taken of the cellular mixture present in the source sample (4, 40). Unfortunately, taxonomic studies were nearly impossible since the information available on each individual was necessarily limited, and traditional classification data, such as optimal growth conditions, carbon source utilization, and GC-content of DNA, were completely unavailable for specimens which could not be cultured (4). As it turned out, these data probably would not have been all that helpful, since even many "well characterized" organisms were found to be wrongly classified once additional sequencing data became available (215). For the most part, researchers could do little but speculate on the basis of what little data were available (215).

Sequencing technology began to change this situation, as it became apparent that biopolymers such as DNA and proteins contained an evolutionary record in the form of mutations acquired over the years (229), and by the 1960's protein sequence data were being used to produce phylogenetic trees of macroscopic organisms in good agreement with the fossil record (54). By the 1970's, new technologies and data lead to drastic changes in perceptions of the complexity of life and its taxonomic organization. Sequencing of 16S rRNA, initially a laborious task but later simplified by rapid PCR based sequencing methodology, became the dominant tool for determining taxonomic classification among bacteria (4, 103, 144). The ribosome, which is conserved across all classes of life, proved to be an ideal tool to observe

relationships between even distantly related species, and this depth of insight had profound consequences for the science of bacterial taxonomy and evolution (144, 215). Pioneering research into the most ancient evolutionary events became possible, and in 1977 Carl Woese famously shattered centuries of prevailing wisdom by redrawing the highest levels of the taxonomy of life and calling into question the entire paradigm of the prokaryotic classification system, primarily on the basis of this data but with other biochemical support (56, 215-217).

Norman Pace began applying 16S RNA sequencing technology to environmental samples in the mid-1980's (191). Later, the same group cloned DNA directly from a marine sample with the intention of isolating 16S rDNA sequences for taxonomic investigation of plankton (177). These experiments and many others confirmed what microbiologists had suspected all along: that there are, in fact, many, many more microbial species present in natural environments than are culturable under laboratory conditions (4, 93, 175, 196, 203). A more surprising finding was that many are not even that closely related to isolated strains (93, 175). It seems that there are not just new species waiting to be discovered, but entire families.

Since that time, many groups have begun producing genomic libraries from environmental samples, and not just with a view towards the study of taxonomy. This particular aspect of metagenomic study came to be dominated by Edward DeLong, a student of Norman Pace. Other groups have joined the fray with an eye towards the pursuit of such disparate interests as the study of symbiotic interactions (72), the discovery of novel or improved biocatalysts (113, 206, 211), environment specific metabolic systems profiling (196), natural products (18-25, 41, 63, 167, 207), and, it seems, some have joined in for the simple joy of sequencing novel DNA to their heart's content (203).

METAGENOMICS AND NATURAL PRODUCT DISCOVERY

The investigators primarily responsible for the trend towards looking for new natural products from metagenomic sources are John Clardy and Jo Handelsman, who teamed up in the late 1990's and early 2000's to produce a flurry of papers on the subject (18-20, 22, 24, 63, 73, 167). Initially, results looked promising. While not all of the natural products so "discovered" were new (20, 63), a new class of acyl-amino acid antibiotics was found, vindicating the technique at least in principle (Fig. 1) (18, 22-24). Other groups found similar results (207). But it was clear that the method had major shortcomings. Enormous libraries were screened to find these few compounds which were all found to be from small gene clusters, typically less

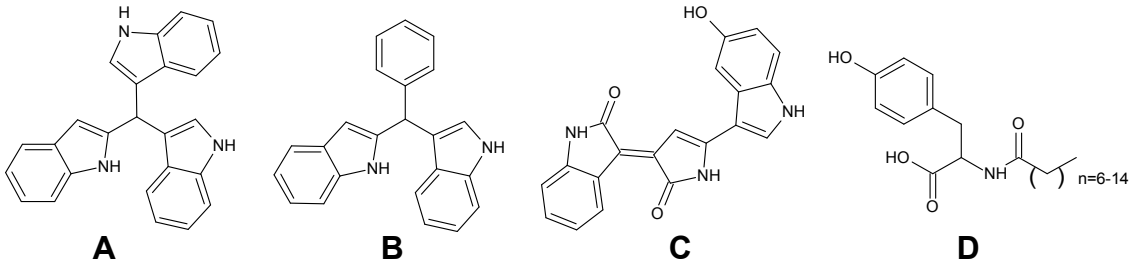


FIG. 1. Representative natural products isolated by Clardy and co-workers from metagenomic DNA expression. (20, 22, 63). (a)Turbomycin A, (b) Turbomycin B, (c) Violacein, (d) N-Acyltyrosine

than 10 kilobases in length. It was clear that many promising pathways were being overlooked in the screening process, which in these initial experiments was primarily through activity screening depending on heterologous expression in *E. coli*. Within a few years it was apparent that the technique would need some major innovations if it was to become widely applicable.

The first shortcoming to be targeted was the insert size of the library. The first libraries to be produced were bacteriophage-λ based libraries, which by necessity are limited in size by the viral vector used for transformation to no greater than 50 kb (169). Many natural product biosynthetic pathways are known to be significantly larger than this (178, 219), so this type of vector system would be limited in its ability to find a significant fraction of the pathways which might be present. A much greater limitation, however, lies in the use of functional assays for screening. Few bacterial hosts are as amenable as *E. coli* towards genetic manipulation with today's technology, but *E. coli* is not necessarily an appropriate host for gene expression for a given pathway. As any enzymologist will confess, a variety of hurdles confront functional enzyme expression under the best of circumstances with specifically engineered expression vectors (118, 190). These include poor codon usage, protein instability or toxicity to the host, and protein misfolding, just to name a few (118, 190). The situation is far more bleak when only single-copy, large-insert vectors can be used and the *E. coli* host is limited to a handful of strains which are capable of propagating such large vectors stably. The probability that any pathway contained in the library, much less one pathway in particular, will be functionally expressed and detected is slim, perhaps explaining the necessity of such large libraries in the previous examples. The technology at this stage did not address the stated problem well because of the inefficiency of this search strategy.

Advances in Natural Product Metagenomics. It did not take long before attempts were made to remedy these shortcomings. First, many groups attempted to switch to bacterial artificial chromosome (BAC) vectors for cloning purposes. DeLong found early success with a

marine microbe library constructed from filtered seawater (10), but others struggled with limited results. The Clardy-Handelsman team did nominally produce a BAC library from soil, their favorite metagenomic target, but due to the fragility of DNA beyond the 50kb size range and the stringent purification necessary for this type of sample, the insert sizes were far below those observed with conventional mammalian and plant genomic library construction (145) and did not justify loss of cloning efficiency in comparison with the λ-based system. Better purification regimens have been published (166), but still there has not been a reported example of a soil metagenomic BAC library which takes full advantage of the capacity of the BAC approach.

Similarly, several groups have performed experiments with modified vectors which were capable of propagation in multiple hosts, including *Pseudomonas putida* and *Streptomyces lividans* (41, 120, 207). Other groups have looked into even broader host range vehicles, including an unusual vector useful in a variety of gram-negative bacteria (113, 206, 211). These experiments met with some success, although the number and utility of pathways found were not particularly compelling considering the amount of effort required to produce these novel vehicles.

To improve screening strategies, a variety of approaches have been tried. In addition to traditional screening by complementation (113, 206), several innovative strategies have been produced, including substrate-induced gene expression (SIGEX) (197), and stable-isotope probing (SIP) experiments (57). In the first, translation of the genes contained within the plasmid is coupled with expression of green fluorescent protein (GFP). The initial library is sorted by flow cytometry, removing clones showing constitutive expression of GFP. The library is then induced with the metabolite of interest, in this case various aromatic substrates, to look for induction of the pathways contained on the plasmid. The cells are re-sorted, this time collecting those cells which fluoresce. The result is that a very large library can be screened quite rapidly to look for pathways which are involved in the metabolism of the inducing substrate. In SIP, the initial sample is grown in the presence of ^{13}C or ^{15}N enriched metabolite of interest. The DNA is then fractionated by gradient centrifugation, allowing DNA enriched for the heavier isotopes to be separated from DNA of organisms which have not incorporated the substrate of interest. In this way, the library produced will be enriched with the DNA of those organisms capable of metabolizing the compound of interest.

Other groups have used nucleotide based targeting and screening. Several groups have focused on the unculturable symbionts of known toxin sources, such as the sponges responsible

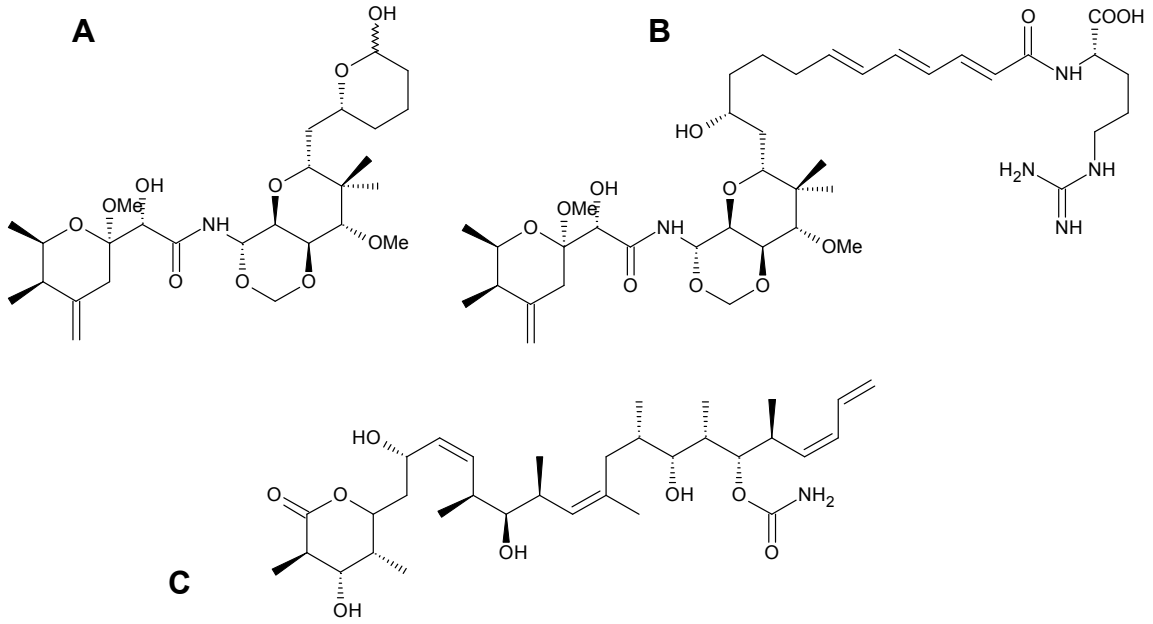


FIG. 2. Polyketides pederin, onnamide A, and discodermolide. (a) Pederin, (b) Onnamide, (c) Discodermolide

for the production of discodermolide (174) and the *Paederus* beetle, the source of the toxin pederin (Fig. 2) (154, 155). Both of these compounds belong to the class of secondary metabolites called polyketides. By narrowing the search to the few bacterial species identified in this specialized environment, the library is enriched for organisms likely to produce pathways of interest, and identification of the bulk of the organisms present can be done without too much difficulty. Once the identities of the organisms are known, it becomes a much simpler procedure to design degenerate primers which target highly conserved regions of related polyketides from other organisms. This method has been successful, but has resulted in some unexpected twists. In the case of Piel and co-workers, who sought after the gene-clusters for the onnamide and pederin pathways, it became clear that the two pathways, one from a Turkish beetle symbiont and one from a Japanese marine sponge, were actually closely related. Although the species harboring the sponge-associated pathway could not be identified, it is likely the two are evolutionarily related based on the fact that the similarities of the structures corresponded to a high degree of similarity in the two pathways despite their vastly different origins. In addition, both used a previously unknown "trans-AT" domain architecture which aided identification. In the case of the search for the discodermolide pathway by the DeLong group, a variety of

polyketide pathways were found, but none actually corresponded to the structure of discodermolide itself.

Similar difficulties have been reported by a number of other researchers. The problem tends not to be one of a paucity of pathways, but confusion at their variety and number. This was true even when pathways were sought from individual organisms known to produce the compound of interest. In fact, a specific approach was proposed to deal with just such a problem on the level of the individual organism (172), but this cannot be readily applied to a mixed sample for which the statistics involved would not apply. As a result, seeking specific pathways through homology in mixed environmental samples has only proven more difficult.

Furthermore, these types of difficulties illustrate the current limitations of metagenomics for the isolation of new chemical species: it is not enough to isolate a DNA sequence if the goal is to produce a useful compound. That one may easily find plenty of "interesting," but ultimately useless DNA only underlines how great is the potential that surrounds us and how little we are able to exploit it. In this sense, the promise of metagenomics has already been realized: there does appear to be great metabolic diversity in the uncultured environment and there are a great many new molecules waiting to be discovered. It even appears that it is not that difficult to get at the sequences which have been hidden for so long. But these new sequences have yet to be translated into new chemical entities on a regular basis and this remains an area of intensive research.

NATURAL PRODUCT PATHWAYS

Despite the dizzying variety of structures classified as "natural products," the biological manufacture of the great majority proceeds through a relatively few distinct molecular assembly processes (Fig. 3) (48). Two related families, the non-ribosomal peptides and the polyketides, are produced by related enzymatic machinery conveniently called non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS), which appear to be evolutionarily derived from, or at least related to, fatty acid synthesis (33, 192). Terpenes are produced by sequential polymerization of the activated 5-carbon units isopentenyl pyrophosphate and dimethylallyl pyrophosphate to produce isoprenoid chains, which then undergo a variety of modifications to produce their respective end products (48). Other classes of natural products, such as the alkaloids, derivatives of shikimic acid, and carbohydrate derivatives, are not as amenable to a generalized cloning through homology scheme due to the non-systematic nature of their production and the difficulty of discriminating these pathways from their close cousins involved in primary metabolism.

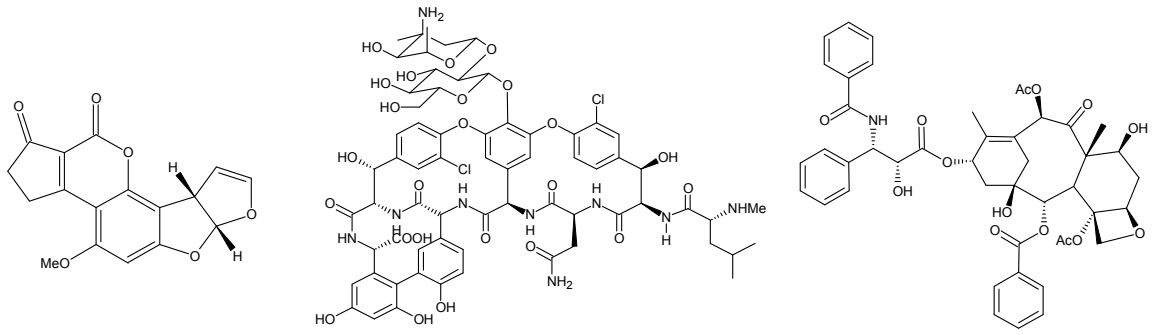


FIG. 3. Typical examples of natual product structures from the polyketide, NRPS and terpene families. (a) Aflatoxin B_1, a polyketide. (b) Vancomycin, a non-ribosomal peptide. (c) Taxol, a terpene.

The Polyketides. Polyketides are a class of natural products characterized by repeating two carbon (C2) units with varying levels of oxidation at alternating carbon positions (Fig. 4) (192). Frequently these compounds take forms ranging from aromatics to macrolactones, polyenes, and saturated chains. Though they can take on vastly different forms, patterns soon emerge in their architecture once their common biosynthetic origin is taken into account.

Polyketides are synthesized in a process very similar to fatty acid synthesis (Fig. 5) (33, 192). Acetate units generated through primary metabolism are loaded onto coenzyme A (CoA) as an activated thioester. These are either directly loaded onto a cysteine residue by transesterification to act as a starter unit, or are converted to malonyl CoA prior to transesterification onto flexible phosphopantetheine arms to act as extender units. Malonyl units are joined to the thioester component of upstream units through a Claisen-like condensation reaction, releasing CO_2 in the process, to generate a new β-keto thioester extended by two carbons. In this manner the chain is extended. In fatty acid synthesis, this β-keto thioester would be reduced to an alcohol, dehydrated to an alkene, then further reduced to the saturated alkane prior to being passed to the next series of domains for the next Claisen chain extension reaction. In polyketide synthesis, some, all, or none of these steps can be eliminated or "skipped," resulting in a structure with varying degrees of oxidation down the length of the growing chain. Eventually the chain is cleaved from the enzyme complex and further modified by other enzymes specific to the molecule in question.

All this chemistry takes place on large, multifunctional complexes (33, 50, 192). The enzymes themselves can be divided into at least two types. Type I synthases use a mammoth, single polypeptide to act as a great, molecular scale assembly line (Fig. 6). This complex molecule is divided into regions and sub-regions based on the chemistry performed. Each

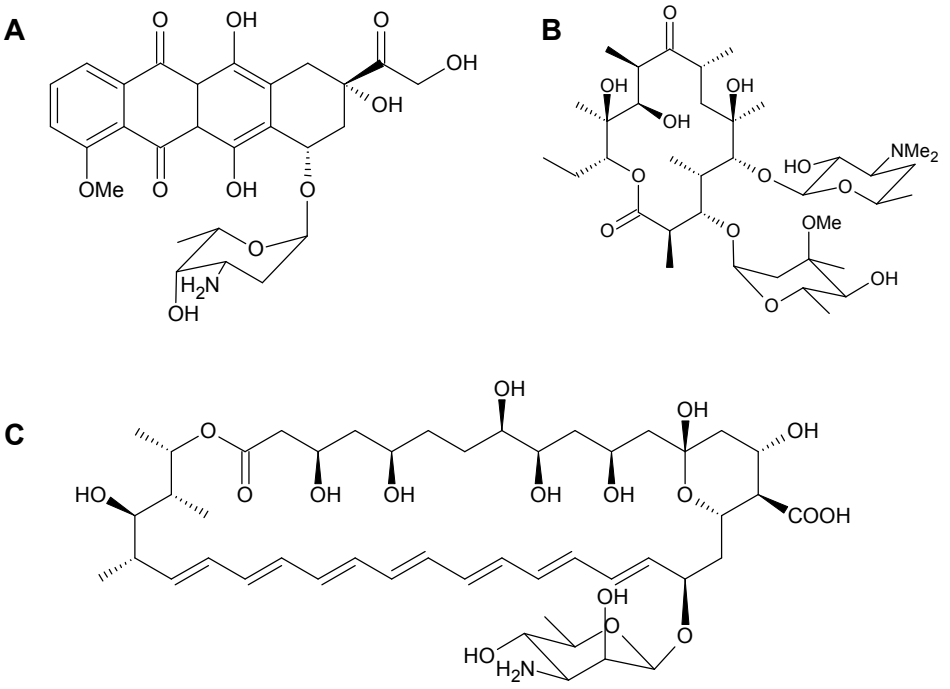


FIG. 4. Selected examples of aromatic, macrolactone, and polyene polyketides. (a) Doxorubicin, an aromatic polyketide. (b) Erythromycin A, a macrolactone. (c) Nystatin A_1, a polyene.

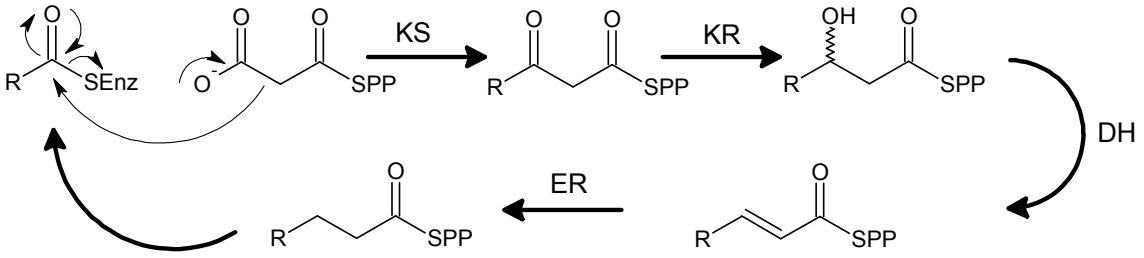


FIG. 5. Fatty acid biosynthesis. Fatty acid biosynthesis procedes through a cyclic mechanism, beginning with a Claisen condensation-like carbon-carbon bond forming reaction, then proceding through a reduction, dehydration, and a second reduction to generate a saturated chain. KS – ketosynthase, KR – ketoreducatse, DH – dehydratase, ER – enoyl reductase, SPP – phosphopantetheine thiol, SEnz – enzyme bound cysteine.

individual reaction is performed by an individual catalytic "domain." A complete set of domains, which carries the substrate through one complete extension reaction and all associated modifications, is called a "module." While within a given module, the substrate chain stays attached to the same phosphopantetheine arm, and when all reactions are complete, the elongated and modified chain is passed to a cysteine present on the first domain of the next module in a

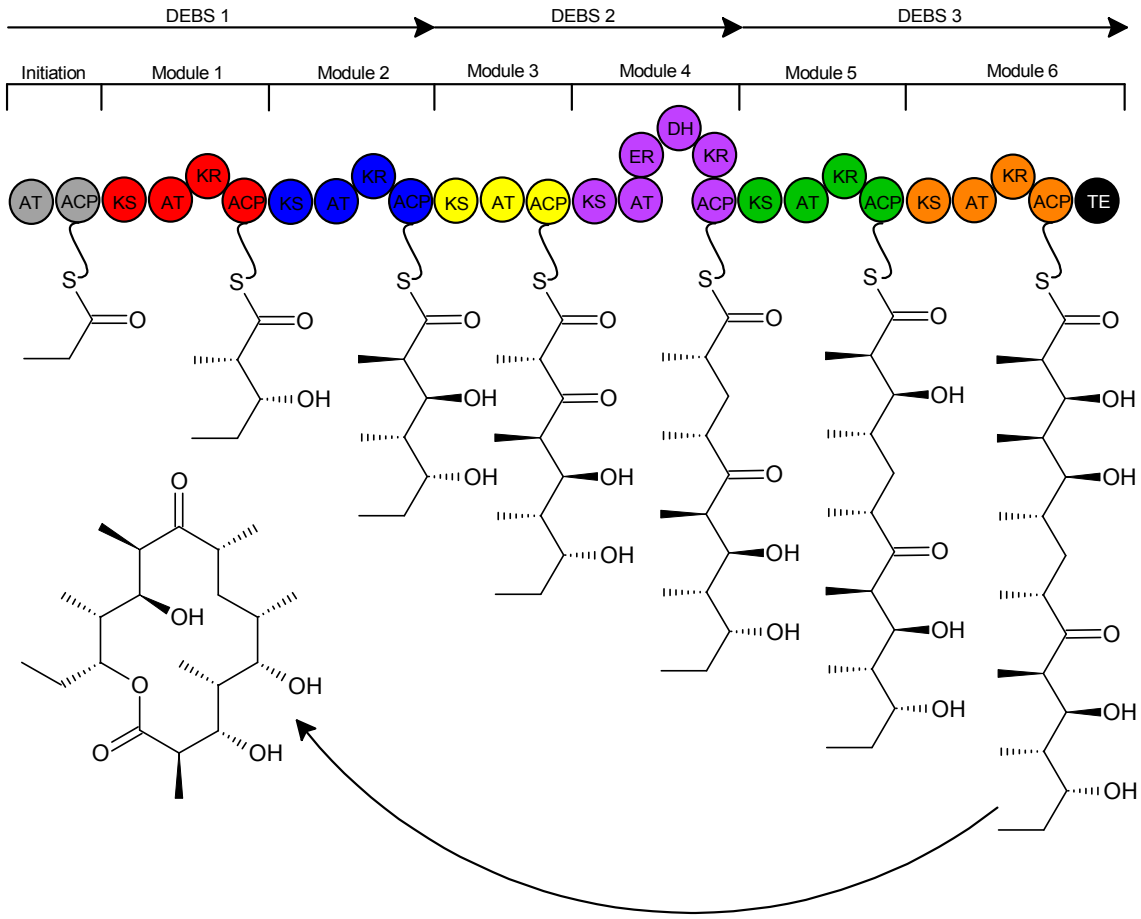


FIG. 6. Modular structure of DEBS (6-Deoxyerythronolide) synthetase. After Staunton, et al. (192).

transesterification reaction to act as the electrophile in the next Claisen condensation. Type II synthases code for each enzymatic activity on a separate polypeptide, and these come together *in situ* to form a large complex which functions similarly to the type I mega-peptide. Between these two extremes lie pathways with varying levels of division. In addition, some type I synthases use a 1-active site, 1-reaction strategy, with each acting in turn to perform its own specific modification and passing the growing chain down the line, while others use an *iterative* strategy, passing the growing thioester around and around through a cycle of active sites like a rotary engine. Most type II synthases function in this manner. As more and more pathways are sequenced, new architectures reveal themselves and the lines between these types have become increasingly blurred, so this terminology might best be viewed and applied a bit loosely (185). A recently discovered set of type III synthases produces the chalcone family of polyketides and

performs condensations of acetyl CoA directly without the use of a covalent enzyme intermediary (6). These synthases show a high degree of divergence from the type I and II systems.

Within the type I and II synthases, the enzymatic domains which perform the chemistry remain the same. The first reaction, the Claisen-like condensation which extends the chain by two carbons, takes place in a ketosynthase (KS) domain. This domain includes the cysteine attachment site for the thioester electrophile. Each phosphopantetheine is loaded with a malonyl nucleophile by an acyl transferase (AT) domain. The other domains, if they are present, are likewise called the ketoreductase (KR), the dehydratase (DH), and the enoyl reductase (ER) domains. The phosphopantetheine arm is attached to a domain called the acyl carrier protein (ACP). The KS, AT, and ACP domains form the minimal core domains required for chain extension and appear in that order. If other domains are present, they are inserted between the AT and ACP domains. When the chain has reached its appropriate size and has completed all the reactions, it is cleaved from the final ACP by a thioesterase (TE) domain. The domains present in the PKS machinery tend to show varying degrees of homology between both species and pathways at the amino acid level. This is particularly true of the KS domain, and several regions show high enough homology that they are good targets for the construction of degenerate probes (41).

The chemistry described thus far clearly has the capability of producing a wide range of compounds, but this is really only a description of the fundamentals. Overlaid atop this is a wide range of variations which have been observed that vastly expand this chemical repertoire. Many organisms possess the ability to incorporate multi-carbon units in addition to C2 units by utilizing substituted malonates or doing the initial loading reaction with a more complex substrate. Because the growing chain contains alternating centers of oxidation, some organisms can further condense the chain through an aldol-like process to produce a variety of cyclic structures and aromatic rings. In addition, there are frequently "post assembly line" modifications, such as additional oxidation, methylation, structural rearrangements, and glycosylation, which can be used to manipulate and "decorate" the core structure. Lastly, there are many polyketides which actually contain components of NRPS pathways, called hybrid pathways, which allow the incorporation of amino acids.

The Non-Ribosomal Peptides. The polyketides and non-ribosomal peptides share many similarities. Like the polyketides, the non-ribosomal peptides are produced through a series of acylation reactions on a large, multifunctional protein complex (53, 180). Amino acid

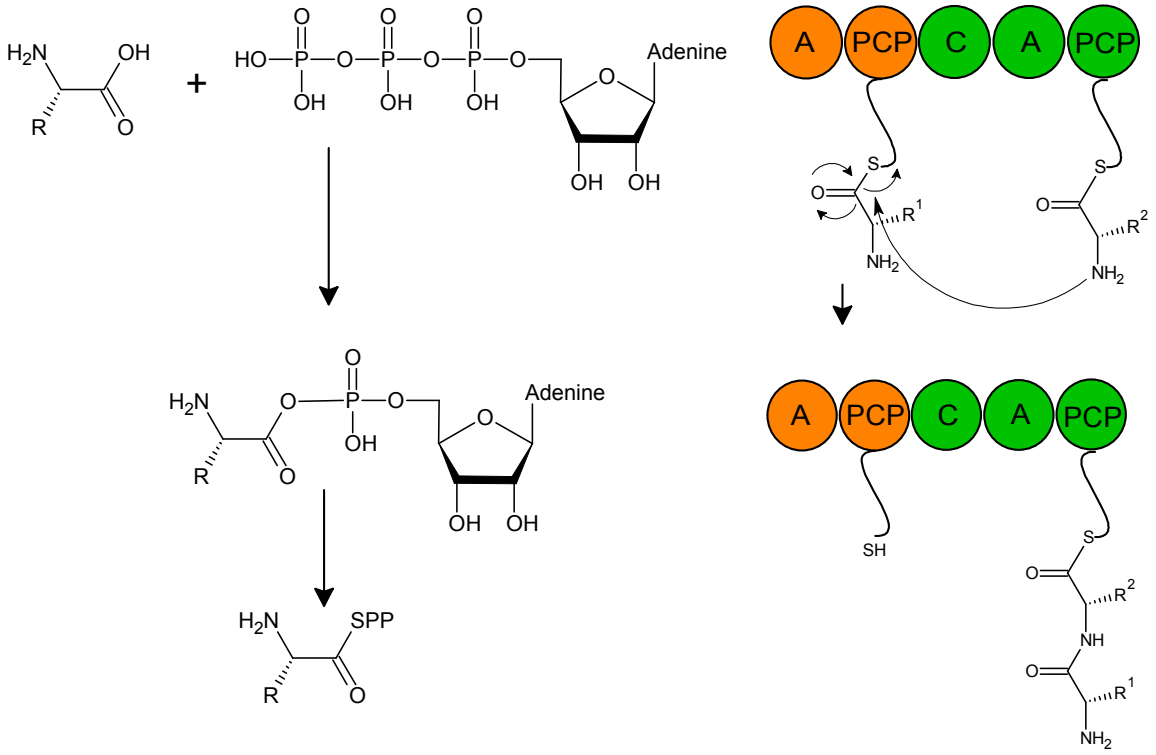


FIG. 7. Chemistry of amino acid activation and chain elongation on a modular NRPS.

substrates are activated to the acid anhydride adenylate, then transferred to the thiol of a phosphopantetheine through a nucleophilic acyl substitution reaction to produce the thioester (Fig. 7). Chain elongation occurs when the amine of another amino acid, similarly stored downstream as a phosphopantetheine thioester, performs a nucleophilic displacement in a second nucleophilic acyl substitution reaction. The new chain then serves as the electrophilic substrate of the next amino acid in line.

As with the polyketides, the non-ribosomal peptides are assembled on large, multi-enzyme complexes, either coded on a single polypeptide or in separate polypeptides which then self assemble (Fig. 8). The required activities are named much like their polyketide counterparts: domains catalyze each step, a complete set of domains for one chain extension and modification forms a module. Amino acid adenylation is performed by adenylation (A) domains. Formation of the phosphopantetheine thioester is performed by the thiolation (T) domain. This domain is also the attachment site for the phosphopantetheine arm and is called the peptidyl carrier protein (PCP) domain by some researchers. The nucleophilic acyl

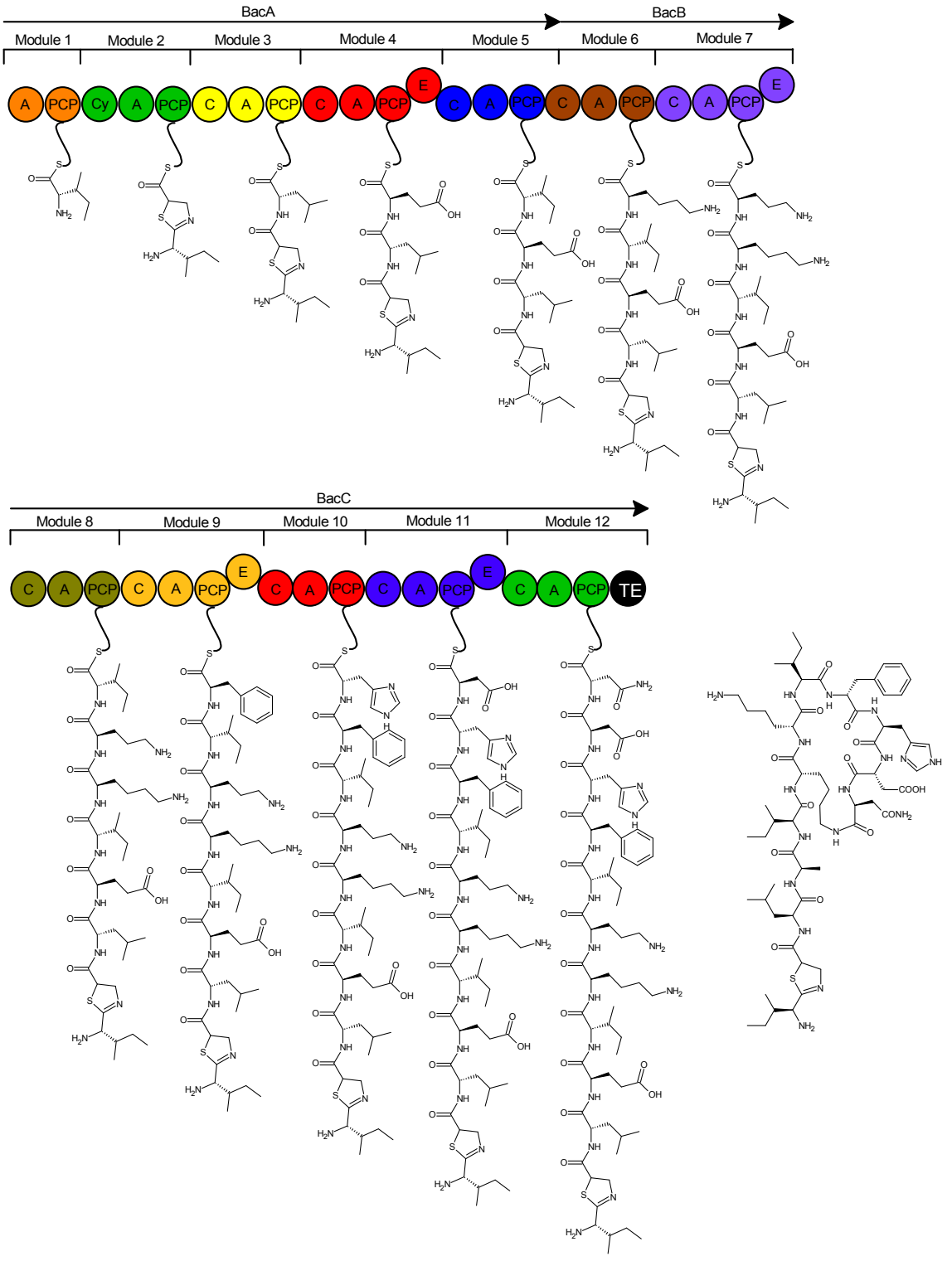


FIG. 8. Modular structure of the bacitracin synthetase complex. After Marahiel, et al. (180).

substitution reaction which results in chain elongation is performed by condensation (C) domains, which also has the effect of transferring the peptide chain to the next module.

Some NRPS systems contain other tailoring domains within this basic structure. Epimerase (E) domains can convert incorporated L-amino acids into their D-counterparts. N-methytransferase (N-Mt) domains can methylate amide residues using S-adenosyl methionine. If the module is responsible for incorporation of serine, threonine, or cysteine, the C domain may be replaced with a cyclization (Cy) domain, which, in addition to forming the amide bond, further condenses these side chains with the amide to produce oxazoline and thiazoline functional groups. If the Cy domain is accompanied by an oxidation (Ox) domain, the heterocycle can be aromatized into the oxazole or thiazole. Like the polyketides, chain termination can result from cleavage by a thioesterase (TE) domain, to produce the free acid if the nucleophile is water or a macrocyle if it is an internal residue, or by a reduction domain (R) to produce an aldehyde. As with the polyketides, hybrid pathways are common and the polyketide domains flow seamlessly with the NRPS machinery. A variety of other, less common domains are also available.

In addition to this variety, NRPS can incorporate a wide array of amino acids which are not used in ribosomal protein synthesis. Some, like ornithine, appear in other metabolic processes, but some appear to be produced exclusively for the purpose of non-ribosomal peptide synthesis. Indeed, in most such cases, the biosynthetic pathways for these compounds can be found adjacent to the NRPS pathway itself. A variety of post-assembly modifications are also common, including glycosylation, acylation, alkylation, and even halogenation.

The Terpenes. Compared to the synthesis of the polyketides and non-ribosomal peptides, terpene biosynthesis looks like a “traditional” biosynthetic pathway. There is no multi-active site, megasynthase assembly-line. Each enzyme acts independently.

Synthesis can begin from one of two sources. The longest known and most well studied pathway begins with acetate and is called the mevalonate pathway (Fig. 9A) (16, 44, 48). Three units are activated to the thioester and condensed in several steps to produce hydroxymethylglutaryl-CoA (HMG-CoA). This compound is dehydrated, reduced, decarboxylated, and phosphorylated in several steps to produce isopentenyl pyrophosphate, which can be isomerized to dimethylallyl pyrophosphate. These two fundamental units are polymerized to produce terpenes and represent the metabolic junction point of the two strategies for terpene synthesis. The second pathway, called the non-mevalonate pathway (Fig. 9B), begins with the condensation of one unit of acetate and one of glyceraldehyde-3-phosphate to

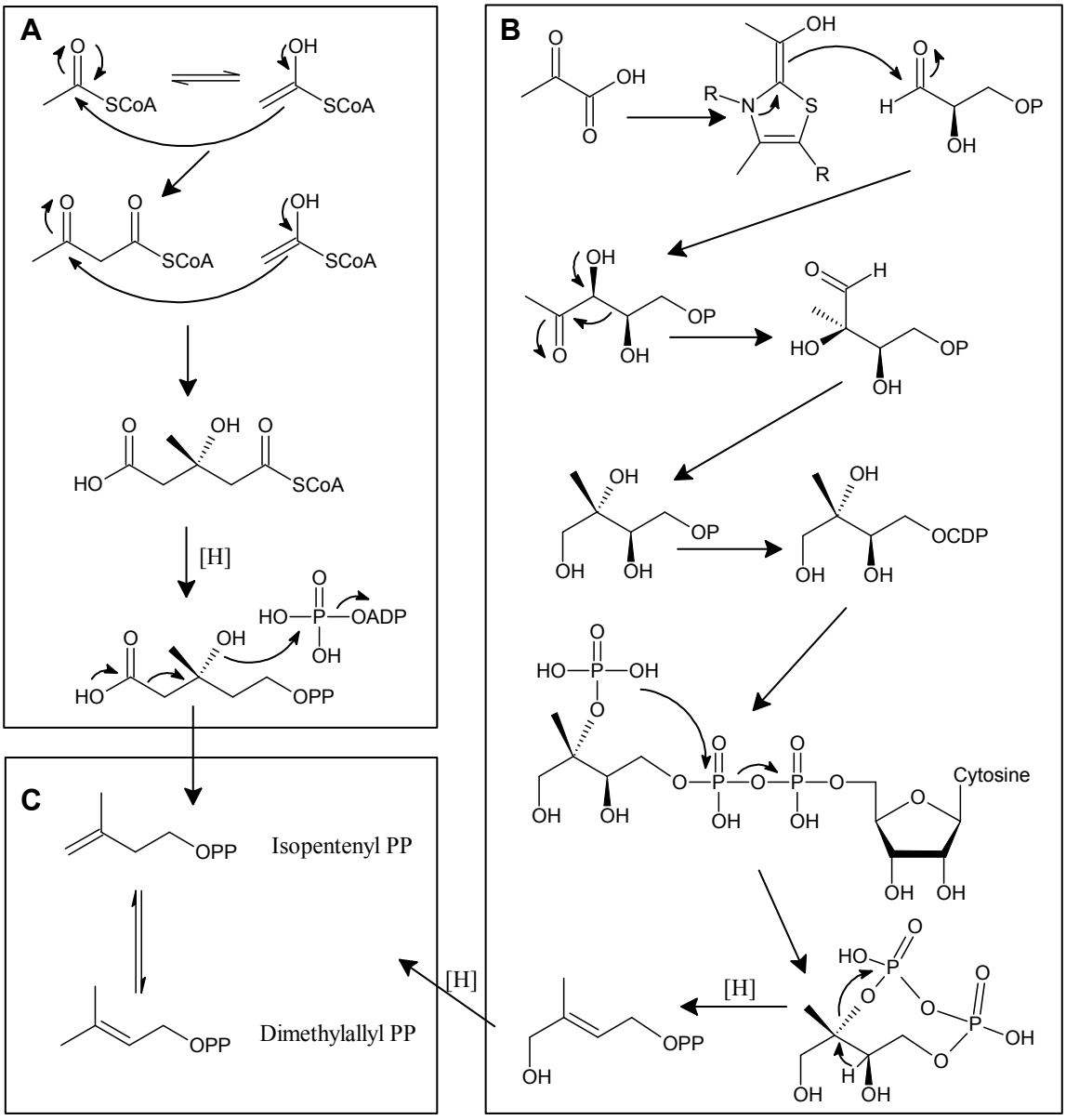


FIG. 9. Mevalonate and non-mevalonate biosynthesis of isopentenyl and dimethylallyl pyrophosphate. After Dewick (48). (a) Mevalonate pathway, (b) Non-mevalonate pathway, (c) Isopentenyl and dimethylallyl pyrophosphate end products.

generate 1-deoxy-D-xylulose 5-phosphate (26, 44, 48). This molecule undergoes several reduction and rearrangement steps to generate isopentenyl pyrophosphate, the same end product as the first pathway.

From here, units are polymerized by a series of enzymes called prenyl transferases (Fig. 10A) (44, 48, 114). Some of these enzymes catalyze specific reactions to generate single products, some catalyze polymerization cascades to generate populations of long chains of

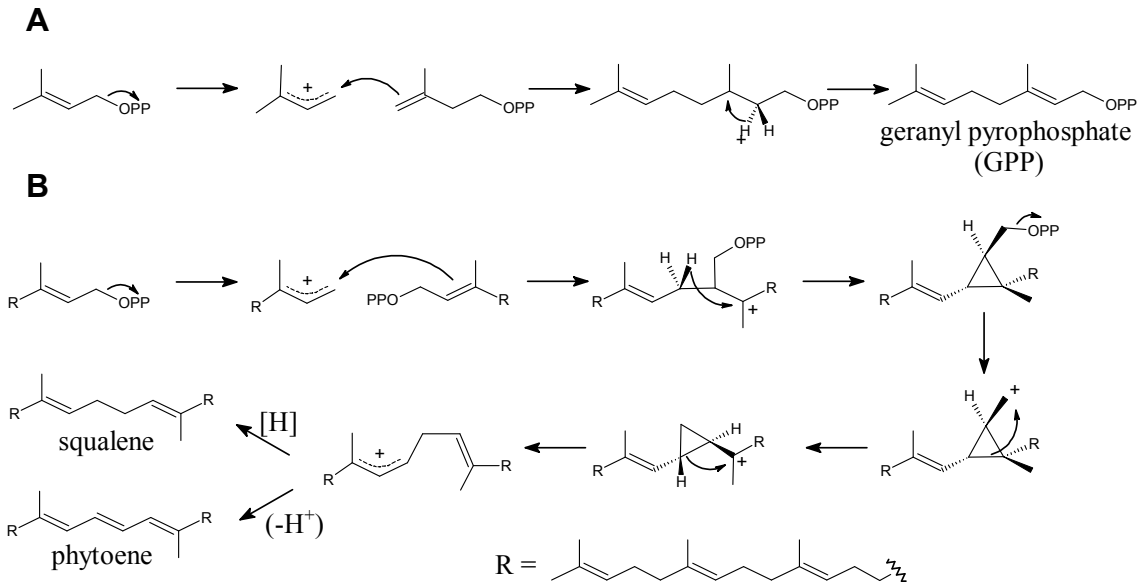


FIG. 10. Head-to-tail and tail-to-tail condensation of isoprene units. After Dewick (48). (a) Head-to-tail condensation of dimethylallyl pyrophosphate and isopentenyl pyrophosphate to produce geranyl pyrophosphate. (b) Tail-to-tail condensation of geranylgeranyl pyrophosphate to generate squalene and phytoene.

varying length, one of the most extreme example being the synthesis of rubber (114). In any case, the reaction begins with enzyme catalyzed ionization of the allyl pyrophosphate to produce a delocalized carbocation. This carbocation is attacked by the terminal alkene of isopentenyl pyrophosphate to produce a tertiary carbocation. The carbocation is quenched by loss of a proton to generate a new allyl pyrophosphate extended by five carbons. The condensation of two units produces geranyl pyrophosphate (GPP), three units gives farnesyl pyrophosphate (FPP), and four gives geranylgeranyl pyrophosphate (GGPP). These enzymes contain several conserved regions which makes them good targets for homology based cloning (114).

Allyl pyrophosphate units can also be joined in a tail-to-tail fashion, as opposed to the head-to-tail reaction used for polymerization (Fig. 10B). This is accomplished by one of two pathways, one involving a net reduction and the other lacking any change of oxidation state. These reactions proceed through a complex series of intermediates, including transiently formed cyclopropyl species (2, 15, 26, 47, 48, 69). The non-reductive pathway generates a conjugated triene at the point of ligation, since the carbon-carbon bond forming reaction occurred between two allylic alcohol positions. These compounds can be desaturated and modified to form the class of compounds called the carotenes, including lycopene and β-carotene (171). The

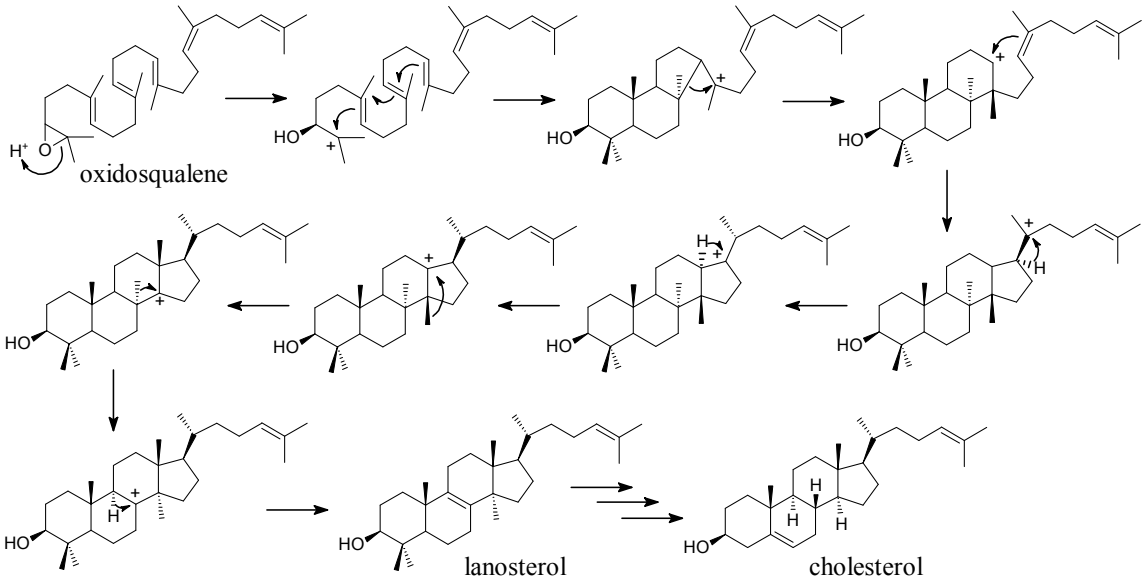


FIG. 11. Polyene cyclization cascade of oxidosqualene to produce lanosterol. After Davis, et al. (44)

reductive pathway results in two saturated methylene units at the juncture. These compounds undergo further metabolism to form a wide variety of compounds, perhaps the most famous of which is the polyene cyclization cascade of squalene oxide which leads to the steroids (Fig. 11) (26).

With respect to secondary metabolism and the generation of natural products, the other major fate of these isoprene oligomers is rearrangement and cyclization by a variety of terpene cyclases (30, 31, 48, 111, 114). A dazzling array of products is generated this way. Mechanistically, these reactions resemble those previously discussed in that the first step is usually ionization of the allyl pyrophosphate to generate a delocalized carbocation. From here, there are many possible paths, including hydride shifts, Wagner-Meerwein rearrangements, interception of the carbocation by various nucleophiles, and quenching through loss of a proton, and it is not uncommon for a cascade of many of these elements to occur. Frequently, one or more of the steps involves nucleophilic attack of a carbocation by an alkene present elsewhere on the chain, generating cyclic molecules.

As with the other natural product classes, it is quite common for terpenes to be further modified once the basic skeleton has been assembled by these fundamental reactions. Most undergo oxidation by various oxidases to generate alcohols and ketones. In addition, they are frequently acylated, glycosylated, and alkylated.

Statement of Purpose. Thus far, biosynthetic research in natural product chemistry has proceeded primarily in the direction of chemicals to genes. Compounds are discovered, their structures provide hints as to the types of pathways that produced them, then the pathways are sought out using the techniques of molecular biology. The advances discussed above have turned this notion on its head, allowing the possibility of performing searches in the direction of genes leading to active compounds. Several natural products families have sufficient genetic homology to make them amenable to cloning through nucleotide based probing of genomic libraries. We will attempt the use of genomic, metagenomic, and mixed culture techniques to explore these possibilities and develop methods of natural product pathway cloning from genomic libraries.

CHAPTER II

PYOCYANIN ISOLATED FROM A MARINE MICROBIAL POPULATION: SYNERGISTIC PRODUCTION BY TWO DISTINCT BACTERIAL SPECIES AND MODE OF ACTION*

INTRODUCTION

Natural products and their structural analogs account for approximately 60% of new small-molecule drugs to enter the market over the past 20 years (136, 138). Notable triumphs include Taxol, produced from the Pacific yew tree and one of the most successful anticancer agents of the past decade (168), as well as vancomycin, often referred to as the drug of last resort, the last line of defense for treatment of drug-resistant *Staphylococcus aureus* infections (214). Both drugs have generated $1 billion in annual sales (100, 214). Traditional methods for mining natural products entail organism collection, extraction, bioassay-guided purification, and structure determination. Among possible sources, the marine environment has provided an abundant supply of natural products with a diverse array of structures and bioactivities, including antimicrobial, immunosuppressive, and anticancer agents (16, 137). Indeed, ziconotide, the synthetic form of the cone snail peptide ω-conotoxin M-VII-A, a neuron-specific N-type calcium channel blocker, has recently entered the market for treatment of severe chronic pain (125).

Sponges and ‘‘culturable’’ microorganisms are common resources for the discovery of marine-derived therapeutics. Microorganisms can offer the added advantage of providing a continuous supply of the natural product, although it is frequently debated as to whether the sponge itself or microorganisms living within the sponge environment are responsible for natural product production (89, 153, 154). The general strategy for identifying microorganisms from marine habitats typically involves screening microbial isolates for bioactivity. Here, we investigate marine microbial communities for antibiotic/antifungal activity, bearing the hypothesis in mind that natural product production, i.e., the regulation and expression of their respective biosynthetic pathways, could be coupled with microbial interactions (5). Indeed,

---

* Parts of this chapter have been reproduced from **Angell, S., B. J. Bench, H. Williams, and C. M. Watanabe.** 2006. Pyocyanin isolated from a marine microbial population: synergistic production between two distinct bacterial species and mode of action. Chem Biol **13:**1349-59.
 DOI 10.1016/j.chembiol.2006.10.012.

TABLE 2. Bacterial strains and plasmids used in Chapter II.

Strain or plasmid	Description[a]	Reference/Source
Strains		
E. coli DH10B	Host strain for DNA manipulations	Invitrogen
Pup14	Marine microbial assemblage isolated in Pupukea, HI	This study
Pup14A	Identified as *Pseudomonas aeruginosa* by CBS	This study
Pup14B	Identified as *Enterobacter sp.* by CBS	This study
Surf1	Identified as *Pseudoaltermonas sp.* by CBS	This study
Hon6	Identified as *Chromohalobacter sp.* by CBS	This study
Pup16	Identified as *Rhodotorula mucilaginosa* by CBS	This study
KM1	Identified as *Enterobacter sp.* by us	This study
Plasmids		
pCR2.1	Amp^r, Kan^r, TA cloning plasmid from Invitrogen	Invitrogen
pTApyolas	Amp^r, Kan^r, .7 kb fragment containing upstream region of phz1 pathway amplified by phz2_up3 primer pair and TA cloned into pCR2.1	This study
pTAphz2las	Amp^r, Kan^r, .48 kb fragment containing upstream region of phz2 pathway amplified by phz2_up3 primer pair and TA cloned into pCR2.1	This study
pTA2-3	Amp^r, Kan^r, 1.1 kb fragment containing upstream region of phz2 pathway amplified by phz2_up2 primer pair and TA cloned into pCR2.1	This study
pTA3-4	Amp^r, Kan^r, 1 kb fragment containing upstream region of phz2 pathway amplified by phz2_up3 primer pair and TA cloned into pCR2.1	This study

[a] Amp^r, ampicillin resistance; Kan^r, kanamycin resistance.

bacteria isolated from marine snow particle-microbial communities have been shown to display potent antagonistic effects, offering some support for this mechanism (71, 115).

MATERIALS AND METHODS

Instrumentation and General Methods. Unless otherwise specified, biochemical reagents were obtained from Sigma Biochemicals (St. Louis, MO). Instant Ocean utilized in marine media preparations was obtained from Petco (College Station, TX). All other media reagents were obtained from VWR International (West Chester, PA). Dichloromethane (CH_2Cl_2), chloroform ($CHCl_3$), and methanol (MeOH) were purchased from EMD, while ammonium hydroxide (NH_4OH) was purchased from EM Science (VWR International, West Chester, PA). All were used without further purification. Flash column chromatography was performed with grade 62 silica gel (EMD, 60–200 mesh). ^{1}H-NMR spectra were obtained with a Varian Inova-300 spectrometer (Palo Alto, CA), in methanol-d_4 (Aldrich; Milwaukee, WI). ^{1}H-NMR (300 MHz) spectra are reported as follows: chemical shifts in ppm downfield from TMS, the internal standard (δ-scale); multiplicity (s, singlet; d, doublet; t, triplet; m, multiplet), integration, and coupling constant (Hz). ^{13}C and other 2D NMR spectra were obtained on a Bruker ARX500 spectrometer. ^{13}C-NMR (125 MHz) spectra were recorded in proton decoupled mode and reported in ppm. Preparative high performance liquid chromatography (HPLC)

separations were conducted on a Varian ProStar 230 solvent delivery module equipped with a Varian ProStar 330 Photodiode Array Detector (Palo Alto, CA) by using a Phenomenex Luna 5μ Silica (250 x 10.0 mm) column (Torrance, CA). Infrared (IR) spectra were obtained on a Perkin-Elmer FT-IR spectrometer (Boston, MA) with absorption peaks reported in wave numbers (cm^{-1}). Mass spectra were recorded on a PE Sciex with an API Qstar Pulsar (Concord, Ontario, Canada). The Affymetrix platform (Santa Clara, CA) was used in all GeneChip experiments. Cell-density measurements were made with a Biotek μQuant plate reader.

Organisms. All marine microorganisms utilized in these experiments were obtained from field collections (Table 2). Hon6 (isolate 6) was acquired from Honokohau, Maui, Pup16 (isolate 16) from Pupukea, Oahu, and KM1 (isolate 1) from the Florida Keys. Pup14A and Pup14B were also obtained from Pupukea (sample 14). The identity of each of the microorganisms was assigned by the Centralbureau voor Schimmelcultures (CBS) in the Netherlands. The yeast *S. cerevisiae* wild-type strain (#404; BY4741; MATα his3[δ]1 leu2[δ]0 met15[δ]0 ura3[δ]0) was obtained from Dr. Michael Kladde at the Department of Biochemistry, Texas A&M University. *E. coli* DH10B was obtained from Invitrogen (Carlsbad, CA).

Field Collection of Marine Microbial Samples. Marine agar plates were prepared with slight modification to standard literature protocols (42). Yeast extract (2 g), bactopeptone (2 g), glucose (4 g), and bacto-agar (Difco, 20 g) were combined in 1 liter of seawater, prepared by mixing Instant Ocean (38 g, Petco) with deionized distilled water. The marine agar plates were further supplemented with either 10 mg/l of miconazole or a combination of chloramphenicol (10 mg/l) and kanamycin (10 mg/l) for culturing of bacteria or fungi, respectively. Culture swabs (Becton Dickinson, supplier VWR International; West Chester, PA) were used to swab submerged rocks, and autoclaved vials were used to collect ocean floor sediments. Samples were applied to both types of marine agar plates and incubated at room temperature for one week. Single colonies were generated by repeated streaking of microorganisms onto fresh plates. To generate marine microbial populations, cell scrapers (Costar, supplier VWR International; West Chester, PA) were applied to the original plates, and the cells resuspended and cultured in marine broth. These experiments were performed by Dr. Coran Watanabe.

IC_{50} Determination. Overnight cultures (5 ml) of *S. cerevisiae* or *E. coli* were generated. *S. cerevisiae* was cultured in YPD broth (30°C, 250 rpm) (208) and *E. coli* in LB medium (169). Cultures were diluted to an OD_{600} of 0.1 and aliquoted (100 μl per well) into a 96-well plate. To each culture was added 1 ml of serial dilutions of the crude extract (solubilized

in ethanol) or ethanol alone. All assays were executed in triplicate. The 96-well plate was incubated at 30°C, 250 rpm for 12 hr. Empty wells were filled with 100 μl of autoclaved water to prevent evaporation during incubation. Cell density was measured on a Biotek μQuant plate reader (Winooski, Vermont). Cellular toxicities were determinied by Dr. Coran Watanabe.

Split/Pool Analysis. Sixty single colonies were streaked and individually numbered on marine agar plates. The cells were each individually resuspended in marine broth and brought to an OD_{600} of 1.0. Initially, the cells were each cultured ten at a time. To 5 ml aliquots of marine broth was added 10 ml of each cell sample in pools of ten, with ten cell samples in 5 ml marine broth. The samples were cultured for 3-4 days. Several pools gave the green phenotype. We focused on one of those pools and repeated the assay again but pooling two samples at a time. As a control, we also cultured each sample individually, but no induction was observed in this case. These experiments were performed by Dr. Coran Watanabe.

Varying the Ratio of Pup14A to Pup14B. Pup14A and Pup14B were streaked onto marine agar plates and grown to confluency. The cells were resuspended in marine broth and brought to an OD_{600} of 1.0. The concentration of Pup14A (200 μl) was held constant and mixed with different amounts of Pup14B (at 1, 10, 50, 100, and 200 μl) and vice versa in marine broth (5 ml). The concentration of Pup14B (200 μl) was held constant and the amount of Pup14A varied (at 1, 10, 50, 100, and 200 μl). This experiment was performed by Dr. Coran Watanabe.

Inhibition Assays. Pup14A and Pup14B were cultured overnight at 30°C, 250 rpm. Cultures were diluted to an OD_{600} of 0.1 and aliquoted 100 μl per well into a 96-well plate. To each culture was added 1 μl of pyocyanin solubilized in ethanol; stock solutions of 1, 10, and 100 μg/ml to give final concentrations of 10 μg/ml, 100 μg/ml, and 1 μg/ml, or ethanol alone. All assays were executed in triplicate. The 96-well plate was incubated at 30°C, 250 rpm for 12 hr. Empty wells were filled with 100 μl of autoclaved water to prevent evaporation during incubation. Cell density was measured on a Biotek μQuant plate reader (Winooski, Vermont). Growth inhibition assays were performed by Dr. Coran Watanabe.

Small-Molecule Induction Assay. Pup14A and Pup14B were cultured on marine agar plates for two days as described above. The agar plates were then sliced into pieces. Agar from one half of each plate was added to separate 2 liter Erlenmeyer flasks containing 1 liter of marine broth. Two sets of each culture, corresponding to a total of four 2 liter flasks, were allowed to grow at 30°C at 250 rpm for 24 and 72 hr, respectively, and subsequently centrifuged. The media from each culture were filtered through a 0.2 μm filter (Nalgene) and transferred to

autoclaved flasks (2 liter). Pup14A cells (24 and 72 hr) were added to the medium of Pup14B (24 and 72 hr) and vice versa. Cultures were again placed in a rotary shaker at 30°C for 3 days. As a control, the other half of each plate (Pup14A and Pup14B) was cocultured in 1 liter of marine medium for 2–3 days until the coloration of the medium turned from yellow to green. This experiment was performed by Dr. Coran Watanabe.

Boyden Chamber Assay. The Boyden chamber assay was performed with Corning transwell plates (0.4 μm pore diameter, 24 mm diameter, 6-well, Corning, NY). Marine broth (1 ml) was added to the top and bottom chambers of the wells of the plate and subsequently inoculated with the appropriate microbial strains. The plates were incubated at 30°C for 2 days without shaking. Following incubation, the culture media from the bottom chamber of the wells was extracted with methylene chloride and centrifuged to give two distinct layers, and the organics were transferred to fresh tubes and concentrated to ~100 μl volume for comparison. This experiment was performed by Dr. Coran Watanabe.

Induction of Blue Phenotype with Different Strains. Microorganisms Pup14A and Pup14B, Hon6, Pup16, Surf1, and KM1 were cultured on marine agar plates to confluency and resuspended in marine medium to give an OD_{600} of 1.0. Pup14A orPup14B (200 μl) was added to 5 ml of marine broth and mixed with 200 μl of Hon6, Pup16, or KM1. As a control, Pup14A and Pup14B (200 μl each) were mixed and cultured. All cultures were grown at 30°C, 250 rpm for a period of a week. Samples were evaporated to dryness, redissolved in dichloromethane, and examined by HPLC. An isocratic solvent system of 87:13 CH_2Cl_2/5% NH_4OH in MeOH was employed. Under these conditions, pyocyanin elutes with a retention time (R_t) of 7.23 min. Pyocyanin production gives a characteristic blue phenotype and is easily detected upon collection. Purified pyocyanin was collected, concentrated in vacuo and its identity confirmed by mass spectrometry. Induction and extraction were performed by Dr. Coran Watanabe, and analysis was performed by Bennie J. Bench.

Purification Conditions. Liquid cultures (2 l) were extracted twice with 1 l of dichloromethane. Fractions were combined, and the solvent evaporated under vacuum. The organic residue was partially purified by flash-column chromatography with silica gel as the stationary phase and 10:1:0.001 CH_2Cl_2/MeOH/NH_4OH as the mobile phase. Fractions were analyzed by TLC utilizing the same conditions. Fractions containing the blue compound were combined and the solvent evaporated to dryness. The remaining semipure residue was dissolved in 1 ml dichloromethane and precipitated with 2 ml hexanes. After standing on ice for 5 min, the

sample was centrifuged at 14,000 rpm at 4°C for 1 min to pellet the blue material and the solvent aspirated off. The precipitation procedure was repeated five times.

HPLC Conditions. Further purification of pyocyanin was performed via high performance liquid chromatography (HPLC). The blue organic residue was dissolved in dichloromethane and purified by HPLC utilizing an isocratic solvent system (87:13 CH2Cl2/5% NH_4OH in MeOH). The purified pyocyanin was collected with a retention time (R_t) of 8.63 min, concentrated in vacuo, and stored at -80°C. Approximately 2 mg of pyocyanin can be generated per liter of cells. HPLC purification and analysis was performed by Bennie J. Bench.

Crystallization Conditions. Pyocyanin (1 mg) was dissolved in 0.5 ml 1:1 CH_2Cl_2:$CHCl_3$ and placed into a test tube equipped with a female ground glass joint. The joint was connected to a Dean-Stark trap which contained 6 ml hexanes in the collection tube. The test tube was submerged in an ice bath. The remainder of the apparatus was maintained at room temperature. Green-blue crystals formed within 48 hr.

Characterization Data. ^{1}H-NMR (300 MHz, methanol-d_4) 8.38 (d, J = 8.4 Hz, 1H), 8.12 (m, 2H), 7.99 (t, J = 8.4Hz, 1H), 7.79 (t, J = 7.0 Hz, 1H), 6.67 (dd, J = 8.4, 4.1 Hz, 2H), 3.95 (s, 3H); ^{13}C NMR (75 MHz, Methanol-d_4) 177.5, 147.3, 146.7, 138.5, 138.1, 136.4, 134.4, 134.0, 128.0, 116.8, 115.5, 95.3, 36.7; IR (neat solid) 2982, 1599, 1454 cm^{-1}; HRMS (ESI) for $C_{13}H_{11}N_2O$ $(M+H)^+$: calcd. 211.0871, found 211.0885. NMR spectral analysis was performed by Dr. Howard Williams, Bennie J. Bench and Scott Angell. X-ray crystallographic analysis was performed by Dr. Joseph Riebenspies.

For other spectral data, see Appendix A. X-ray data are available through the Cambridge Crystallographic Data Centre (deposition number: CCDC 626277).

Detection of Phenazine Biosynthetic Loci- and Quorum-Sensing Regions. *P. aeruginosa* strain Pup14B was cultured as detailed above and genomic DNA isolated from 125 mg of the cell pellet. The pellet was washed twice with 1.5 ml of TE buffer (10 mM Tris, 1 mM EDTA [pH 7.5]) and resuspended in 2 ml of TE buffer. To this was added 2 ml of lysis buffer (0.5 M EDTA, 1% w/v SDS [pH 9.0]) and 1 mg of proteinase K (Sigma-Aldrich). The sample was incubated at 50°C for 16 hr. The solution volume was doubled with water, extracted with chloroform, and the DNA isolated by spooling (169). The DNA was resuspended in 500 µl of TE and treated with 10 ml of 1 mg/ml RNase A for 2 hr at 37°C. This solution was extracted with an equal volume of chloroform, the DNA precipitated with isopropanol, and the pellet washed with 70% ethanol. The DNA pellet was dried and respended in 250 ul of TE buffer.

TABLE 3. Primers used in Chapter II.

Primer Name	Sequence	Melting temperature (°C)	Purpose
Sequencing primers			
PCR2.1seq_for	CAGGAAACAGCTATGACCATG	54	Sequencing primer for pCR2.1
PCR2.1seq_rev	AGATGCATGCTCGAGCGGCCGCC	69	Sequencing primer for pCR2.1
3-4 Left	CTTTTTAATTGCCTCAGAAT	46	Chromosome walking primer for sequencing pTA3-4
2-3 Left	AGATATGCGGTAATTATGGA	48	Chromosome walking primer for sequencing pTA2-3
3-4 Right	AACCGTAGAACGGCTCTC	54	Chromosome walking primer for sequencing pTA3-4
2-3 Right	GTGATCCAGAGCAGGAAG	52	Chromosome walking primer for sequencing pTA2-3
Pseudomonas genome fragment amplification primers			
pyo las F	GTAACCCGAGAAGTACCCAAGCG	59	Amplify .7 kb upstream region of phz1 pathway, including las box
pyo las R	TTTCCCTGTACCGCTGACCGTTC	61	Amplify .7 kb upstream region of phz1 pathway, including las box
phz2 las F	ACGCCATCGGCCTGCTCAACTG	65	Amplify .48 kb upstream region of phz2 pathway, overlapping with other phz2 amplified fragments
phz2 las rev	GGTAAACCCTTTCAACCGTTGG	57	Amplify .48 kb upstream region of phz2 pathway, overlapping with other phz2 amplified fragments
phz2_up2 for	CTCTTCAGCCTCGTTTCGTC	56	Amplify 1.1 kb upstream region of phz2 pathway, overlapping with other phz2 amplified fragments
phz2_up2 rev	CCCGGAAAGTTGCACTAGC	56	Amplify 1.1 kb upstream region of phz2 pathway, overlapping with other phz2 amplified fragments
phz2_up3 for	GGAACCAACTGTTCCAGCAT	56	Amplify 1 kb upstream region of phz2 pathway, overlapping with other phz2 amplified fragments
phz2_up3 rev	CCTGGGTAATTGGACAGGAA	54	Amplify 1 kb upstream region of phz2 pathway, overlapping with other phz2 amplified fragments
RT-PCR analysis primers			
YBR244Wfor	ATGACCACATCTTTTTATGATTTAG	50	RT-PCR forward primer for analysis of glutathione peroxidase transcription levels
YBR244Wrev	TCATTTACTTAACAGGCTTTGGATT	53	RT-PCR reverse primer for analysis of glutathione peroxidase transcription levels
YCL064Cfor	CATCTTTGTCCTCTATTACAACGGA	55	RT-PCR forward primer for analysis of catabolic serine dehydratase transcription levels
YCL064Crev	GGTAGATAAAATCAGGAACACCGGT	57	RT-PCR reverse primer for analysis of catabolic serine dehydratase transcription levels
YEL039Cfor	CTATTTGGCAGCCTTTGTCATATAA	54	RT-PCR forward primer for analysis of Iso-2-cytochrome c transcription levels
YEL039Crev	ATGGCTAAAGAAAGTACGGGATTCA	56	RT-PCR reverse primer for analysis of Iso-2-cytochrome c transcription levels
YGL184Cfor	ATTGTGTTCTTCTAGAGTCT	48	RT-PCR forward primer for analysis of cystathionine β-lyase transcription levels
YGL184Crev	TATTGAACGATTTATGCAGC	49	RT-PCR reverse primer for analysis of cystathionine β-lyase transcription levels
YJR009Cfor	ACATTGACATCGCCATTGACTCCAC	60	RT-PCR forward primer for control analysis of GAPDH transcription levels
YJR009Crev	TTTCATCGTAGGTGGTTTCCTTGTT	57	RT-PCR reverse primer for control analysis of GAPDH transcription levels

The 700 bp sequence upstream of the phzA1 gene was amplified by PCR from the Pup14B genomic DNA with the pyo las F (5'-GTA ACC CGA GAA GTA CCC AAG CG-3') (Table 3) and pyo las R (5'-TTT CCC TGT ACC GCT GAC CGT TC-3') primer set. The reaction contained (in 50 μl) 1X ThermoPol Reaction Buffer, 2.5% DMSO, 250 mM of dNTP's, 2 U Taq DNA polymerase (New England Biolabs, Ipswich, MA), 50 pmol of each primer, and 20 ng of genomic DNA template. Thermal cycling was performed with a PTC-200 thermal cycler (MJ Research, Waltham, MA) by using an initial 2 min denaturing step at 94°C, followed by 30 cycles of 94°C for 30 s, 30 s at the annealing temperature of 55°C, and 1 min at 68°C. After cycling, the reaction was held at 68°C for 7 min. Similarly, the 2041 bp region upstream of the phzA2 gene was amplified in overlapping sections from genomic DNA by using three primer sets, phz2 las F (5'-ACG CCA TCG GCC TGC TCA ACT G-3') and phz2 las R (5'-GGT AAA CCC TTT CAA CCG TTG G-3') (480 bp), phz2_up2 F (5'-CTC TTC AGC CTC GTT TCG TC-3') and phz2_up2 R (5'-CCC GGA AAG TTG CAC TAG C-3') (1102 bp), and phz2_up3 F (5'-GGA ACC AAC TGT TCC AGC AT-3') and phz2_up3 R (5'-CCT GGG TAA TTG GAC AGG AA-3') (937 bp). The same reaction conditions and temperature cycles were used, except that the annealing temperature was 59°C for the phz2_las primer pair and 50°C for the phz2_up2 and phz2_up3 primer pairs. The PCR products were gel purified and TOPO-TA cloned (Invitrogen, Carlsbad, CA).

Cloned fragments were end sequenced with primers PCR2.1seq for (5'-CAG GAA ACA GCT ATG ACC ATG-3') and PCR2.1seq rev (5'-AGA TGC ATG CTC GAG CGG CCG CC-3'). Gaps were filled by chromosome walking with the primers 3-4 left (5'-CTT TTT AAT TGC CTC AGA AT-3'), 3-4 right (5'-AAC CGT AGA ACG GCT CTC-3'), 2-3 left (5'-AGA TAT GCG GTA ATT ATG GA-3'), and 2-3 right (5'-GTG ATC CAG AGC AGG AAG-3'). Sequencing reactions were performed with the Perkin Elmer ABI BigDye Terminator kit according to the manufacturer's instructions and analyzed on an ABI 31303/Genetic Analyzer by the Gene Technology Lab Core Facility of Texas A&M University. Primer design for these procedures was aided by Primer 3 software (source code available at http://fokker.wi.mit.edu/primer3/) (126). Sequence data can be found in Appendix A.

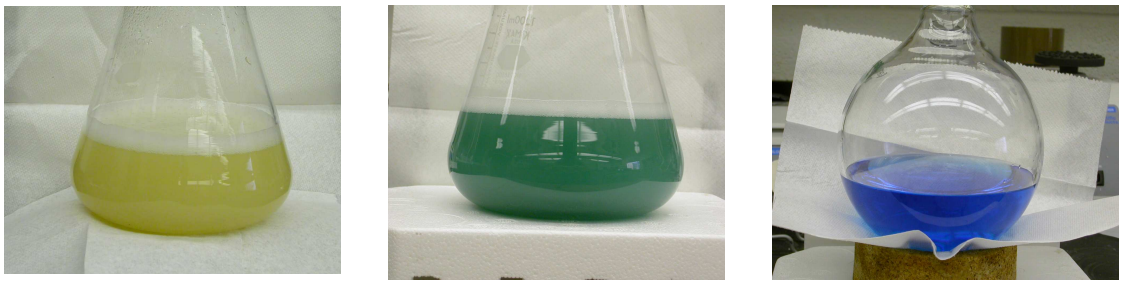
Yeast Sample Preparation and GeneChip Evaluation. Yeast was cultured overnight in YPD broth (208) to an OD_{600} of 1.0. Total RNA was extracted with an SDS/hot phenol extraction method (208) and treated with DNase as follows. Total RNA (30 μl) was incubated at room temperature with 5 μl of first strand buffer (GIBCOBRL cDNA Superscript Choice Kit; Carlsbad, CA) and 1 μl of DNase I (RNase-free, Ambion). Samples were heat inactivated by

heating the samples at 75°C for 15 min and subsequently purified with an RNeasy kit (QIAGEN, Valencia, CA). The samples (15 μg of total RNA per sample) were provided to the GeneChip facility where they were amplified, biotinylated, and hybridized to GeneChips according to the protocol detailed by Affymetrix (Santa Clara, CA). Microarray data are available through the NCBI Geo Database (accession number: GSE6185). Microarray analysis was performed by Dr. Coran Watanabe.

RT-PCR, Validation of GeneChip Results. RT-PCR was used to quantify the mRNA levels of the following, with primer pairs and annealing temperatures indicated in parenthesis: glutathione peroxidase (YBR244Wfor, 5'-ATG ACC ACA TCT TTT TAT GAT TTA G-3', YBR244Wrev, 5'-TCA TTT ACT TAA CAG GCT TTG GAT T-3', anneal at 44°C) catabolic serine (threonine) dehydratase (YCL064Cfor, 5'-CAT CTT TGT CCT CTA TTA CAA CGG A-3'; YCL064Crev, 5'-GGT AGA TAA AAT CAG GAA CAC CGG T-3', anneal at 45°C), iso-2-cytochrome c (YEL039Cfor, 5'-CTA TTT GGC AGC CTT TGT CAT ATA A-3', YEL039Crev, 5'-ATG GCT AAA GAA AGT ACG GGA TTC A-3', anneal at 45°C), and cystathionine β-lyase (YGL184Cfor, 5'-ATT GTG TTC TTC TAG AGT CT-3', YGL184Crev, 5'-TAT TGA ACG ATT TAT GCA GC-3', anneal at 45°C). Total RNA was extracted from control cells and cells treated with pyocyanin (25 μg/ml and 250 μg/ml). Following extraction, total RNA (10 μg) was transcribed into single-stranded cDNA with the Superscript Choice system (GIBCO-BRL). PCR reactions were conducted with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as an endogenous amplification standard (YJR009Cfor, 5'-ACA TTG ACA TCG CCA TTG ACT CCA C-3', YJR009Crev, 5'-TTT CAT CGT AGG TGG TTT CCT TGT T-3', anneal at 55°C). PCR conditions were optimized so that amplification of both GAPDH and the cDNA of interest were in the exponential phase (estimated by varying the number of PCR cycles for each template or gene of interest). PCR cycles consisted of: step 1, 5 min denaturation at 95°C; step 2, 1 min of denaturation at 95°C; step 3, 1 min of primer annealing at the appropriate temperature; step 4, 1 min of extension at 72°C. Steps 2 through 4 were repeated for the requisite number of cycles. PCR products were stained with Sybr Green and fluorescence was measured with a microplate reader. These experiments were performed by Dr. Coran Watanabe.

RESULTS

Isolation of Pup14. Samples were collected from the north shore of Oahu, Hawaii. Ocean floor sediments (up to a depth of 40 ft) were obtained as well as culture swabs of submerged rocks. The microorganism specimens were plated on bacterial or fungal plates



A B C

FIG. 12. Induction of pyocyanin production. (a) Uninduced co-culture of Pup14A and Pup14B after 24 hours of growth. (b) Induced co-culture after 72 hours of growth. (c) Methylene chloride extract of induced co-culture.

employing standard literature protocols (42). Microbial communities were inoculated into liquid media and cultured in shake flasks. Among the microorganism samples examined, Pup14 (Pupukea sample 14, bacterial plate) gave an unusual green phenotype following ~72 hr of culturing; while the cells remained white, the medium turned green (Fig. 12). Extraction of the medium resulted in an aqua blue color. Crude extracts of the Pup14 sample were found to be moderately active against *E. coli* and yeast (IC_{50} = 25 μg/ml). The Pup14 sample showed that cells propagated from single colonies and grown in culture could not generate the green phenotype, suggesting that natural product production was in some way coupled with microbial interactions. These experiments were performed by Dr. Coran Watanabe.

Establishing Microbial Synergism of Pup14A and Pup14B. To investigate the possibility that microbial interplay was necessary for stimulating natural product production, we initially took a split-pool approach. Sixty single colonies were picked and maintained individually on marine agar plates. Colonies were first pooled ten at a time, assayed for the green phenotype, and upon confirmation, the colonies were further split and pooled until it was determined that two microorganisms, designated here as Pup14A and Pup14B, were necessary for production of the blue metabolite.

At this point, two basic scenarios were envisioned. Either one strain is the inducer and the other is the producer, or one strain generates a metabolic precursor that is subsequently converted by the other to give the final product. To distinguish between these possible mechanisms, we performed a series of experiments. We first examined the natural product for toxicity against these strains. Overnight cultures of Pup14A and Pup14B were generated, diluted to an OD_{600} of 0.1, and aliquoted into a 96-well plate. To each culture was added ethanol vehicle (serving as a control) or the purified blue metabolite to give final concentrations of 10 μg/ml,

100 μg/ml, and 1 mg/ml. The cultures were incubated with shaking for an additional 12 hr, and cell density was measured with a Biotek microplate reader. Cellular cytoxicity was not observed with either Pup14A or Pup14B.

Varying the ratio of Pup14A and Pup14B inoculated revealed that equal volumes of each microorganism were most effective at inducing natural product production. Skewing the ratio in favor of Pup14A over Pup14B resulted in a loss of induction, while inoculating with negligible amounts of Pup14A versus that of Pup14B still gave the green phenotype and similar levels of production but with delayed onset. This suggested that Pup14A is the inducer and Pup14B is the producer. Moreover, as the natural product was produced in similar quantities whether equal or skewed cell ratios were utilized, the mechanism of activation does not likely involve the generation of an intermediate by one strain followed by metabolite transfer and biosynthesis by the other strain to give the final product.

A small-molecule exchange assay was performed where the cells of one were introduced into the media of the other. The experiment was carried out at two different time points, 24 and 72 hr, respectively. Two sets of Pup14A and Pup14B cultures were generated and harvested by centrifugation at the appropriate time point, and the media were sterile filtered. The cells of Pup14A were introduced into the media of Pup14B, and Pup14B cells introduced into the media of Pup14A. Induction was not observed in either case. Therefore, small molecule production by Pup14A alone was not sufficient for activation. Induction of the green phenotype might require cell-to-cell contact, or more complex mechanisms could be at work, such as small molecule exchange between microorganisms (29, 78, 164).

To address this issue and decipher between these two mechanisms, a Boyden chamber assay, consisting of two wells separated by a membrane was carried out. The producing strain, Pup14B, was inoculated into the bottom chamber, and Pup14A, the activating strain, into the top chamber. The cells were incubated at 30°C in an incubator, without shaking, for 2 days, extracted with methylene chloride, and compared visually to a set of controls where Pup14B was inoculated into the bottom chamber of one well, the top chamber containing nothing, or Pup14A and Pup14B were mixed together with no membrane separation. Activation of the blue phenotype was observed when Pup14A and Pup14B were mixed together or when separated by a membrane. In fact, activation was slightly enhanced when there was no cell contact. As expected, no activation was observed when Pup14B was cultured alone. Additionally, to ensure that the microorganisms could not penetrate the membrane, both Pup14A and Pup14B were each added to the top chamber of separate wells with the bottom chamber containing medium alone

and incubated for 2 days. In neither case was translocation to the bottom chamber observed. The medium in the bottom chamber remained clear in both cases. The results from these experiments favor a mechanism in which metabolite exchange between the two microorganisms is necessary to induce the green phenotype.

To determine whether the interaction was specific between Pup14A and Pup14B or whether any microorganism could activate Pup14B, Pup14B was cultured with an assortment of other microorganisms. Pup14B was cultured for a week with microorganisms Hon6, Pup16, Surf1, and KM1. Pup14A was also cultured with these same microorganisms serving as a control set. While none of the Pup14A mixed cultures gave the green phenotype, Pup14B cultured with KM1 resulted in activation of the strain, although to a lesser extent than when cultured with Pup14A. Microorganisms were sent to the Centrallbureau voor Schimmelcultures in the Netherlands (CBS) for bacterial classification. The results are depicted in Table 2.

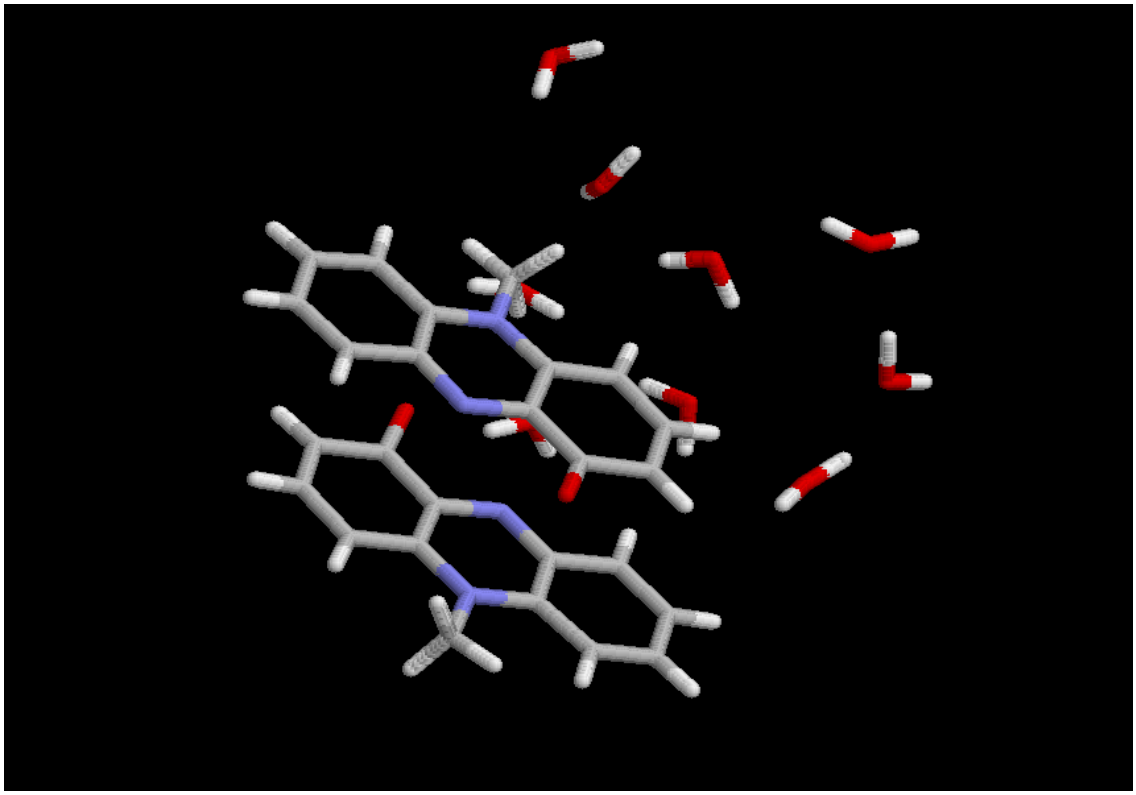
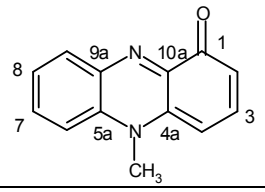


FIG. 13. X-ray crystal structure of pyocyanin.

Pup14A was identified as *Enterobacter sp.* and Pup14B as *Pseudomonas aeruginosa*, a quorum-sensing bacterium (62, 64, 193). These experiments were performed by Dr. Coran Watanabe.

Characterization of the Natural Product. Concurrent to these investigations, we obtained an X-ray crystal structure of the blue natural product (Fig. 13). Literature evaluation of *Pseudomonas aerginosa* taken together with the X-ray crystal structure data revealed that the identity of the compound was likely pyocyanin; however, full characterization on the compound was lacking. Therefore, in addition to providing X-ray analysis, 1D ^{1}H-NMR, and ^{13}C-NMR data, we performed 2D-NMR analyses of the compound including H,H-COSY, HMBC, HSQC, and DEPT (Table 4). The results revealed that some of the ^{1}H-NMR and ^{13}C-NMR signals in the spectra of the compound had been misassigned in the 1989 report (59). The proton at carbon 3 was assigned to carbon 6, and vice versa. H,H-COSY showed coupling between protons 2, 3, and 4. Since protons at positions 3 and 6 have similar chemical shifts, 8.12 and 7.99 ppm, respectively, the data were misinterpreted. Moreover, ^{13}C-NMR peaks corresponding to carbons 5a and 9 and carbons 3 and 10a were also reversed in their assignments. Carbon positions 3 and 9, located at 147.3 and 134.0 ppm, were readily assigned on the basis of HSQC and DEPT

TABLE 4. NMR spectral data for pyocyanin in methanol-d_4



Position	^{1}H δ, mult; *J* (Hz)	^{13}C δ	H→H COSY	HMBC (C→H)	HSQC	DEPT
1		177.5				C
2	6.68, d; 8.7	115.5	3	4	2	CH
3	7.99, t; 8.4	147.3	2,4		3	CH
4	6.66, d; 8.4	95.3	3	2,3	4	CH
4a		136.4		5-Me		C
5a		134.4		5-Me		C
6	8.12, m	116.8	7	8	6	CH
7	8.12, m	138.1	6,8	9	7	CH
8	7.79, t; 7.0	128.0	7,9	6,7	8	CH
9	8.38, d; 8.4	134.0	8		9	CH
9a		138.5		6,7,8		C
10a		146.7		2,4		C
5-Me	4.29, s (3H)	36.7			5	CH_3

experiments, which revealed the presence of an attached proton at these sites. This revealed an inversion of peak assignments between these carbons and carbons 10a and 5a that have similar chemical shifts, 146.7 and 134.4 ppm, respectively. Spectral data on the compound are provided in Table 4 with corrections in place and J-values listed. Isolation and spectral characterization experiments were performed by Dr. Howard Williams, Bennie J. Bench, Dr. Joe Riebenspies, and Scott Angell.

Sequence Analysis of the Phenazine Gene Cluster Regulatory Elements. Pyocyanine belongs to a class of natural products known as the phenazines. We examined Pup14B for the presence of the phenazine gene cluster and its quorum-sensing region by PCR analysis, employing oligonucleotide primers flanking both the las box and qscR regions of the two biosynthetic loci and their respective upstream regions. PCR products were obtained in both cases and sequencing confirmed that both of these regions are intact in the Pup14B strain. As with *P. aeruginosa* PAO1, these results might suggest that the presence of two differentially regulated phenazine biosynthetic operons might give the bacterium the ability to more easily modulate phenazine production and adapt to environmental signals or growth phase (121).

Transcriptional Response to Pyocyanin Treatment. We investigated the mode of action of the compound by profiling the effects of the compound on the yeast transcriptome with Affymetrix oligonucleotide microarrays (Yeast S98 Array). Data were obtained in duplicate at two concentrations of agent, 25 and 250 μg/ml, at 6 hr of drug exposure (Table 5). Differential mRNA gene expression (~2.5-fold) was observed in 45 transcripts of the approximately 6200 genes that comprise the yeast genome. The results are consistent with the proposed role of the drug as an inducer of oxidative damage (76, 130, 140, 160). The primary transcriptional response consisted of genes involved in oxidative damage and stress, cell-wall maintenance and synthesis, and DNA damage. Secondary effects consisted of genes involved in protein synthesis, drug resistance, transport, stress response, as well as metabolic genes, cell-cycle-related effects, and several unclassified transcripts. Individual transcriptional changes are summarized in Table 5. Transcriptional profiling experiments were performed by Dr. Coran Watanabe.

Confirmation of Selected Gene Changes by RT-PCR. Validation of the transcriptional effects induced by pyocyanin treatment was demonstrated by semiquantitative reverse transcriptase (RT)-PCR of an arbitrary subset of the gene changes, including glutathione peroxidase (YBR244W), catabolic serine (threonine) dehydratase, iso-2-cytochrome c (YEL039C), and cystathionine β-lyase (YGL184C). The procedure, while less sensitive and quantitative than real-time PCR, is capable of confirming a transcriptional response, the presence

TABLE 5. Selected results of transcriptional array profile.

Probe Set	ORF/Gene Description	Fold Change vs. Control at: 25 μg/ml	250 μg/ml
Protein-Synthesis Related			
10210_at	YLR136C/TIS11,tRNA-specific adenosine deaminase; transcription factor	—	4.9
10226_at	YLR107W/REX3, RNA exonuclease involved in RNA processing/degradation	—	2.8
9413_f_at	YMR230W/RPS10B, ribosomal protein	—	3.2
5957_at	YMR322c/SNO4, possible chaperon and cysteine protease	—	2.8
9667_at	RRN6P,RRN7P/rRNA transcription factors (RNAP I specific initiation factor)	—	3
10976_at	YJR047C/ANB1 translation initiation factor lF5A.2	—	-4.7
Oxygen-Stress Related			
5697_at	YER042w/MXR1, responsible for the reduction of methionine sulfoxide	—	2.6
7109_at	YBR244W/GPX2, glutathione peroxidase	—	4.7
7847_at	YPL088W/similar to aryl alcohol dehydrogenase	2.4	3.1
8988_at	YNL134C/putative alcohol dehydrogenase implicated in oxygen stress response	—	3.7
4924_at	YGR088W/ctt1 cytoplasmic catalase T	-2.6	—
5751_at	YEL039C/cyc7 iso-2-cytochrome c	-3.3	24.3
10977_at	YJR048W/cyc1 cytochrome c isoform 1	-2.6	—
General Drug Resistance			
10358_at	YLL028W/TPO1, vacuolar polyamine/H+ antiporter	—	4.1
4573_at	YHL035C/VMR1, multidrug resistance protein, ABC protein	—	2.5
4778_at	YGR213C/RTA1, integral membrane protein involved in 7-aminocholesterol resistance	—	5.9
8923_at	YNLO65W/AQR1, membrane transporter	—	4.7
9117_at	YNL231C/PDR16, protein involved in lipid synthesis and multidrug resistance	—	2.5
9328_at	YMR279C/member of a family of multidrug resistance proteins	10.6	—
Transport Proteins			
4936_at	YGR055W/MUP1, high affinity methionine permease	—	3
10852_at	YJR150C/DAN1 protein only expressed during anaerobic growth	—	-2.9
10795_at	YKL217W/JEN1 lactate and pyruvate permease	-2.5	—
9065_at	YNL194C/integral membrane protein localized to cell periphery	-2.3	-4.4
8528_at	YOR11w/AUS1 ATP-binding cassette transporter	—	-2.8
4388_at	YHR139C/SPS100, sporulation-specific wall maturation and cell-differentiation protein	—	3.2
6113_at	YDR371w/CTS2, sporulation-specific chitinase involved in cellular differentiation	—	4.7
7409_at	YBL043W/ECM13 involved in cell-wall synthesis	—	2.5
10853_at	YJR151C/DAN4 cell-wall mannoprotein, similar to YKL224c, Sta1p	—	-2.6
8526_at	YOR009W/TIR1 cell-wall mannoprotein, involved in cell-wall maintanance	—	-4.0
8527_at	YOR010C/TIR2 cold shock induced protein	—	-3.3
9628_at	YMR006C/PLB2 lysophospholipase/phospholipase B	—	-3.6
DNA Related			
5970_at	YDR501w/PLM2, plasmid maintenance protein; mutant shows 2mu-m plasmid instability	8.6	7.3
4045_s_at	YIL066C/RNR3, ribonucleotide reductase, large subunit expressed only after DNA damage	8.4	—
Amino Acid Metabolism			
5196_at	YGL184C/STR3, cystathionine beta-lyase	2	3.6
6915_at	YCL064C/CHA1, L-serine/L-threonine deaminase	6.1	12.6
Isoprenoid Metabolism			
8642_at	YOL101C/IZH4 implicated in zinc homeostasis, membrane protein	—	-8.1
9854_at	YLR450W/HMG2 3-hydroxy-3-methylglutaryl-coenzyme A reductase A	—	-2.6

TABLE 5, cont.

Probe Set	ORF/Gene Description	Fold Change vs. Control at: 25 μg/ml	250 μg/ml
Cell Cycle		Fold Change vs. Control	
8380_at	YOR178c/GAC1 ser/thr phosphoprotein phosphatase, cell-cycle checkpoint	—	-2.5
Other			
4194_at	YIL117C/PRM5, pheromone-regulated protein, induced during cell-integrity signaling	—	2.6
5372_at	YFL014W/HSP12, heat-shock protein	—	2.6
8338_at	YOR226C/mitochondrial matrix protein required for synthesis of iron sulfur proteins	—	3
8399_at	YOR152C/protein of unknown function localized to ER	—	2.9
8399_at	YOR107W/RGS2, negative regulator of glucose induced cAMP signaling pathway	—	2.7
9305_at	YMR299C/DYN3, light intermediate chain of dynein	—	3.8
9567_at	YMR081C/ISF1 involved in suppression of mitochondrial splicing defect	—	-2.7

Table was generated by Dr. Coran Watanabe.

or absence of a transcriptional effect. Normalization of the RNA/cDNA was based upon glyceraldehyde- 3-phosphate dehydrogenase (GAPDH) that was unaffected by drug exposure. This analysis was performed by Dr. Coran Watanabe and a graph of the results can be found in Appendix A.

DISCUSSION

Biosynthesis of Pyocyanin and Quorum Sensing Control. Two seven-gene phenazine biosynthetic loci have been identified in *Pseudomonas aeruginosa* PAO1 (121). The two operons, designated as phzA1B1C1D1E1F1G1 and phzA2B2C2D2E2F2G2, exhibit high homology to phenazine biosynthetic operons from *P. fluorescens, P. aureofaciens*, and *P. chlororaphis* PCL1391 (35, 122, 156). Additionally, Southern hybridization analysis for the presence of the core phenazine biosynthetic genes phzA-G in 30 bacterial strains, including other *P. aeruginosa* strains, showed that the cluster could be detected in 21 of these strains (121). Two additional genes, phzM and phzS, located in the vicinity of the phzA1B1C1D1E1F1G1 operon in strain PAO1 have been shown to convert phenazine-1-carboxylic acid (PCA) to pyocyanin (121). Upstream of the phzA1B1C1D1E1 F1G1 operon is a well-conserved putative promoter element required for quorum control, the las box, but does not precede the phzA2B2C2D2E2F2G locus (213). Instead, a gene, qscR, encodes a protein that negatively regulates expression of a number of quorum-sensing controlled genes including the phzA1B1C1D1E1F1G1 operon (37, 110).

Quorum sensing is a communication strategy used by diverse bacterial species to transmit information intercellularly, enabling the organism to express specific genes in a coordinated fashion, leading to a rapid and full-blown virulence cascade (78, 164). A complex

network of regulatory factors govern production of secondary metabolites in *Pseudomonas aeruginosa* that involve two quorum-sensing systems, las and rhl (49, 62, 150, 151, 204). At least three signaling molecules have been identified that act as intercellular communication signals. The acyl homoserine lactone signals, N-(2-oxododecanoyl)homoserine lactone and N-butyryl homoserine lactone, are each autoinducers, together controlling 6%–11% of the *Pseudomonas* genome (179, 204, 205, 213). The third intercellular signal is a quinolone compound, 2-heptyl-3-hydroxy-4-quinolone, that serves as a regulatory link between the las and rhl quorumsensing systems and controls the synthesis of multiple virulence factors (49, 204). In many strains of *Pseudomonas aeruginosa*, a basal level of autoinducer is produced at low culture densities, and as the population grows, autoinducer concentrations increase. On reaching a threshold concentration, the autoinducers activate the virulence cascade, including that of secondary metabolite production (60, 62). Pup14B on the other hand, the *Pseudomonas aeruginosa* strain we have isolated, did not induce the blue phenotype in culture without activation by another strain, reflecting some likely evolutionary differences with this strain. KM1, like Pup14A, is of the genus *Enterobacter*, accounting for its ability to activate production of the compound. Both *Pseudomonas sp.* and *Enterobacter sp*. have been identified in microbial biofilms among a variety of virulent pathogens, some of which have been shown to be regulated by acylhomoserine lactone (AHL) quorum-sensing signals (86). Moreover, cell-free supernatants of *Pseudomonas aeruginosa* and *Enterobacter agglomerans*, were shown to contain diketopiperizines (DKPs) capable of activating an N-acylhomoserine lactone (AHL) biosensor (82).

Transcriptional Effects of Pyocyanin Exposure in Yeast. The oxidative damage, cell-wall biosynthesis, and DNA damage effects observed with pyocyanin treatment are each fully consistent with the agent's documented ability to redox cycle and oxidize glutathione (L-γ-glutamylcysteinylglycine) (130, 141). A map correlating pyocyanin's biochemical actions with its transcriptional effects is depicted in Fig. 14. Under aerobic conditions, pyocyanin is reduced by NADH or NADPH by electron transfer, passing those electrons to O_2 to give superoxide radical ($O_2^{\bullet-}$) and hydrogen peroxide (H_2O_2). These oxidative species result in the oxidation of glutathione (GSH), the primary oxygen detoxification species and chemical modulator of redox potential within the eukaryotic cytosol, (124, 128, 212) to its oxidized state (GSSG). We observed that glutathionine peroxidase (GPX2) was upregulated by 4.7-fold. Glutathione peroxidase neutralizes destructive peroxides by catalyzing their reduction using glutathione as an electron source (8). Pyocyanin can also directly oxidize glutathione to form pyocyanin radical

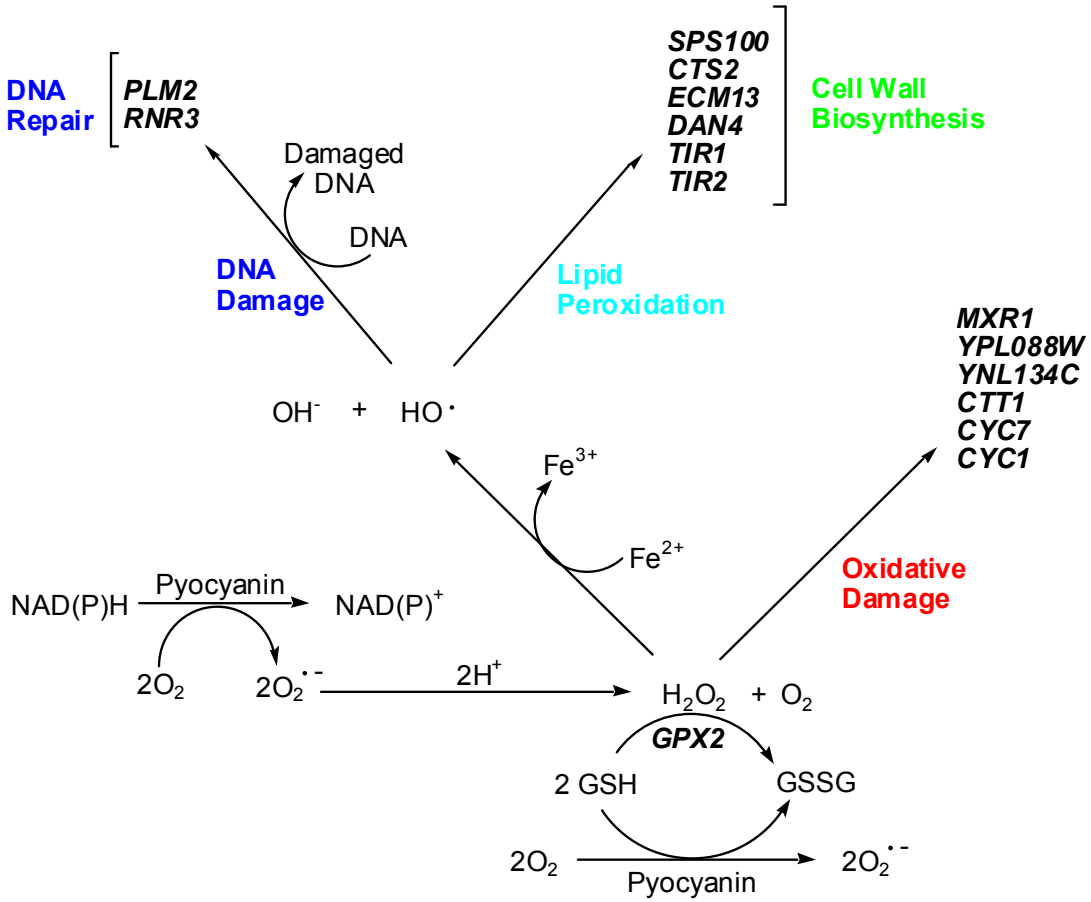


FIG. 14. Correlation of gene changes generated from yeast transcriptional profile data. Graphic produced by Dr. Coran Watanabe.

and superoxide radical ($O_2^{\bullet -}$), placing further oxidative stress on the cell (141). Reactive oxygen species generated through pyocyanin exposure can also oxidize residues of proteins and elicit oxidative stress-defense systems. Methionine sulfoxide reductase, for example, an oxidative damage-repair protein that acts by catalyzing the reduction of methionine sulfoxide and other methyl sulfoxides to methyl sulfides, was upregulated in its expression by 2.6-fold, consistent with this notion (129). Other oxidative stress proteins upregulated with pyocyanin treatment included two alcohol dehydrogenases, YPL088W and YNL134C, differentially expressed by 3.1-fold and 3.7-fold, respectively, at the high dose (65). Although their exact substrates are unknown, both have previously been associated with oxygen stress response. Likewise, we observed upregulation of glutathione S-transferase by 2.4-fold, a class of enzymes recruited for xenobiotic detoxification (36, 124). Somewhat paradoxically, three genes normally upregulated by increases in oxygen concentration were found to be downregulated. These were ctt1, a

cytoplasmic catalase (22.6-fold) (75) and two cytochrome c encoding transcripts, cyc7 (23.3-fold and 24.3-fold at low and high doses, respectively) and cyc1 (22.6-fold at high dose), which function in respiration and electron transport (27). Downregulation of cytochrome c peroxidase genes has been observed with treatment of *Rhodobacter sphaeroides* with hydrogen peroxide (221). The mechanism by which organisms cope with damage inflicted by reactive oxygen species varies from organism to organism, whereby nonessential genes are downregulated (221). The involvement of catalase and superoxide dismutase (SOD) have both been detected in yeast exposed to pyocyanin at 0.5–1.5 hr post treatment (160). The absence of these genes from the transcriptional profile (measured at 6 hr) could thus reflect the timing of the response or half life of the mRNA of these genes.

Further evidence of the oxidative damage induced by pyocyanin was seen in the upregulation of genes associated with DNA damage. Although hydrogen peroxide itself does not inflict DNA damage, it is readily converted to hydroxyl radicals by transition metals such as iron via the Fenton reaction (Fig. 14), resulting in the production of DNA radicals (95, 148). Pyocyanin itself has also been shown to bind double stranded DNA with no base specificity and inhibit DNA template controlled RNA synthesis (106). Two DNA damage-response proteins, RNR3 and PLM2, were upregulated with pyocyanin treatment. RNR3, the large subunit of ribonucleotide reductase, was upregulated by 8.4-fold. Ribonucleotide reductases catalyze the reduction of nucleotide diphosphates to their corresponding deoxynucleotide diphosphates, playing a central role in DNA synthesis (91) and in response to DNA damage (51, 228). PLM2 is a plasmid-maintenance protein and was upregulated in its expression up to 8.6-fold. Expression of this gene product is coupled with DNA damage (94) and has been shown to activate genes involved in DNA damage checkpoint and repair (84).

The damaging effects of hydroxyl radicals on DNA notwithstanding, lipoperoxidation can also result, causing enhanced cell-wall permeability (79, 170, 181, 221). Interestingly, a variety of cell-wall proteins were differentially expressed with pyocyanin treatment. While these effects have not been previously documented with pyocyanin exposure, they are consistent with the ability of the compound to redox cycle and generate reactive oxygen species. For example, SPS100, a protein that plays a protective role during the early stages of spore-wall formation, was enhanced in its expression by 3.2-fold (108). CTS2, a sporulation-specific chitinase expressed during spore-wall synthesis was upregulated by 4.7-fold. Deletion of CTS2 has been shown to decrease surface density of the spore (52). ECM13, another protein involved in cell wall biosynthesis, was increased in its expression by 2.5-fold, although the exact function of this

gene has yet to be established (117). Cell-wall mannoproteins, DAN4, TIR1, and TIR2, were shown to be downregulated with pyocyanin treatment (22.6-fold, 24.0-fold, and 23.3-fold, respectively). Each of these genes is responsible for the synthesis of cell-wall mannoproteins that are necessary for anaerobic growth (1). Similarly, PLB2 a phospholipase involved in lipid metabolism was also downregulated. Presumably, enhanced levels of these highly oxidative species caused further downregulation of these genes compared to standard aerobic conditions (61).

Collectively, the gene changes resulting from pyocyanin support the purported role of this compound as an inducer of oxidative damage. Although the transcriptional effects of the natural product have not been previously reported, we were intrigued to find that differentially expressed transcripts roughly paralleled those observed with hydrogen peroxide treatment (77, 112, 147, 158, 221, 227). Pyocyanin has also been screened against a yeast deletion library as a means to identify potential protein targets (160). Targets identified in this study also supported the ability of this compound to redox cycle and promote oxidative stress encompassing proteins involved in the cell cycle, electron transport and respiration, epidermal cell growth, protein sorting, vesicle transport, and the vacuolar ATPase (160).

Conclusions. In summary, we probed the hypothesis that microbial interactions in the marine environment can impact natural product production. Marine microbial populations were cultured and examined for antibiotic activity. Among the mixed cultures examined was a bacterium that was found to act synergistically with select microbial species to produce a blue compound. The bacterium was identified as *Pseudomonas aeruginosa*, and the blue metabolite as pyocyanin. Full structural analysis of pyocyanin, including X-ray analysis and 2D-NMR, was performed. Corrections were made to ^{1}H-NMR and ^{13}C-NMR assignments of the compound that were misreported in the chemical literature. Taken together, the data gleaned from these investigations suggest the importance of investigating not only organic extracts generated from single colonies, but also those from mixed microbial populations for bioactivity. The strategy could lead to the discovery of otherwise unexpressed compounds. While the procedures detailed here were for the microbial production of a blue metabolite, which certainly facilitated our analyses, the approach is broadly applicable toward the investigation of metabolites of other organisms. The primary procedural modification would involve examination and comparison of extracts generated from microbial communities and pure cultures by TLC and LC/MS.

CHAPTER III

CONSTRUCTION AND ANALYSIS OF A BAC LIBRARY OF *CHROMOHALOBACTER SP.*

INTRODUCTION

Recent advances in molecular microbial ecology have revealed that microbial diversity in nature far exceeds that reflected in laboratory strain collections since the majority of naturally occurring species either cannot withstand the conditions imposed by traditional culture media, or do not grow sufficiently to produce observable colonies under these conditions (55, 93, 224). Cultivation-independent analyses such as microbial rRNA sequencing have dramatically altered phylogenic definitions of major microbial classes and suggest that a large fraction of indigenous microbial communities are composed of new but uncultivable microbial groups (4, 10, 93, 175, 195, 224). Therefore, improving organism culturability could unveil previously untapped resources and natural products with novel structural frameworks. Some success has been achieved using high-throughput culturing procedures that utilize the concept of extinction culturing to isolate cells in very low nutrient medium and small volumes (40).

An alternative strategy is to "shotgun" clone environmental genomic DNA and express these whole genomic bacterial communities within heterologous hosts. Recent studies with soil samples and bacterial symbiont genomes are beginning to reveal the potential of this technology for both natural product and enzyme discovery (18-24, 41, 63, 167). Molecular screening of recombinant environmental libraries with degenerate probes recognizing specific biosynthetic gene sequences have also proved quite fruitful for the identification of natural product biosynthetic pathways (41, 153, 154, 174, 175). In some instances, heterologous expression of these clones has resulted in metabolite production (176).

To improve and expand on these approaches, it would be desirable to generate libraries with even larger genomic inserts. Cosmids are limited to approximately 30-50 kilobases of genetic material, and many pathways are known to be significantly larger than this (169, 178, 219). Larger inserts would increase the probability of cloning intact pathways and would allow a larger representation of genetic material in a given library size. It would also be advantageous to expand the host range of the library to eukaryotes such as *S. cerevesiae*, increasing the probability of observing expression of fungal pathways.

Here we provide details on the construction of a large construct BAC library from a previously undescribed species of *Chromohalobacter* collected from Honokohau, Maui. We modified the BAC-yeast shuttle vector pClasper so that it contains the lacZ multiple cloning site of pBelloBAC11, producing pClasperlacZ, for the construction of our library, which allows blue-white screening as well as propagation in a fungal host.

MATERIALS AND METHODS

General. Pulsed Field Gel Electrophoresis (PFGE) was performed on a BioRad CHEF DR II apparatus (Hercules, CA). Electroporation of cells was carried out with a BioRad MicroPulser. The *E. coli* strain DH10B was obtained from Prof. Dennis Gross, Texas A&M University, Department of Plant Pathology. *N. crassa* was obtained from the lab of Prof. Deborah Bell-Pederson (Texas A&M University, Department of Biology). Flow cytometry experiments were carried out on a Becton-Dickinson FACS Calibur flow cytometer (Franklin Lakes, NJ).

Construction of pClasperlacZ. The lacZ multiple cloning site and its promoter were amplified from the vector pBeloBAC11 with primers, ASC1LACZF (5'-GGC GCG CCA ATT TTT TTA AGG CAG TTA TTG GTG-3', the AscI site is underlined) and MLUILACZR (5'-ACG CGT TTG GCG GGT GTC GGG GCT GGC TTA A-3', the MluI site is underlined), tagged with restriction enzyme sites AscI and MluI, respectively. The resulting DNA fragment was cloned into the cloning vector pCR 2.1 TOPO (Invitrogen; Carlsbad, CA). The plasmid was miniprepped and digested with the enzymes AscI and MluI. The pClasper vector was also digested with AscI and MluI, phosphatase treated, and subsequently ligated to the lacZ construct with T4 ligase. The resulting colonies were miniprepped, screened by digestion with AscI and MluI, and the respective clone confirmed through sequencing. pClasperlacZ construction was performed by Dr. Coran Watanabe.

Isolation of Hon6. The Hon6 strain was identified from marine microbial isolates obtained from Honokohau, Maui. Seawater and sediment were plated onto marine bacterial plates. Marine agar plates were prepared with slight modification to standard literature protocols (42). Yeast extract (1 g; Difco), Bactopeptone (1 g; Difco), glucose (2 g), and Bactoagar (10 g; Difco) were combined in 500 mL of seawater, prepared by mixing instant ocean (19 g, Petco) with deionized distilled water. The agar suspension was autoclaved, cooled to 50°C, to which miconazole (10 mg/l) was added. Hon6 was identified from single colonies generated through repeated streaking of microorganisms onto fresh plates. Taxonomy of the strain was assigned by

the Centrallbureau voor Schimmelcultures (CBS) in the Netherlands. Hon6 was isolated by Dr. Coran Watanabe.

Cytotoxicity Assay. Jurkat cells were cultured in RPMI medium (Gibco BRL) supplemented with 10% FBS (Gibco BRL) and subsequently plated at a density of 200,000 cells/mL into a 48-well plate (Falcon, VWR International). To each culture was added 10 μL of Hon6 crude extract (solubilized in DMSO; stock solutions of 1, 10, and 100 mg/mL to give final concentrations of 10 μg/ml, 100 μg/ml, and 1 mg/ml) or DMSO alone. All assays were executed in duplicate and allowed to incubate for 24 h at 37°C under a 5% CO_2 atmosphere. IC_{50} values were estimated with a hemacytometer (VWR International). Cytotoxicity assays were performed by Dr. Coran Watanabe.

Library Construction. DH10B cells containing pClasperlacZ were grown in 2 L culture and harvested by centrifugation (222). The plasmid was isolated using a Qiagen Maxiprep kit (Valencia, CA). The vector (10 μg) was digested and dephosphorylated (222). Following dephosphorylation, the vector was further purified by self-ligation and electrophoresis (152) and isolated from the surrounding agarose with a Qiagen gel-extraction kit.

Hon6 cultures (500 ml; 1 l Erlenmeyer flask) were grown for 20 h at 30°C and 250 rpm. The cell pellet was harvested by centrifugation, washed with TE buffer (169), and resuspended in an equal volume of TE. To the suspension was added an equal volume of molten agarose (1.5% m/v agarose, 10% m/v sucrose, and 40% v/v glycerol). This suspension was streaked out onto a Petri dish cooled on ice and allowed to solidify. The solidified agarose was immersed in wash buffer (222) containing 0.15% v/v β-mercaptoethanol and incubated on ice for 20 h. The wash buffer was drained and the agarose incubated with fresh wash buffer lacking β-mercaptoethanol for one hour at 37°C. The wash buffer was replaced with lysis buffer (222) and lysed at 50°C for 48 h. The lysed agarose material was embedded in a pulsed field agarose gel and run with settings of 15 second initial and final switch times, 6 V/cm, 14°C, and a 12 h run time in 0.5X TAE buffer (169). The agarose material was cut out with a razor blade and stored in 0.5 M EDTA, pH = 8 at 4°C until use.

Restriction digest optimization was performed (152). Once optimal conditions were established, a preparative digest and double size fractionation was carried out (152) employing eight agarose plugs (150 μl), 250 U of HindIII per digest, and 0.5X TAE buffer instead of 0.5X TBE. The resulting high molecular weight Hon6 genomic DNA was isolated by electroelution with dialysis tubing (222).

The genomic DNA was ligated to the dephosphorylated pClasperlacZ with T4 ligase (two 300 ng scale ligation reactions were necessary to produce the required number of clones, 1536) and electroporated (2 kV; 0.2 cm gapped cuvette, BioRad) into electrocompetent DH10B cells. Electrocompetent DH10B cells were prepared as described in Current Protocols in Molecular Biology (7). Colonies (1536) were arrayed by hand into four 384 well microtiter plates and stored at -80°C.

Insert Size Characterization. Individual clones (24 clones) were cultured in LB broth (3 ml) supplemented with chloramphenicol (15 mg/l) and miconazole (3 mg/ml). Clones were mini-prepped and the nucleic acid pellet dissolved in TE buffer (7). Samples (10 μL) were digested with 2 U of I-SceI and analyzed by PFGE (initial switch time: 20 sec.; final switch time: 6 V/cm; 14 °C; run time: 12 h). MidRange PFG Marker II from New England Biolabs (Beverly, MA) was utilized as a size reference.

FACS Analysis of Genome Size. Overnight cultures (3 ml) were prepared for all species except *Neurospora*, which was cultured on an agar slant. *E. coli* and *S. cerevesiae* were grown for 16 h according to standard procedures in LB and YPD liquid media, respectively (7, 169). *Neurospora crassa* was propagated on an agar slant of Vogel's media according to standard procedures (45). Liquid cultures were pelleted by centrifugation. Pellets (50 μL portions) were resuspended in 1 mL of 60% Galbraith buffer in EtOH (90). *Neurospora* multinucleate spores were harvested by scraping the surface cultures with a spatula. The cells

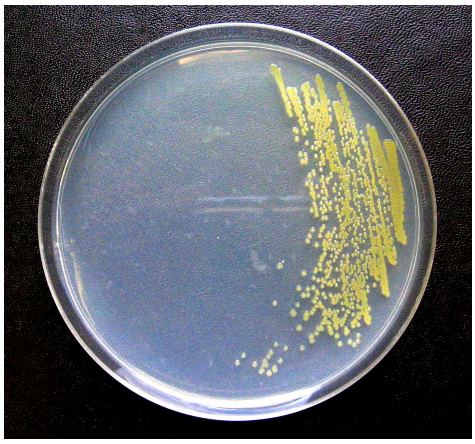


FIG. 15. Photograph of the Hon6 strain cultured on a marine agar plate.

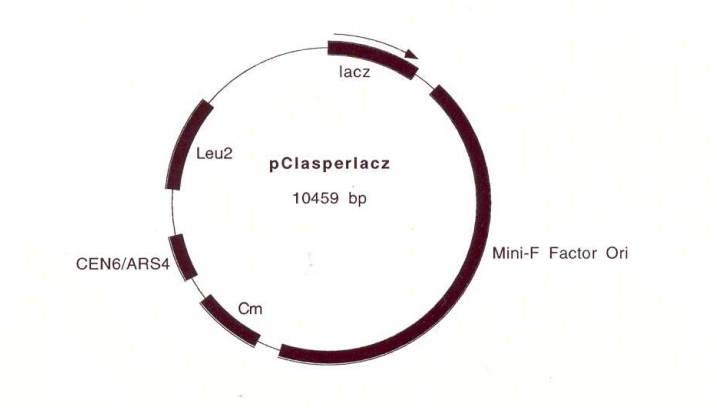


FIG. 16. Plasmid map of the shuttle vector pClasper*lacZ.*

were resuspended in 60% Galbraith buffer as detailed above and filtered through a 20 μm nylon mesh.

Samples were incubated at 4°C overnight, pelleted by centrifugation and resuspended in 500 μl Galbraith's buffer. The samples were subsequently incubated at 37°C for 1 h, pelleted by centrifugation and resuspended in 500 μl of TE buffer containing propidium iodide (0.1 mg/ml). Following 1 h of incubation, the cells were again pelleted, resuspended in 500 ml of TE and analyzed by flow cytometry. Genome sizing experiments were performed by Dr. Roger Smith, Thom Kelly and Scott Angell.

RESULTS

Isolation and Strain Identification of Hon6. Hon6 was identified from marine microbial isolates obtained from the coast of Honokohau, Maui. The Centrallbureau voor Schimmelcultures (CBS) in the Netherlands identified the strain as a new species, 98% identical to *Chromohalobacter* (Fig. 15). In cell-based assays, crude extracts of the microorganism showed moderate activity against Jurkats (IC_{50} = ~100 μg/ml), a cancerous T-cell line, while no effect was observed against *E. coli*, *B. subtilis*, and *S. cerevisiae*. Isolation of Hon6 was performed by Dr. Coran Watanabe.

Construction of pClasperlacZ. pClasper (17), a vector that can maintain large DNA fragments in both bacteria and yeast was modified to install the lacZ gene within the multiple cloning site (MCS region) of the vector (Fig. 16). The lacZ MCS was successfully subcloned by PCR amplification and digestion, then ligation into the cloning region of pClasper. The lacZ

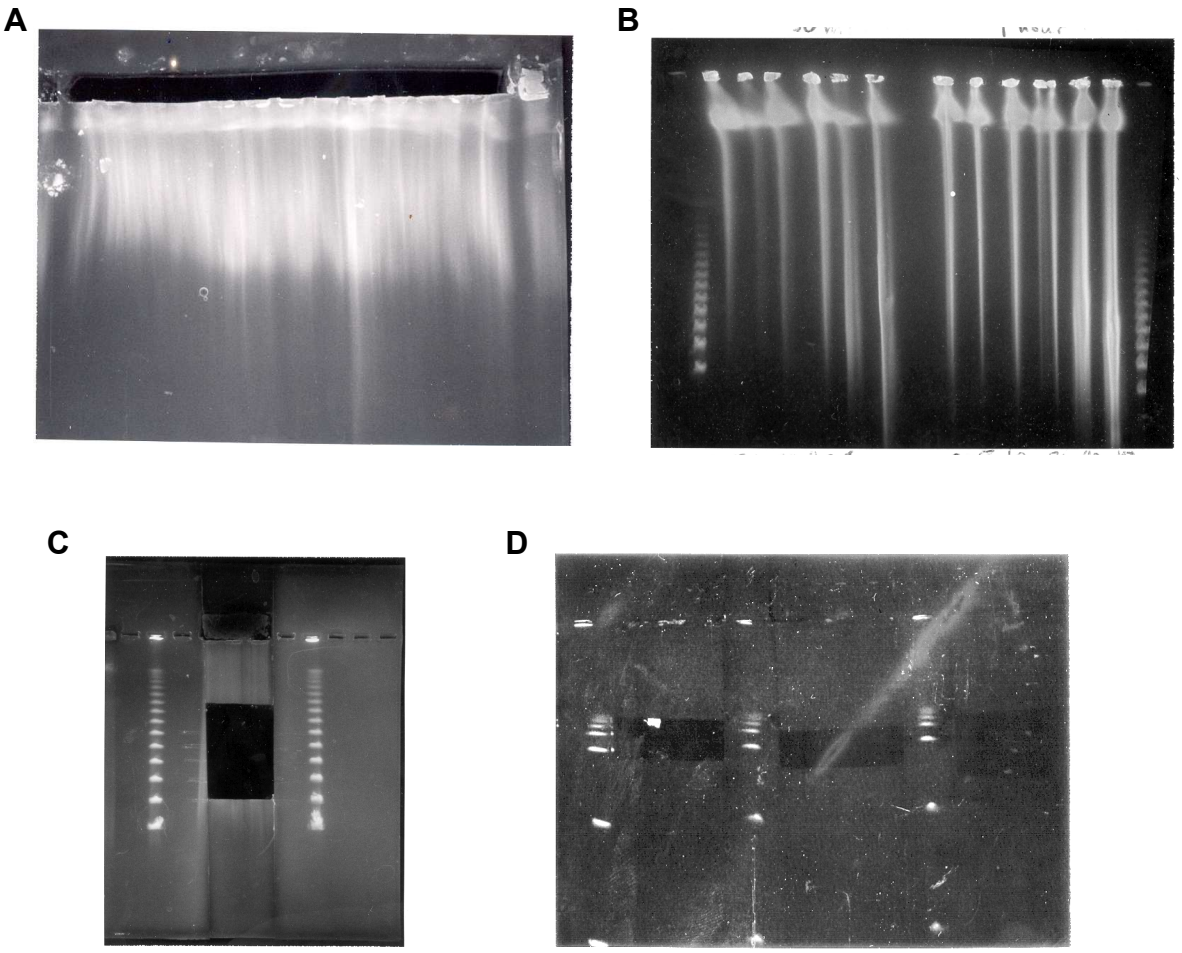


FIG. 17. DNA preparation for ligation into BAC vector. (a) High-molecular weight genomic DNA isolated from Hon6, showing removal of contaminating low-molecular weight fragments. Agarose embedded high-molecular weight DNA has been cut out of the large, rectangular hole in the upper part of the gel. (b) Optimization of restriction enzyme partial digestion with HindIII. The ladder is a NEB MidRange PFG markers, which consist of concatamers of λ-virus. Each band represents an increase of 48 kb. Increasing concentrations of enzyme were applied, from left to right. (c) Bulk digestion and first size fractionation with HindIII partially digested DNA. Ladder is the same as in (b). DNA was removed as three roughly equal sized horizontal blocks from the center region between 100 kb and 400 kb. (d) Second size selection to eliminate small fragments with altered CHEF gel conditions. DNA of the optimum size is compressed near the "zone of retention," which is clearly indicated by the compression of the bands of the λ-ladder. The desired region is cut out, allowing removal of lower-molecular weight DNA without significant dilution of the desired fraction. The ladder is the same as in the previous gels. Migration differences are due to altered running conditions.

modification allows for blue/white colony selection and elimination of empty constructs during library construction. In the long term, the vector might allow for the housing and expression of biosynthetic pathways in both prokaryotic and eukaryotic hosts. Construction of pClasperlacZ was performed by Dr. Coran Watanabe.

Hon6 BAC Library Construction and Initial Characterization of Clones. High molecular weight genomic DNA was isolated utilizing agarose as a DNA stabilizer. Cultures of Hon6 were pelleted by centrifugation, embedded in molten agarose, washed and lysed with

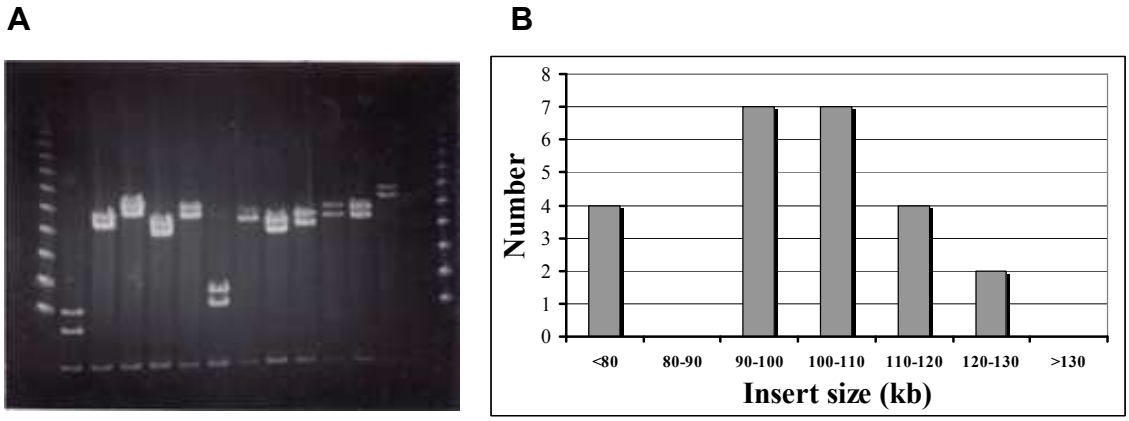


FIG. 18. Library insert size analysis.(a) Representative clones from the Hon6 BAC library cut with I-SceI. (b) Graph depicting the number of clones with specific insert size ranges, of the 25 clones analyzed.

detergents. DNA degradation occurred to a lesser extent when the embedded cells were either streaked in a thin layer or cut into thin sheets prior to lysis. We attribute this to an increased surface area, which enhances the diffusion of nuclease inhibitors, detergents, and proteases into the agarose matrix. Isolation of megabase DNA was achieved by pulsed field gel electrophoresis (Fig. 17A), which eliminated small fragments of DNA from the gel plugs. The agarose plugs containing the megabase DNA were fragmented and digested with the restriction enzyme HindIII. Restriction digest optimization (Fig. 17B) showed that between 200 and 400 units of HindIII was optimal for digestion of 150 µl of macerated gel fragments, and the bulk digestion was performed with 250 units per 150 µl aliquot. A double size selection by pulsed field gel electrophoresis was carried out to generate a narrow insert size distribution (Fig. 17C and Fig. 17D). Following ligation and transformation into DH10B cells the clones were analyzed (Fig. 18A). Four (16%) were found to contain no insert or an insert of negligible size, while those containing inserts, averaged 105 kb. The size distribution shows a large percentage of clones in the 100 kb size range with a few scattered individuals at sizes below 50 kb and a strong, roughly exponential decrease in the population of clones as the size increases beyond 110 kb (Fig. 18B).

Estimation of Genome Size by Flow Cytometric Analysis. Calculating the number of clones required to allow good coverage of a genomic library requires prior knowledge of the genome size of an organism. We therefore measured the genome size of Hon6 by flow cytometric analysis. The nuclear DNA content of the Hon6 strain was evaluated by flow cytometry after PI staining relative to *Neurospora crassa*, which served as a standard (Fig. 19A and Fig. 19B). Genome sizes of *E. coli* and *S. cerevisiae* were also measured to assess the

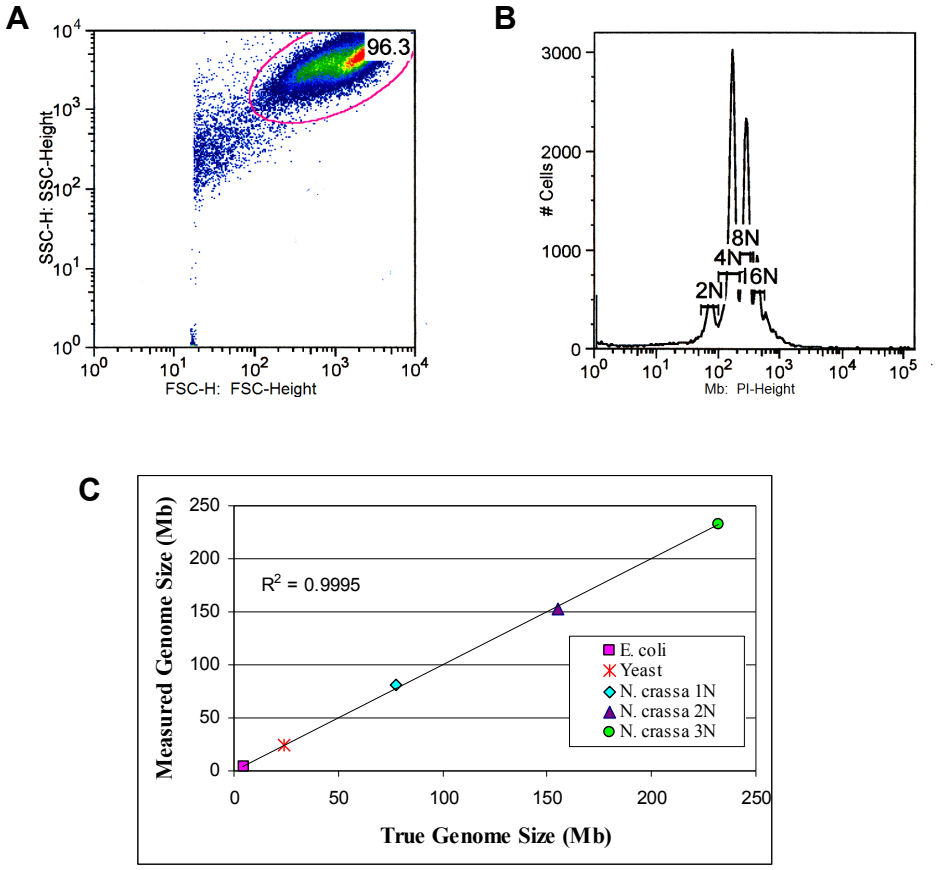


FIG. 19. FACS analysis of genome size. (a) *N. crassa* cell population, graphed by forward and side light scattering of the cytometer laser. (b) Genome size histogram for *N. crassa* multinucleate spores in various nuclear states. Fluorescence intensity is plotted on the horizontal axis, versus cell couts on the vertical axis. (c) Standardization curve of organisms with established genome sizes.

degree of accuracy. The technique can be employed to estimate the genome size of an organism since fluorescence intensity is directly proportional to the DNA content of the cells (11, 80). A calibration curve was generated resulting in the linear graph shown in Fig. 19C and was used to calculate the genome size of *S. cerevisiae* and *E. coli*, which are also plotted on the graph for comparison. Measured values were accurate to within 7% of the known standards (Fig. 19D). For example, the genome size of *E. coli* was calculated to be 4.33 Mb while its true genome size is 4.63 Mb. Likewise, the genome of *S. cerevisiae* was measured to be 24.7 Mb while its established genome size is 24.14 Mb. The nuclear DNA content of Hon6 was estimated by this flow cytometric approach to be 3.89 ± 0.27 Mb (Table 6).

DISCUSSION

Construction of the Library. Of several organisms which were candidates for BAC library construction, this particular strain of *Chromohalobacter,* which we have designated Hon6, proved to be the most amenable to manipulation. It was one of only two strains to thoroughly lyse after being embedded in agarose, which could be visualized by the embedded cell mixture turning from turbid to clear, and was the only strain to give high yields of high molecular weight DNA. Most others gave little DNA or showed extensive degradation to low-molecular weight fragments. In addition, finding proper digestion conditions was very difficult, and it took large quantities of restriction enzyme to achieve a usable level of digestion.

The quality of the vector preparation was also a key to the construction of this library. Particularly troublesome were the low levels of production due to its single copy F-factor origin of replication by *E. coli*, which resulted in large cultures being required to produce tiny quantities of vector which were highly impure, and difficulties with dephosphorylation. The first problem was later solved by inserting linearized pUC19 into the HindIII site of pClasperlacZ to generate pUClasperlacZ, which replicates at high-copy number due to the presence of the pUC origin of replication. A single 5 ml culture is sufficient for a 10 μg preparation of this plasmid in much higher quality. The plasmid can then be digested and dephosphorylated, and the pUC19 fragment removed by electrophoresis. Imperfect dephosphorylation led to two different, equally difficult, scenarios. Excessive exposure to CIP led to the generation of large quantities of false positive clones that show a white phenotype in the blue-white screen but actually contain no inserts. Too little dephosphorylation led to excessive numbers of blue colonies. Religation of the vector followed by electrophoretic purification of linearized vector from circularized and concatamerized vector species helped with insufficient dephosphorylation, but was unable to improve batches which were excessively treated. Well prepared vector actually gave very few

TABLE 6. Genome sizes of standards and Hon6.

Standards	Measured Genome Size (Mb)	True Genome Size (Mb)	Error (%)
E. coli	4.33	4.63	6.5
S. cerevesiae	24.7	24.14	2.3
N. crassa (2N)	81.2	77.3	5
N. crassa (4N)	153	155	1.3
N. crassa (6N)	232	232	-
Hon 6	3.89	-	-

blue colonies, so the lacZ site and blue white screening is not necessary if vector preparation is sufficiently good. It was useful for the rapid determination of the quality of the vector preparation, however.

Insert Size Analysis. Insert sizes were highly satisfactory. Most were greater than 90 kb, with the exception of a scattered few which were much smaller. It is tempting to say that these are stray low-molecular weight fragment impurities which were retained in the high-molecular weight fraction. However, DeLong reports similar observations with the marine BAC library that they produced (10). Smaller inserts were associated with 16S rDNA sequences, and overall the isolation of 16S rDNA was rare within the library. This is an indication that either these sequences or sequences which are closely linked were toxic to the host, resulting in rearranged plasmids and cloning bias against these types of sequences. This explanation is more consistent with what we observe, since low-molecular weight contaminants were extremely small and scattered in distribution relative to the major fraction. A contaminating fragment should be expected to be at least within the same size range as the major fraction. Confirmation of the rearrangement hypothesis would have required extensive sequencing, however, and this was not performed.

Genome Size Analysis. Flow cytometry performed very well for genome size estimation of our bacterial strain. We were interested in a rapid technique to provide size estimates with good accuracy, but not necessarily high precision. The estimated genome size would indicate the number of colonies required to generate a high coverage library (38, 169), but only two to three significant figures would be necessary for the estimate, so a rapid, simple analysis was more important than precision. There are a variety of techniques for the estimation of genome size including densitometry (74) and the preferred method of PFGE analysis (116), but these tend to be time consuming and somewhat unreliable. While other researchers might criticize flow cytometry for its lack of precision, for our purposes its major shortcomings are instrument expense and the expertise required, so while it may not be available to all researchers it is certainly the preferred choice for our purposes if facilities and expertise are available.

Conclusions. While the construction of this particular library was eventually successful and the analysis went well, it must be concluded that this approach is not generally viable for the search for novel natural products without significant improvements. The methods were neither general across species nor were they simple enough for widespread application. In addition, it appears likely from the insert size analysis that significant cloning bias may be associated with the technique, and screening of the library by several means for pathways of interest (data not

shown) was unsuccessful. BAC library construction has been very useful for the investigation of mammalian and plant genomes and, in general, large insert library construction has been useful for the investigation of many biosynthetic pathways in the past. But these types of experiments usually employ λ-based libraries, which are easier to construct and manipulate, and these investigations are done on organisms which are already known to harbor interesting pathways, immediately justifying their effort and expense.

CHAPTER IV

CONSTRUCTION OF FOSMID LIBRARIES OF GENOMIC AND METAGENOMIC DNA

INTRODUCTION

Construction of λ-bacteriophage based libraries has been a core technique for molecular biologists for over 30 years (169). A wide variety of vector systems have been constructed based on this system which primarily differ in the amount of the original viral genome and function still contained in the DNA fraction of the system. The viral head which carries the DNA construct is limited by volume to contain no more than 50 kb of DNA, so the less viral sequence contained in the DNA the larger the foreign insert can be. Currently, the most popular systems for large insert libraries contain only the "cohesive ends," or *cos* sites which are recognized by the viral proteins for packaging, and are called cosmid libraries. Replication and maintenance functions within the bacterium are performed by standard origin and selectable marker sequences and viral protein extracts are generated in a separate step. Modifying the cosmid with an F-factor origin of replication lowers the copy number of the generated plasmid, increasing its stability (99). A cosmid library utilizing such an origin is called a fosmid library. Because the minimal DNA sequence to encode these functions requires ~ 7-10 kb, these vectors are capable of carrying approximately 40 kb of foreign DNA.

The viral machinery for packaging and transfection is highly efficient and, because the process does not require manipulation of DNA over 50 kb, the danger of degradation due to shearing is very small and the sample can be handled roughly and purified more easily and thoroughly. While the fosmid cannot compete with the BAC library or the yeast artifical chromosome (YAC) library in terms of the size of the insert it can propagate, these vectors are no match for the fosmid in terms of flexibility, robustness, and ease of application. While our previous efforts to produce BAC libraries failed on several of our environmental samples, fosmid libraries succeeded. Here we describe their construction.

MATERIALS AND METHODS

Isolation of Environmental DNA. Mesocosm samples were obtained from Dr. Richard Long, Department of Oceanography. Briefly, a 200 liter carboy of sterile synthetic seawater was inoculated with a 30 liter sample of ocean water obtained from the Gulf of Mexico which had been filtered through a 163 μm Nytex filter to remove macroscopic organisms. The carboy was

incubated for 19 days at 24°C under 14:10 light:dark cycle to simulate ambient temperature and light conditions. The carboy was then cooled to 4°C in the dark overnight to cause deposition of the bulk of the culture to the bottom. The supernatant salt water was pumped off the surface and the residual amount was transferred to a 25 liter carboy. The deposition process was repeated until the volume was sufficient for centrifugation in a tabletop centrifuge. A 100 ml sample was centrifuged for 5 min at 2750 rpm. The pellet was washed twice with marine resuspension buffer (5% w/v instant ocean sea salt, 5% w/v sorbitol, 100 mM EDTA, pH = 8) and flash frozen in liquid nitrogen in droplets. The droplets were ground with a mortar and pestle and stored at -80°C.

Biofilm was obtained from a field collection by Dr. Dan Thornton, Department of Oceanography, Texas A&M Univeristy, in the Gulf of Mexico. The biofilm was initially washed with 19 g/L instant ocean seawater to remove dirt and debris, then cut with a razor blade into small pieces. The pieces were flash frozen on liquid nitrogen, then ground with a mortar and pestle.

Isolation of genomic DNA from both samples was identical. Two cubic centimeters of the ground material was suspended in 5 ml of TE buffer. To this was added an additional 5 ml of SDS lysis buffer (1% SDS in .5 M EDTA, pH = 9), prewarmed to 50°C. 1 mg of Proteinase K was added and mixed gently and the solution was incubated at 50°C for 24 hours. After lysis, the mesocosm solution cleared noticeably, but the biofilm sample appeared unchanged. The volume was tripled with water and an equal volume of chloroform was added. The phases were mixed gently but thoroughly, then the sample was centrifuged at 2750 rpm for 10 min. The upper aqueous phase was transferred to a clean tube. To this was added 1/10 volume of 2 M sodium acetate, and 5 ml was transferred to a new clean tube. 15 ml of ethanol was gently overlaid on top of the aqueous phase and the DNA was spooled out at the interface. The biofilm sample did not produce enough DNA for spooling, so the entire sample was mixed with 2 volumes of ethanol and centrifuged.

The DNA pellet was rinsed in 70% ethanol, dried under a stream of nitrogen, then resuspended in 1 ml of TE buffer (10 mM tris, 1 mM EDTA, pH=7.5). The solution was treated with 5 µl of 1 mg/ml RNAse A for 4 hours at 37°C, then extracted with 1:1 phenol:chloroform. 100 µl of 2 M sodium acetate was added and the DNA precipitated with 1 volume of isopropanol. The pellet was rinsed with 70% ethanol, dried under a stream of nitrogen, then resuspended in 50 µl of TE buffer.

Isolation of *Pseudoalteromonas* DNA. An unidentified strain of *Pseudoalteromonas*, Surf1, which was collected by Dr. Coran Watanabe at Surfside Beach, TX, was streaked onto marine agar plates as described for the *Chromohalobacter* BAC library and grown at 30°C for 48 hours. Cells were harvested by scraping from the plate surface with a cell scraper and transferred to a clean tube. The cell mass was suspended in marine resuspension buffer (100 mM EDTA, 5% Sorbitol, 5% Instant Ocean) and collected by centrifugation. The washed cells were resuspended in an equal volume of TE and embedded in agarose, washed and lysed as described for the *Chromohalobacter* BAC library.

Agarose embedded DNA was embedded in the gel of a Bio-Rad CHEF DR II PFGE apparatus described for the *Chromohalobacter* BAC library. The gel was run at 14°C with settings of 6 V/cm, a linear ramp from 2 to 3.5 sec switch times for 10 hours. NEB MidRange II PFG markers were used as a sizing standard and placed in flanking lanes. After electrophoresis, the flanking ladders were cut out and stained by immersion in 100 ml of 100 μg/l ethidium bromide and destained in water. The regions of the gel corresponding to 50-65 kb were marked with a razor and realigned with the remainder of the gel. The agarose embedded *Pseudoalteromonas* DNA sample corresponding to this size range was cut out and isolated with a Qiagen gel extraction kit with the following modifications. Centrifugation was done at 4000 rpm and the purified DNA incubated at 37°C for 15 minutes before elution to prevent additional shearing. The remainder of library construction was the same as for the environmental samples.

Library Construction. Each DNA sample was sheared to the proper size range by repeated passage through a 200 μl micropipette tip. A 2 μl sample was quantified on a gel and the sample diluted to give 200 ng/μl by comparison of fluorescence with the ladder standard. 34 μl of this solution was end repaired using the End-It kit by Epicentre.

After incubation, 10 μl of glycerol and 5 μl of loading dye was added to the reaction. The reaction was gently mixed, then loaded into two wells of a Bio-Rad CHEF DR II PFGE apparatus. The gel was run at 14°C with settings of 6 V/cm, a linear ramp from 2 to 3.5 sec switch times for 10 hours. NEB MidRange II PFG markers were used as a sizing standard and placed in flanking lanes. After electrophoresis, the flanking ladders were cut out and stained by immersion in 100 ml of 100 μg/l ethidium bromide and destained in water. The regions of the gel corresponding to 50-65 kb were marked with a razor and realigned with the remainder of the gel. The agarose embedded environmental DNA sample corresponding to this size range was cut out.

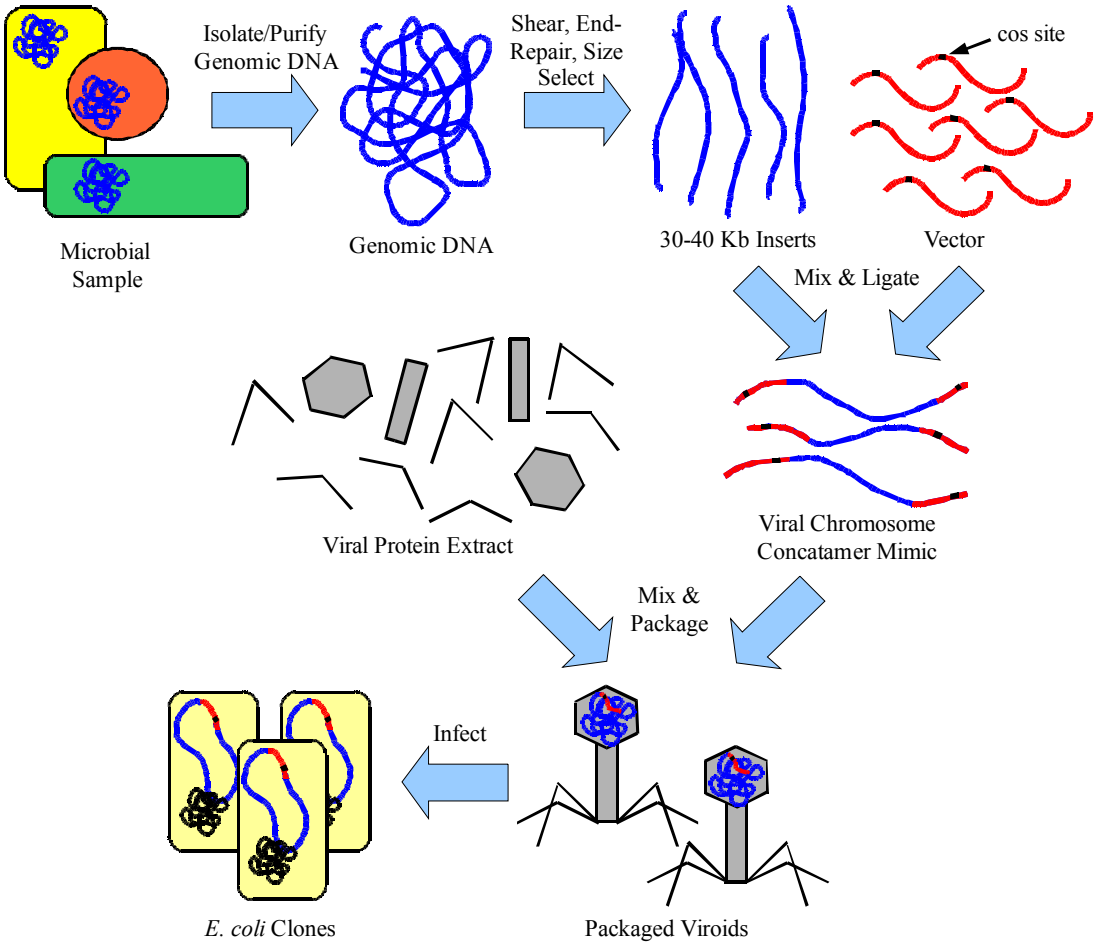


FIG. 20. Schematic representation of fosmid library construction.

The sample of agarose was cut down to fit into 10 mm 6000 molecular weight cut-off dialysis tubing. Minimal TE was added to displace air and the ends were sealed with clips. The DNA was electroeluted on the Bio-Rad CHEF DR II apparatus at 12°C with settings of 6 V/cm and a fixed switch time of 30 sec for 6 hours. The electroeluted buffer was collected and the DNA precipitated with 1/10 volume of sodium acetate and one volume of isopropanol. After resuspension in 10 µl of TE buffer, the DNA was quantified to give a concentration of 50 ng/µl. Ligation of the DNA to the vector, packaging into viral heads, and transfection into *E. coli* was performed according to the Epicentre Copy Control Fosmid Library Production Kit (Fig. 20).

Arraying and Analysis of Library Inserts. After transformation of the library into *E. coli* EPI300-T1^{R} provided with the kit, 25 randomly selected clones were picked and the plasmid

DNA isolated by alkaline lysis according to standard procedures (7). The plasmids were digested with NotI and analyzed by PFGE (Fig. 21).

Colonies were picked and arrayed on 384-well plates. In total, nine plates were arrayed for each environmental library, giving a total of 3456 members in each library. Three plates were arrayed for the *Pseudoalteromonas* library.

RESULTS

Library Titer and Insert Size Analysis. The titer of the phage packaging reactions was low for the environmental samples. A typical ligation and packaging reaction gave a titer of 1 cfu (colony forming unit) per microliter, at least three orders of magnitude lower than what is typically seen for a library preparation of this type. Three separate preparations were required to produce the requisite number of colonies. Insert size was well within the acceptable range, however. A few small insert and empty plasmids were observed in the sample (Fig. 21).

The *Pseudoalteromonas* library performed much better. Titer was in the range of 10^5 cfu per microliter, and a one-hundred-fold dilution was necessary for plating. Insert sizes were comparable with the environmental libraries.

DISCUSSION

Library Construction. Manipulation of environmental DNA proved to be highly problematic. Solutions of DNA were off-colored regardless of the purification technique used.

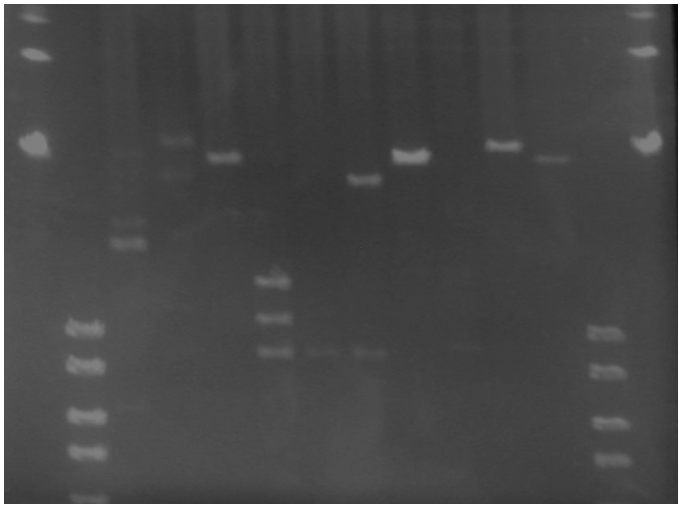


FIG. 21. PFGE sizing of fosmid inserts. Lanes 1, 14 - NEB MidRange II PFG Markers. Lanes 2, 13 – NEB 1 kb Ladder. Lanes 3-12 – NotI digests of mesocosm fosmid clones. Most contain inserts ranging from 30 to 40 kb, while lanes 7 and 10 show empty plasmids.

The *Pseudoalteromonas* Surf1 library performed exceptionally well, and was the only one for which DNA was subjected to the Qiagen gel-purification procedure prior to enzymatic modification. This is indicative that the major difficulty with the manipulation of these types of DNA is one of purity, and that these impurities interfere with enzymatic modification of DNA. Cellular debris, media, and environmental contaminants can contain enzyme inhibitors, and it is a common occurrence that DNA samples will be resistant to enzymatic modification due to the presence of these compounds (67, 169). Working with mixed cultures, such as environmental samples, only increases the complexity of the mixture and the probability that inhibitory substances will be present. These DNA samples were also used for the attempted construction of BAC libraries (data not shown), but all attempts failed because in every case either the DNA could not be digested with restriction enzymes, or once fragmented it failed to ligate to the vector. The environmental samples came with the additional problem of low DNA yields and extensive degradation when lysed while embedded in agarose.

DNA purification is a routine problem reported with the construction of large insert libraries, and several groups have proposed unusual purification schemes (87, 166, 189). Purification for these samples is a particularly vexing problem due to the properties of the DNA in question. Although DNA is reasonably resilient to chemical degradation such as oxidation and hydrolysis, it does have weaknesses. Because it is a high molecular weight polymer, it tends to shear if subjected to mechanical stresses, and because it is a common biological currency it is susceptible to degradation by many enzymes and microbes present in the environment (169). High molecular weight DNA is also susceptible to damage by UV light and intercalating dyes which are frequently used for routine DNA detection in the lab (152, 169). Unfortunately, no solution has proven particularly effective for a wide range of samples, so the process remains unstandardized and often tedious and time consuming.

On the positive side, library construction was successful and insert sizes were of good quality. The two environmental libraries contain on the order of 138 Mb of DNA, or about 30 times the amount of sequence present in *E. coli*. This is a substantial amount of sequence, and should be sufficient to contain at least one pathway of interest. The surf1 library contains 46 Mb, or about 10 fold coverage of this organism's closest relatives (104). This is more than adequate coverage for cloning a sequence of interest.

CHAPTER V

ANALYSIS AND SIMULATION OF THE PROBABILITY OF CLONING INTACT BIOSYNTHETIC PATHWAYS IN LARGE INSERT GENOMIC LIBRARIES

INTRODUCTION

If large insert genomic libraries of cultured and environmental samples are to be used for activity based screening of natural product pathways, several important conditions must be met. The most obvious is that if a pathway is in fact cloned, it must also be expressed. Expression requires that several biochemical hurdles be overcome (118, 190). The transcriptional machinery of the heterologous host must be able to recognize and initiate transcription from the promoter sequences found on the alien DNA fragment, and once the mRNA is produced, it must be successfully translated into functional protein. There are several pitfalls on the pathway from mRNA to functional protein. Translation initiation efficiency may be low due to secondary structure, poor recognition of ribosomal binding sites, or interference with translation by host regulatory mechanisms. The host's translational machinery may not be equipped to handle the codon usage of the message. Even if the protein is successfully produced, achieving the proper folding for activity can be a problem in a different environment, particularly if chaperonins or foldases are involved, and if the proper cofactors are not present or the modification enzymes lacking, the protein will not be able to function. Proteases and RNAses can also degrade the protein and mRNA, respectively, and reduce the level of protein present. Assuming the protein overcomes all these hurdles, adequate substrates may not be present in the host's metabolic repertoire for function.

These difficulties aside, a slightly more obvious hurdle is that the pathway must, in fact, be cloned in its entirety to function properly. In the past, this fact has often been overlooked in cloning experiments, and with good reason. Upon production of their first "clone bank," Clark and Carbon considered the problem of cloning probabilities and generated the formula that has found the most widespread use (38, 169). Their analysis was simple and straightforward, and as a result was highly applicable to many situations. One of their implicit assumptions was that the "region of interest," whether it is a gene, pathway, or in the case of a sequence library, a single base pair, has a size which is negligible in comparison to the size of the cloned insert. This is a good approximation for most circumstances, such as looking for genes from a BAC or cosmid

library, or calculating appropriate coverage for a small insert sequencing library. To take into account the deviation from their ideal as a result of this assumption was unnecessary, since the difference is negligible and biases and errors due to other assumptions, such as the true randomness of the fragmentation method, would far outweigh this effect (182, 183). Because the region of interest is so small, the probability that one end of the insert or the other would actually fall within the region is effectively nearly zero, so the region will either be cloned intact or not at all by any particular cloning event. To try to take this probability into account is analogous to including the probability of a coin landing on its edge when calculating the probability of it landing heads or tails.

However, the cloning of intact pathways in large insert libraries warrants a closer inspection of this problem. Pathways are frequently nearly as large as BAC inserts themselves (178, 219), so the probability of the cloning event capturing only a part of the pathway becomes significant. The coin has effectively become an elongated cylinder, and the probability of landing on the "edge" cannot be dismissed so casually. It is clear that the library must be enlarged to maintain a high probability of cloning pathways intact, but at first glance it is not apparent just how much larger the library must be.

One approach would simply be to expand the library to the extent possible given the available resources. This approach might be acceptable as a first time application of the technology to see if it is feasible, but in the long term this is an unacceptable strategy. Many of the potentially useful screening methods which might be employed for these experiments are resource intensive, such as the activity screening of organic extracts or the so-called hyphenated methods like LC-MS. Screening an excessively large library would result in an excessively large number of samples for screening and a waste of resources at the very least. It could very well also lead to confusion due to an excessively large number of positive clones, which could foil the popular split-and-pool approaches specifically designed to handle such large libraries and minimize resource use.

For the sake of resource and cost conservation, as well as establishing a mathematically sound description of library coverage for large cloning targets, we describe a mathematical analysis of large-target library coverage and the derivation of a useful formula for calculating cloning probabilities. We also test the accuracy of this formula with a spreadsheet driven computer simulation.

MATERIALS AND METHODS

Instrumentation and General Methods. All calculations were performed on a Dell Inspiron 6400 laptop personal computer running on a Windows Vista operating system equipped with Excel 2003 and Microsoft VisualBasic for Applications. Simulations were done using the random number generator which comes with the Excel program.

Description of Calculation. Library parameters are input via user interface. For each library simulation, the target is assigned a random position by choosing a random number between 1 and the input size of the genome. The ends of the target sequence are assigned as this random number and a second number equal to the first number plus the input size of the target. Clones are generated by choosing a random number between 1 and the size of the genome, and the ends of the clone are assigned similarly. Each clone is scored either positive or negative. A positive clone must have its lower end assigned to a number less than the target's lower number, and a higher end assigned to a number larger than the target's higher number. Positive clones from each simulated library are counted, and the library simulation is repeated. The calculation and simulation were performed by Scott Angell and Jeff Janes (Genomics Institute of the Novartis Research Foundation).

RESULTS

Derivation of Formula. The Clarke and Carbon formula can be derived by using the following analysis. For a genome of size G, the number of randomly selected subclones N which should be collected to ensure a probability P of containing a sequence of interest can be found with equation 1,

(1) $$N = \frac{\ln(1-P)}{\ln(1-f)}$$

where f represents the fraction of G contained in an individual subclone of length I, or f = I/G. This equation is a simple rearrangement of equation 2,

(2) $$(1-P) = (1-f)^N$$

which can be interpreted as stating that the probability of not cloning a fragment, (1-P), is equal to the product of the probabilities that each individual clone will not contain the fragment, (1-f). N is the total number of clones and is therefore the exponent. Since f represents the fraction of G

contained on the interval spanned by I, it is also the probability that a sequence of interest will be contained in I, and (1-f) is the probability that it will not.

This equation only holds under the approximation that the sequence of interest is negligibly small in comparison to the insert size of the subclones. In the event that the desired sequence is in the same size range as the insert, the probability that the sequence will be cloned intact is considerably less than the probability f. To correct for this scenario, we have inserted the variable c in front of f, where c is an expression for the probability that an insert which contains a fixed, negligibly small part of the sequence of interest will contain all of the sequence of interest. The probability that a randomly selected clone will contain the entire pathway is therefore cf, and equation 1 becomes

(3) $$N = \frac{\ln(1-P)}{\ln(1-cf)}$$

In order for an insert to contain the entire sequence of interest, its sequence must obviously begin before the sequence of interest begins and end after the sequence of interest ends. In the case of a negligibly small sequence, the interval of G contained in I may begin at any of the positions I base pairs upstream of the sequence if interest. For a larger sequence, we can consider the beginning point to be constrained to lie on a smaller interval which is I – B base pairs long, where B is the length of the sequence of interest, in our case a biosynthetic pathway. This will be true as long as the genomic insert is actually longer than the sequence desired, and c can be expressed by equation 4, within these limits.

(4) $$c = \frac{I-B+1}{I}$$

Researchers frequently neglect exact descriptions of libraries in terms of insert size and number of clones in favor of the more succinct description of redundancy, R, where R = NI/G or Nf. This approximation is justified by using the Poisson distribution to describe coverage and probabilities instead of equations 1 or 2. The Poisson description is actually the limit of an infinite binomial, and in this case the term $(1-f)^N$ is the binomial being approximated. So long as N is reasonably large, the limit function provides an adequate approximation of this term, though for smaller values of N the description will break down.

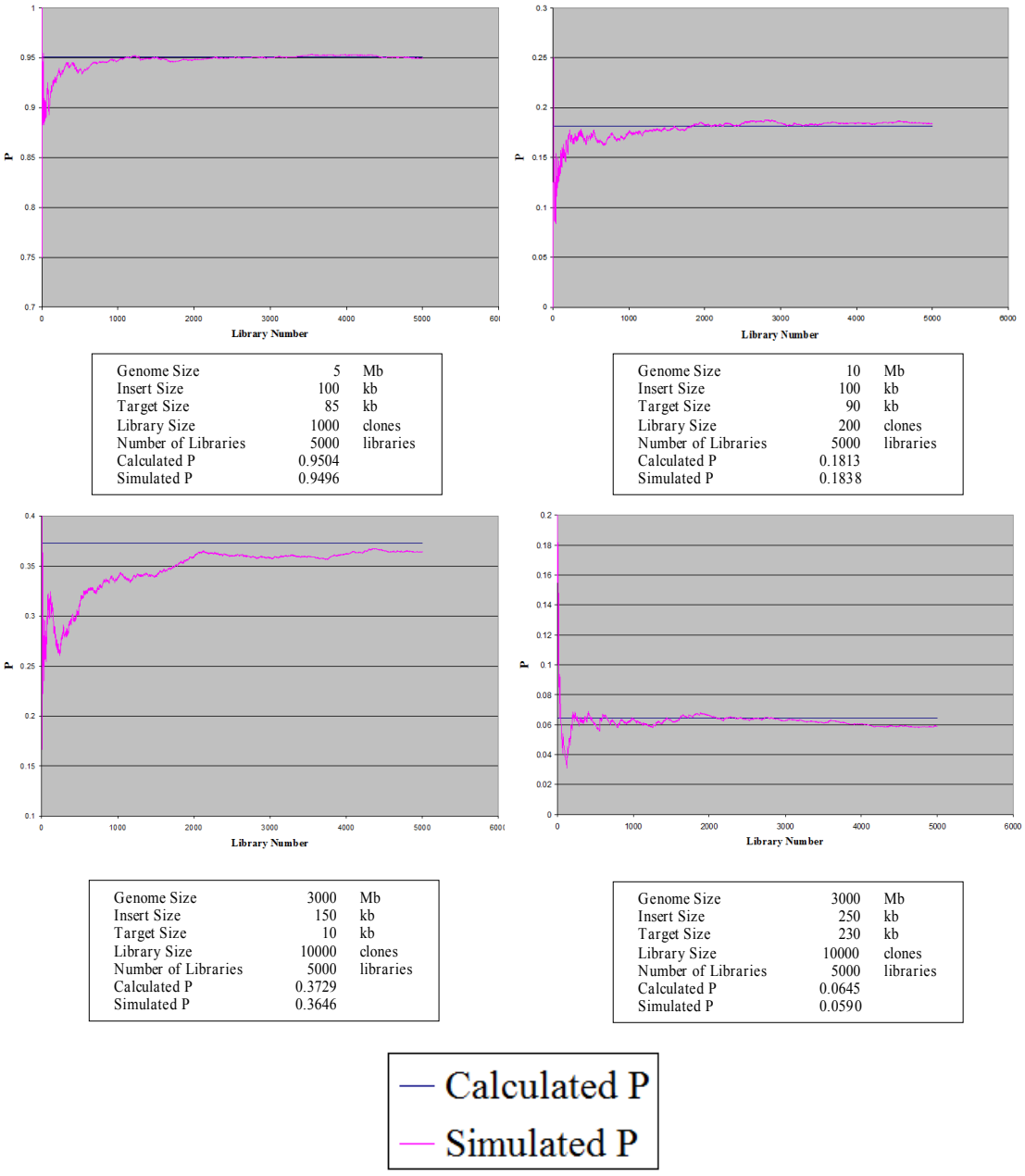


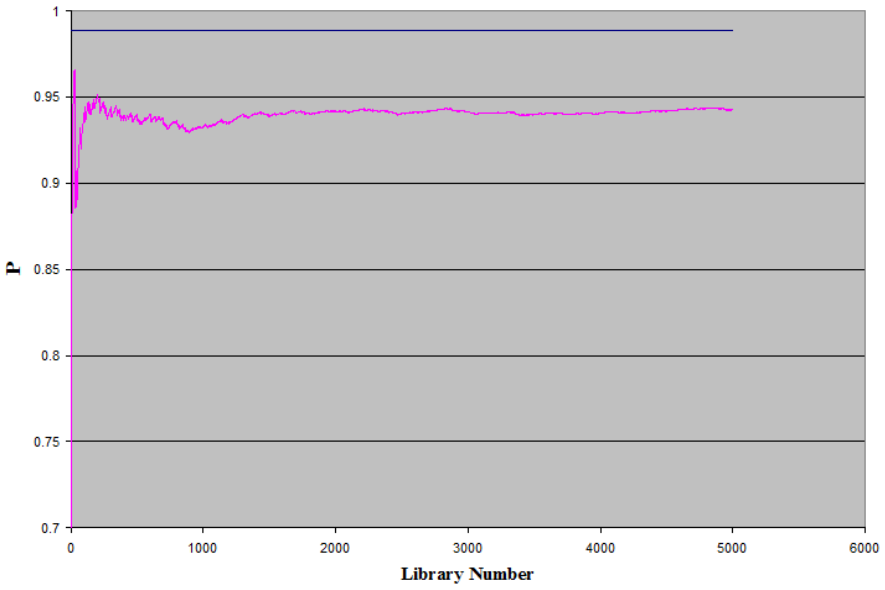
FIG. 22. Computer simulations of cloning probability.

Using the Poisson description, the Clarke and Carbon equation becomes

(5) $$P = 1 - e^{-R}$$

which gives probabilities very close to those generated by equations 1 and 2 over a wide range of insert size and clone numbers. The calculation was generated by Scott Angell and Jeff Janes (98).

Computer Simulation. A cumulative average was used to demonstrate visually the convergence of the simulation towards the predicted value (Fig. 22). For all realistic situations tested, the simulation converged on the prediction generated by our equation, although the simulation could be thrown off by using insert sizes that were on the order of the size of the genome itself (Fig. 23). Such a "library" would contain more than a tenth of the genome in each clone and contain very few members, and therefore would not be a candidate for the application



Genome Size	1	Mb
Insert Size	250	kb
Target Size	50	kb
Library Size	20	clones
Number of Libraries	5000	libraries
Calculated P	0.9885	
Simulated P	0.9426	

Calculated P
Simulated P

FIG. 23. Erroneous simulation convergence at high insert-to-genome ratio, low sampling.

of random sampling statistics. Other extreme situations, such as using inserts very close to target sizes, very small target sizes relative to inserts, and very small inserts relative to genome size, did not give erroneous results, although the latter situation did require more simulations to converge. We therefore conclude that the computer simulations confirmed the predictive value of our model, at least for realistic values. The computer simulation was performed by Scott Angell and Jeff Janes (98).

DISCUSSION

Formula Derivation. In order to make metagenomic searches for natural product biosynthetic pathways widely applicable and useful, it must become a more efficient methodology. Displacement of the current technology will only occur when it is clear that the new technology provides more output per dollar invested, and despite all its problems of strain and compound reisolation, the current technology is much cheaper and will continue to outperform so long as metagenomic technology remains inefficient and expensive. Thus it is clearly important to economize the technique, and accurate mathematical descriptions of the library size necessary for adequate coverage are a valuable pursuit so long as libraries continue to be the method of choice for bioprospecting.

Our derivation of the formula is simple and straightforward. The Clarke and Carbon formula was only slightly modified with the coefficient c, which is easily justified. In order for a particular insert to span a region of interest, it must begin before the region starts (e.g. begin 5'

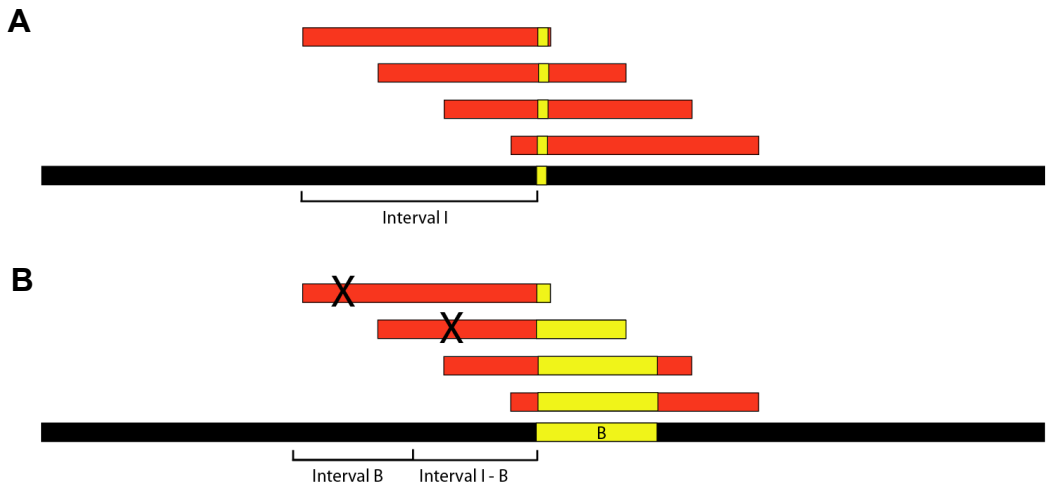


FIG. 24. Derivation of the coefficient c = (I-B+1)/I. (a) Depiction of cloning events which capture a target of negligible size. All clones whose insert begins on the interval I upstream of the target will clone the target. (b) Depiction of cloning events which capture a target of non-negligible size. Successful cloning events must begin on the interval I-B bp upstream of the target sequence.

to the pathway), and end after the region stops (e.g. end 3' to the pathway). For a negligibly small pathway, an insert of size I may begin at any interval within I bp of the beginning of the region to contain the region. For a region of non-negligible size B, the pathway must begin close enough to the beginning of the pathway to contain the entire pathway; therefore the interval is now reduced to size I-B+1 bp in size. The fraction of those clones which will contain the entirety of a pathway of size B relative to the total number which would contain the negligibly small region of the Clarke-Carbon model is (I-B+1)/I (FIG. 24).

It should be noted that many groups performed such analyses long before we did. Much in depth analysis has been done on the application of statistical sampling to the coverage of genomes, particularly with respect to sequencing projects (102, 165, 182, 183, 209), and Veiga and co-workers performed virtually the same analysis with respect to individual genes (200).

The strength of any model is not only in its predictive value, but also its ability to convey a meaningful description of the system it describes. Our model is only a slight modification of the existing Clarke-Carbon formula, and makes the modification with a coefficient in a way that is both conceptually meaningful and portable in that it is directly applicable to the Poisson model and easily incorporated into the logic and vernacular used to describe library coverage. For example, most descriptions of cloning probability focus on library coverage or redundancy. One can describe the redundancy of a library for B bp coverage as $R_B = c_B Nf$, where c_B is the coefficient calculated for the pathway and insert size in question, and R_B is the associated redundancy for this size genetic element. This R_B factor can be directly used in the Poisson model of equation 5 to calculate accurate probabilities. A library can then be described in terms of its coverage for different purposes, allowing for a universal descriptive system for comparison, such as 10-fold redundant sequencing coverage, but only 4-fold redundancy for 85 kb pathway coverage. Better comparisons can then be made between libraries for different applications.

From a conceptual point of view, this redundancy description is more useful than any of the equations because it makes sense to think of an expansion of the redundancy necessary to compensate for the loss of probability due to the cloning of incomplete pathways. If we are only one-third as likely to clone an intact pathway as we are a negligibly small sequence, we need to collect three times as many clones to have the same probability of finding this larger sequence. Such a conclusion is both mathematically sound and readily apparent from our description.

Computer Simulation. The computer simulation was designed to mimic cloning a target from a circular chromosome, since this is the form that most bacterial genomes take. This

is accomplished by allowing both the target and clones to "run off the end" of the genome, presumably back to the beginning. The presence of multiple circular chromosomes, as are present in some bacterial species, should not change this analysis, and although our analysis could not be explicitly applied to linear chromosomes, such as for *Streptomyces* and most eukaryotes, the correction would be expected to be small since end effects would only apply to a small proportion of targets and clones. Again, such a correction would add little in the way of predictive value while undermining the model's simplicity.

Overall, this analysis shows that the cloning of intact pathways is feasible for insert sizes which are at least a third larger than the size of the desired target. Beyond this, as the target approaches the size of the insert, the number of clones required to maintain the same level of probability of cloning increases rapidly, making the required libraries unworkably large. From a practical standpoint, the model becomes unsuitable for libraries produced by partial restriction digests as one approaches this extreme, since the probability of suitable restriction sites being present this close to the ends of the target will become small enough to become non-negligible. For these types of libraries, the inserts should be safely larger than the desired pathway, considering that the expected distance between restriction sites is approximated by 4^n, where n is the number of base pairs in the recognition sequence. Inserts should be some multiple of this value larger than the desired target to ensure a reasonable probability of adequately positioned restriction sites.

CHAPTER VI

CONSTRUCTION OF A DITERPENE COMPLEMENTATION SYSTEM IN *STREPTOMYCES COELICOLOR* A3(2)

INTRODUCTION

The *Streptomycetes* and their relatives in the bacterial order *Actinomycetales* are renowned for their capacity to produce antibiotics and other secondary metabolites (13). The first antibiotic isolated from a streptomycete was streptothricin, isolated in 1942, and for more than a half-century since then they have proven to be a copious source of drugs and drug leads (83). Although their share of the total number of new secondary metabolites discovered has declined in recent years, these organisms retain a central role in bioactive natural products research (13).

In 2005, the Ueda group reported the ability of *Streptomyces coelicolor* A3(2) to generate carotenoids in response to exposure to low wavelength light (Fig. 25) (194). The carotenes are a family of colored organic compounds produced by a variety of organisms which have anti-oxidant properties as a result of their possession of highly conjugated π-electron systems. These compounds can help reduce photo-oxidative damage that results when chemically sensitive compounds undergo photo-excitation to unstable, highly reactive excited states. The production of carotenoids had not been previously observed due to the constitutive production of two powerful chromophores by this *Streptomyces* strain, actinorhodin and prodigiosin. These two intensly colored compounds obscured the orange color resulting from carotene production. Ueda and co-workers happened to isolate a plasmid sequence which suppresses production of the interfering chromophores, allowing them to study the regulation of carotene production.

This phenomenon is of interest because the detection of carotene production is very simple and one of the intermediates in this pathway is crucial to the production of a large family of secondary metabolites, the diterpenes (Fig. 25A). Geranylgeranyl pyrophosphate (GGPP) is produced by the head-to-tail condensation of farnesyl pyrophosphate (FPP) and isopentenyl pyrophosphate (IPP), a reaction catalyzed by geranylgeranyl pyrophosphate synthase (GGPPS). Two molecules of GGPP can then undergo head-to-head condensation to produce phytoene, which is desaturated to varying degrees and further modified to generate the various members of

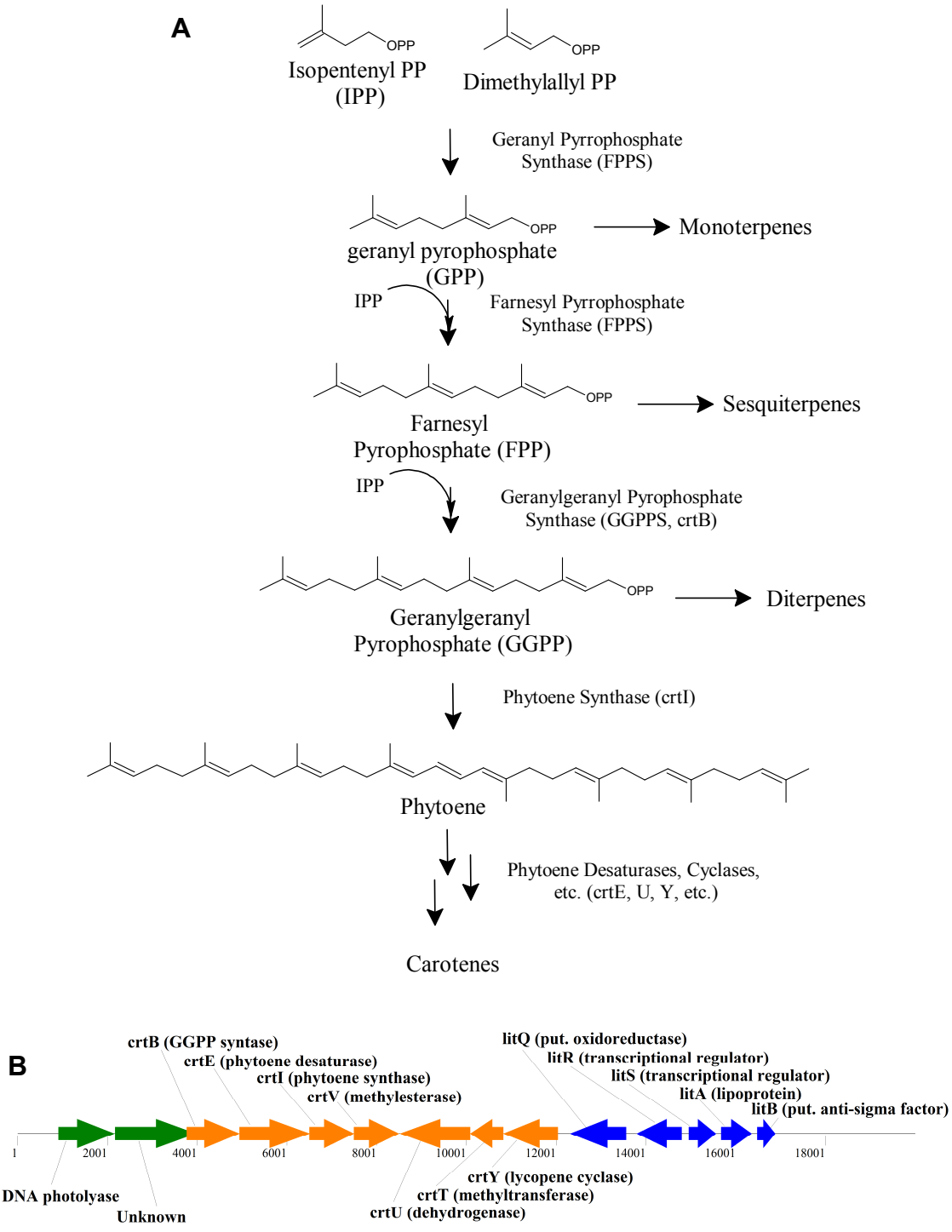


FIG. 25. Carotene biosynthesis in *Streptomyces coelicolor*. (a) Enzymes and chemical intermediates of carotene biosynthesis and branch points for terpene production. (b) Map of the light-inducible carotene biosynthetic pathway in *Streptomyces coelicolor* A3(2).

the carotene family of natural products, or a single molecule can undergo various cyclizations and other modifications to generate a wide variety of diterpene natural products, such as taxol.

We reasoned that this carotene pathway might be exploited for the purpose of constructing a useful visual screen for the presence of GGPP. If the GGPPS gene were knocked out of a strain, it might be possible to detect complementation of the knockout by heterologous DNA hosted in a large-insert plasmid. This would allow easy visual inspection of a BAC or λ-bacteriophage based library hosted in the knockout strain for the presence of GGPPS activity. Because the genes for a biosynthetic pathway in bacteria are usually clustered, complementation of the knockout would indicate a high probability that a diterpene synthase pathway was present on a plasmid which displayed complementation. This would help to simplify library screening, which has been a major hurdle for the isolation of secondary metabolite pathways from metagenomic libraries thus far. It would also make use of a *Streptomyces* host, which could potentially be a better host for heterologous expression than *E. coli* due to its larger biosynthetic repertoire, particularly if the source DNA for the library was enriched for *Actinomycete* DNA.

MATERIALS AND METHODS

Strains and Plasmids. *Streptomyces coelicolor* CH999, *E. coli* strain ET 12567, and plasmids pIJ86 and pKC1139 were obtained from Dr. Rongfeng Li of the Townsend Group at Johns Hopkins University. Routine DNA manipulations were performed in *E. coli* DH10B.

Construction of the Knockout Plasmid. A 1.4 kb region of the *Streptomyces coelicolor* CH999 genome containing the crtB gene (as it is designated in the NCBI genomic database for the organism; most other sources refer to it crtE) was amplified by PCR using the primers SCcrtBH3KOfor (5'-ATA AGC TTG TCA GTG CAT CCA CAA CAG G-3') and SCcrtBH3KOrev (5'-GGA AGC TTA CCG GTG TCC AGC TCG TA-3'), which install HindIII sites for restriction digest, from the genomic DNA of *Streptomyces coelicolor* CH999, which was obtained by lysis of protoplasts (Fig. 26). Protoplasts were prepared according to Hopwood (96). Briefly, a 3 cm^2 agarose slab from a sporulated culture lawn cultivated on an MS plate (Mannitol Soy, 20 g soy flour, 20 g mannitol, 20 g Bacto Agar, diluted to 1 liter) was inoculated into 25 ml liquid R2YE (103 g sucrose, 2 g Bacto Triptone, 4 g Bacto Peptone, 4 g Yeast Extract, .25 g K_2SO_4, 4.7 g $MgCl_2$, 10 g glucose, .5 g Casamino Acids, 50 mg KH_2PO_4, 3 g $CaCl_2 \cdot 2\ H_2O$, 3 g L-proline, 2 ml Trace Elements Solution (40 mg/l $ZnCl_2$, 200 mg/l $FeCl_3 \cdot 6\ H_2O$, 10 mg/l $CuCl_2 \cdot 2\ H_2O$, 10 mg/l $MnCl_2 \cdot 4\ H_2O$, 10 mg/l $Na_2B_4O_7 \cdot 10\ H_2O$, 10 mg/l $(NH_4)_6Mo_7O_{24} \cdot 4\ H_2O$), 27 ml 1M Tris • HCl, pH = 7, diluted to 1 liter) media containing 1%

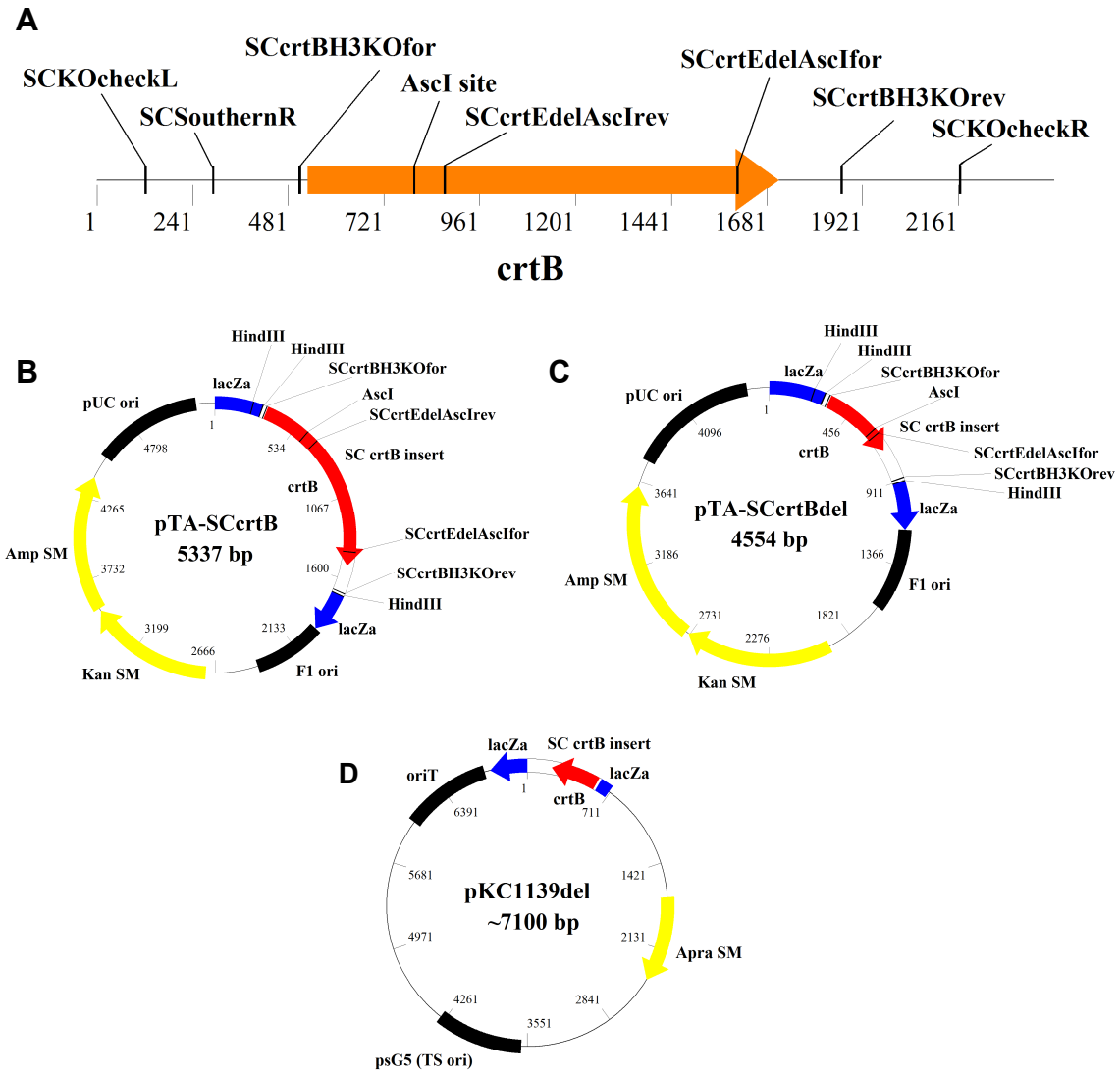


FIG. 26. Construction of the knockout plasmid pKC1139del. (a) Map of the crtB region, including locations of priming sites and the AscI site used for deletion. (b) Plasmid map of the TA-clone of the crtB region amplified from the genomic DNA of *Streptomyces coelicolor*. (c) Plasmid map of the TA-clone after inverse PCR deletion of an internal region of the crtB gene. (d) Plasmid map of pKC1139del.

glycine in a 250 ml Erlenmeyer flask equipped with two-hundred 3 mm glass beads. The culture was grown at 30°C, 250 rpm for 48 hours, then the mycelial pellet was harvested by centrifugation at 3750 rpm for 5 min. The pellet was washed twice in 20 ml of 10% sucrose, then resuspended in 7 ml of protoplast buffer (103 g sucrose, .25 g K_2SO_4, 2 g $MgCl_2$• 6 H_2O, 5 mg KH_2PO_4, .5 ml of .5 M $CaCl_2$, 2 ml Trace Elements Solution, 50 ml 1M Tris • HCl, pH = 7, diluted to 1 liter and sterilized) containing 2 mg/ml lysozyme. After incubation for 45 minutes at

37°C with occasional agitation, the mycelial clumps had dissipated, leaving a fine suspension. The suspension was filtered through sterile cotton and the protoplasts harvested by centrifugation at 3000 rpm for 10 min. Protoplasts were lysed by resuspension in 1% SDS, or resuspended in protoplast buffer lacking lysozyme and stored in 100 µl aliquots by slow freezing to -80°C. The PCR reaction contained 40 µl of water, 2.5 µl of DMSO, 5 µl ThermoPol Reaction Buffer (10X), .5 µl dNTP's (25 mM each), .5 µl Taq DNA Polymerase (5 U/µl, New England Biolabs, Ipsich, MA), 2 µl each primer (100 pmol/µl) and 1 µl of genomic DNA (50 ng/ul). Thermal cycling was performed with a PTC-200 thermal cycler (MJ Research, Waltham, MA) using an initial 2 min denaturing step at 96°C, followed by 20 cycles of touchdown conditions of 96°C for 30 s, 30 s at the annealing temperature which began at 60°C and decreased .5°C at each cycle, and 1.5 min at 68°C. This was followed by 20 standard cycles of 96°C for 30 s, 55°C for 30 s, and 68°C for 1.5 min. After cycling, the reaction was held at 68°C for 7 min, then stored at 4°C until the time of analysis. The reaction was analyzed by gel electrophoresis, the band excised, and the DNA isolated with the Qiagen gel extraction kit. The band was TA cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA).

The in-frame internal deletion was generated through an inverse PCR reaction performed on the cloned fragment. The primers used for this reaction were SCcrtEdelAscIfor (5'-ATC GGG CGC GGC CAC GGA CAT GAG CCT CCG CGC AC-3'), which installs an AscI site, and SCcrtEdelAscIrev (5'-GTC GTC GTG GAC CAG AGC -3'), which amplifies just upstream of a naturally occurring AscI site in the gene. This reaction contained 40 µl of water, 2.5 µl of DMSO, 5 µl ThermoPol Reaction Buffer (10X), .5 µl dNTP's (25 mM each), .5 µl 1:10 PFu Turbo (Stratagene) :Taq DNA Polymerase (.5 U:5 U/µl), 2 µl each primer (100 pmol/µl) and 1 µl of template plasmid DNA (20 ng/µl). Thermal cycling began with an initial 2 min denaturing step at 96°C, followed by 20 cycles of touchdown conditions of 96°C for 30 s, 30 s at the annealing temperature which began at 60°C and decreased .5 °C at each cycle, and 5.5 min extension at 68°C. This was followed by 20 standard cycles of 96°C for 30 s, 53°C for 30 s, and 68°C for 5.5 min. After cycling, the reaction was held at 68°C for 7 min, then stored at 4°C until the time of analysis. The product was gel purified, the band excised, and the DNA isolated with the Qiagen gel extraction kit. The isolated DNA was digested with 20 U of AscI for 2 hours, then extracted with phenol/chloroform and ethanol precipitated. The digested DNA was self-ligated in a 50 µl reaction with 40 U of T4 DNA ligase (New England Biolabs, Ipswich, MA) at 16 °C overnight. A sample of the ligation product was transformed into *E. coli* and the plasmids of ten colonies analyzed by restriction digest with EcoRI. One clone containing a plasmid with

the proper restriction pattern was kept. The fragment was transferred to pKC1139 by digestion of both plasmids with HindIII, isolation of the desired fragment by gel electrophoresis, and ligation of the two fragments to generate pKC1139del.

Gene Knockout in *S. coelicolor* CH999. Prior to transformation into *Streptomyces*, DNA must be non-methylated to escape highly efficient restriction systems present in these organisms. pKC1139del was transformed into *E. coli* strain ET 12567 which lacks the capacity to methylate DNA. Non-methylated plasmid was prepared from a 250 ml culture of the transformed bacteria by alkaline lysis followed by RNAse degradation of the RNA and phenol/chloroform extraction and ethanol precipitation (169). The plasmid was then gel purified to separate it from the pUZ80002 plasmid also present in strain ET 12567. Protoplasts of *Streptomyces coelicolor* CH999 were transformed using the rapid protocol of Hopwood (96). Briefly, a 100 µl protoplast aliquot was thawed quickly at room temperature from storage at -80°C and pelleted by centrifugation at 3000 rpm for 7 min in a microcentrifuge. The protoplasts were resuspended in the remaining drop of protoplast storage buffer. To this suspension was added 7 µl of the non-methylated plasmid (100 ng/µl concentration) dissolved in TE buffer (169), all of which was quickly mixed with 100 µl of transformation buffer (protoplast buffer containing 25% w/v PEG 8000) by pipetting up and down 3 times in a 100 µl pipette tip. This mixture was spread plated on an SPRM plate (Soya Protoplast Recovery Media, 103 g sucrose, 1 g Bacto Triptone, 2 g Bacto Peptone, 2 g Yeast Extract, .25 g K_2SO_4, 4.7 g $MgCl_2$, 10 g glucose, .5 g Casamino Acids, 50 mg KH_2PO_4, 3 g $CaCl_2 \cdot 2\ H_2O$, 3 g L-proline, 2 ml Trace Elements Solution, 27 ml 1M Tris • HCl, pH = 7, 20 g soy flour, 20 g Bacto Agar) which had been pre-dried at 30°C for 5 hours to remove any trace condensation that could osmotically rupture the protoplasts. The plate was incubated for 48 hours at 30°C to allow recovery of the protoplasts, then selection for transformants was applied by dropping 1 ml of 1 mg/ml apramycin solution dissolved in water across the surface of the plate and spreading by agitation to achieve uniform coverage. The plate was allowed to dry for 3 hours, then inverted and placed back into incubation at 30°C. Transformed colonies were visible after 48 hours.

The plasmids from three transformants were prepared and analyzed to confirm their presence as follows. A 25 ml culture was grown in R2YE supplemented with 1% glycine and 50 µg/ml apramycin as described for protoplasting above. The mycelial pellet was harvested by centrifugation and a 150 mg portion resuspended in 500 µl TSE buffer (.3 M sucrose, 25 mM EDTA, 25 mM Tris • HCl, pH = 8) containing 2 mg/ml lysozyme in a microcentrifuge tube. This was incubated at 37°C for 1 hour, then lysis was induced by adding 500 µl alkaline lysis

solution (.3 M NaOH, 2% m/v SDS) and gently mixing. The lysate was incubated at 50°C for 20 min, then immediately transferred to an ice bath for 10 min. 500 µl 7.5 M NH_4OAc was added and mixed, and the precipitate removed by centrifugation at 14,000 rpm for 10 min at 4°C in a microcentrifuge. A 750-µl portion of the supernatant was transferred to a clean tube to which was added 800 µl of isopropanol. The solution was mixed and the nucleic acid pellet collected by centrifugation at 14,000 rpm for 10 min at 4°C in a microcentrifuge. The pellet was washed with cold 70% ethanol, then resuspended in 500 µl of TE buffer. RNA was removed by digestion with RNAse, followed by phenol/chloroform extraction and ethanol precipitation, as described above. The remaining DNA was resuspended in 50 µl of TE buffer and and 5 µl aliquot was analyzed by gel electrophoresis. Two of the three transformants contained the plasmid, while the DNA preparation appeared to fail for the third.

Single-crossover recombinants were obtained by spreading a lawn of the second transformant onto a MS plate containing 50 µg/ml apramycin for selection and incubation at 41°C. Plasmid pKC1139 contains a temperature sensitive origin of replication, which causes non-recombinants to lose the apramycin resistance gene when the plasmid fails to replicate. After 3 days of growth, a lawn of sporulated culture was obtained, and the spores from this culture were spread to a second MS plate with apramycin selection, which was allowed to grow to sporulation. Spores from this culture were then plated on a non-selective MS plate and grown at 41°C for 3 days until the culture sporulated. Spores from this plate were transferred to a new, non-selective MS plate, and the plate was grown at 41°C for 3 days until sporulation occurred. This culture contained a mixed population of both single and double crossover recombinants because growth occurred at the non-permissive temperature without selection, allowing spontaneous double crossover mutants to survive and any recircularized plasmid to be lost. This culture was converted to a spore stock by flooding the plate with 2 ml of sterile water and dislodging spores with a plate spreader. The suspended spores were aspirated from the top of the plate to a microcentrifuge tube and centrifuged at 800 rpm for 1 min to remove large suspended bodies, and the supernatant transferred to a clean tube. Spores were harvested by centrifugation at 14,000 rpm for 10 min, the supernatant aspirated off, and the spore pellet resuspended in 1 ml of clean, sterile water.

Eight serial 100-fold dilutions were plated on non-selective MS plates and grown over a strong fluorescent light for 2 days until colonies were visible. The 10^{-4} dilution was used to inoculate four new non-selective MS plates with approximately 100 colonies each. These were allowed to grow for 3 days until sporulation was just visible, then the plates were replica

stamped to MS plates containing 50 µg/ml apramycin using a flat, steel stamp covered with a velvet cloth as described by Hopwood (83, 96). The original plates were refrigerated while the replica plates grew for 2 days over fluorescent light. Once colonies were visible on the selective plates, the pairs of replica plates were compared to identify colonies which had not survived selection. Thirty of these double-crossover recombinants were transferred to new MS plates and incubated at 30°C for 3 days with repeated spreading to generate sporulating lawns.

Screening of the Double-Crossover Recombinants. A PCR screen was used to discriminate between double-crossover events which resulted in deletion mutants and double-crossover events which restored the wild type. Cultures were grown to produce protoplasts similar to the preceeding description, but at a 5 ml scale in 25 ml Erlenmeyer flasks. The cultures were harvested, washed and digested with lysozyme as described before but on a smaller scale. After digestion, the crude protoplasts were collected by centrifugation at 3000 rpm for 7 min in a microcentrifuge, and the supernatant decanted away from the pellet. The protoplasts were resuspended in the remaining drop, then lysed with 500 µl of a 2% solution of SDS. The lysate was purified as previously described above for the isolation of genomic DNA.

Individual 25 µl PCR reactions were subdivided from a stock reaction containing 120 µl of water, 7.5 µl of DMSO, 15 µl of ThermoPol Reaction Buffer (10X), 1.5 µl of dNTP's (25 mM each), 1.5 µl of Taq DNA Polymerase (.5 U:5 U/µl), and 2 µl of each primer (100 pmol/µl). The primers used in this screen were SCKOcheckL (5'-TAC GAC TGG GTC ATG CTC-3') and SCKOcheckR (5'-ATC GAC ACG AAC TTC GAC T-3'). Once the stock was subdivided into individual tubes, 1 µl of the isolated genomic DNA from the recombinants was added. To an additional tube was added 1 µl of wild-type genomic DNA to serve as a control reaction. Thermal cycling began with an initial 2 min denaturing step at 96°C, followed by 40 cycles of 96°C for 30 s, 30 s at the annealing temperature of 60°C, and extension at 68°C for 2 min, 20 s. After cycling, the reaction was held at 68°C for 7 min, then stored at 4°C until the time of analysis. The reactions were analyzed by gel electrophoresis, which indicated that there was a deletion mutant among the first six recombinants analyzed. Spore stocks and protoplast stocks were produced for the mutant, as described above for the wild-type.

Southern Blot Confirmation of Chromosomal Deletion. A Southern blot was used to confirm the deletion mutant, as described by Sambrook et al (169). Briefly, 2 µg of genomic DNA of CH999 wild-type strain and the deletion mutant were digested with 100 U of ApaI in a 50 µl reaction for 8 hours. The two digestions were stopped by phenol/chloroform extraction and loaded into adjacent lanes of a minigel. The minigel was run at 100V for 100 min, and the

lanes cut from the gel. The gel slice was equilibrated with alkaline transfer solution (.4 M NaOH, 1.0 M NaCl), then the DNA transferred by capillary action to a charged nylon membrane (BrightStar-Plus, Ambion) as described by Sambrook et al (169). A ^{32}P-radiolabeled probe was generated in a PCR reaction containing 80 µl of water, 5 µl of DMSO, 10 µl ThermoPol Reaction Buffer (10X), 1 µl dNTP's (30 mM each except dATP, which was 8 mM), 1 µl Taq DNA Polymerase (5 U/µl), 2 µl each primer (100 pmol/µl), 1 µl of genomic DNA (100 ng/ul), and 2 µl of ^{32}P labeled dATP (50 µCi at an activity of 6000 Ci/mmol). The primers used in this reaction were SCKOcheckL, mentioned earlier, and SCSouthernR (5'-CGT ACA CCA CCG GGT ACT-3'). Thermal cycling began with an initial 2 min denaturing step at 96°C, followed by 40 cycles of 96°C for 30 s, 30 s at the annealing temperature of 60°C, and extension at 68°C for 30 s. After cycling, the reaction was held at 68°C for 7 min, then stored at 4°C until the time of probe isolation. The probe was gel purified and isolated using the Qiagen gel extraction kit.

Neutralization of the membrane and hybridization of the probe to the membrane were performed according to Ausubel et al (7). Briefly, the membrane was neutralized in neutralization buffer (1.25 M NaCl, .5 M Tris • HCl, pH = 7.5), then soaked and prehybridized in hybridization solution (480 ml of formamide, 240 ml of 20X SCC (3M NaCl, .3M sodium citrate • $2H_2O$, pH = 7.0), 10 ml of 2 M Tris • HCl, pH = 7.6, 10 ml of 100X Denhardt's Solution (10 g of Ficoll 400, 10 g of polyvinylpyrrolidone, 10 g of bovine serum albumin, diluted to 500 ml and filter sterilized), 50 ml of water, 200 ml of 50% w/v dextran sulfate, and 10 ml of 10% w/v SDS) in a sealed glass hybridization tube in a rotisserie oven for 1 hour at 41°C. The probe was diluted to 100 µl in TE buffer and mixed with 100 µl of 10 mg/ml sonicated herring sperm DNA (Sigma, St. Louis). The solution was heated to 95°C for 10 min, then transferred quickly to ice. One milliliter of the hybridization solution was mixed with the chilled probe and the entire volume transferred to the hybridization tube. The probe was allowed to hybridize with the membrane for 8 hours.

The membrane was rinsed twice with 10 ml of low stringency wash buffer (2X SCC, .1% w/v SDS), then once with high stringency wash buffer (.2X SCC, .1% w/v SDS). 20 ml of high stringency wash buffer was added to the hybridization tube and placed in the rotisserie oven for 45 min at a temperature of 75°C. It was then rinsed twice with low stringency wash buffer and wrapped in cellophane. The banding pattern was determined by autoradiography using a Fujifilm BAS-5000 Imaging Plate Scanner.

Media Conditions for Expression of Mutant Phenotype. The complex media used for manipulation of the deletion mutant resulted in a phenotype indistinguishable from the wild-

type. To find a medium which gave differential expression of the carotene pigment, five different minimal media formulations were prepared. Minimal media (MM) lacking a carbon source was prepared which contained 5 g of Bacto Casamino Acids, 2 g of $(NH_4)_2SO_4$, .5g of $MgCl_2$• 6 H_2O, 1 ml of trace element solution, 1.5 g of NaH_2PO_4, .75 g of KCl, and 20 g of Bacto Agar per liter of water. One-half liter of this solution was prepared and subdivided into 100 ml aliquots. To each was added a different carbon source. To MMG was added 1 g of glucose, to MMCS was added 1 g of cornstarch, to MMSS was added 1 g of soluble starch, to MMGCS was added .5 g of glucose and .5 grams of cornstarch, and to MMGSS was added .5 g of glucose and .5 g of soluble starch. The aliquots were autoclaved and poured to create solid media plates. 10 µl of spore stock from wild type and mutant were plated to each solid media type on opposite halves of the plates and incubated for 3 days over intense fluorescent light.

Minimal Protoplast Recovery Media (MPRM) was prepared containing 103 g of sucrose, 3 g of Bacto Tryptone, 5 g of Bacto Peptone, .25 g of K_2SO_4, 10 g of $MgCl_2$ • $6H_20$, 10 g of glucose, 1 g of Bacto Casamino Acids, 50 mg KH_2PO_4, 3 g of $CaCl_2$ $2H_20$, 3 g of proline, 2 ml of trace element solution, 27 ml of 1 M Tris HCl, pH = 7, and 20 g of Bacto Agar diluted to 1 liter in water. This mixture was autoclaved and poured to produce solid media plates. Protoplasts were transformed with non-methylated pIJ86 as described above and recovered on these plates at 30°C for 48 hours, then treated with apramycin as described before and incubated above an intense fluorescent light for 3 days.

RESULTS

Construction of the Knockout. *Streptomyces coelicolor* CH999 is a strain of *Streptomyces* developed by Khosla and co-workers for investigations into polyketide metabolism and heterologous expression of natural products (32, 88, 123). It is a derivative of strain A3(2), the model strain which was the subject of a recent genome sequencing effort (12). This strain was engineered to contain no endogenous plasmids, a chromosomal deletion of the actinorhodin gene cluster, and a mutation which prevents production of prodigiosin (123). It was therefore an ideal choice for our knockout experiment. This strain clearly showed production of carotenoids when cultured over intense fluorescent light on MS plates, which provide a white background for clearer viewing of the yellow to orange pigments. The color was usually visible after 48 hours of incubation and increased in intensity with longer culture times. After approximately 4 days, colonies entered the sporulation phase of their life cycle, and the orange color became obscured by the bright white aerial hyphae and eventually grey-black spore bodies upon extended culture. Culture on nutrient rich SPRM plates, however, resulted in delayed sporulation and an extended

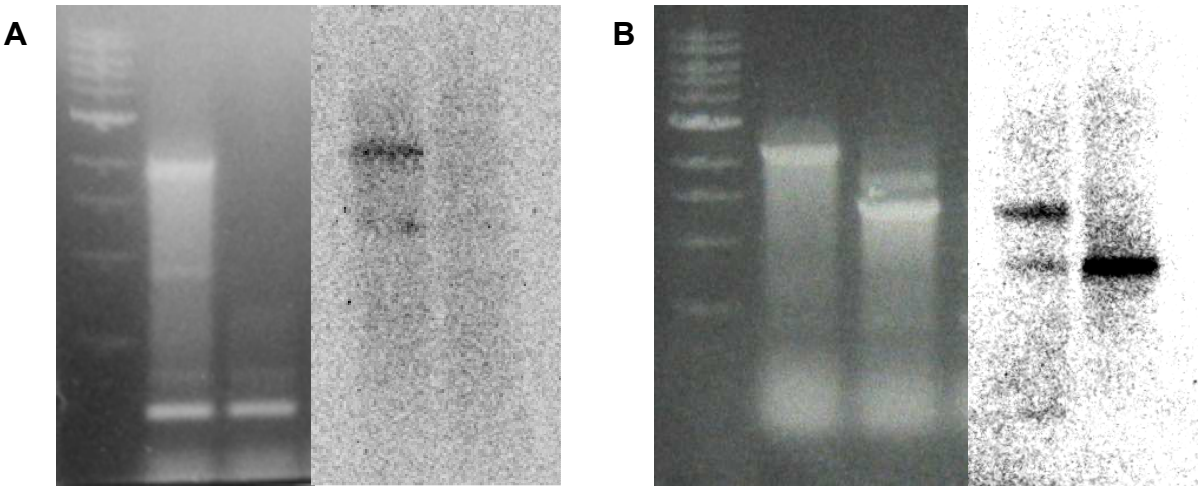


FIG. 27. PCR and Southern Blot confirmation of knockout. (a) PCR and Southern Blot of chromosomal deletion mutant. Gel 1 - PCR analysis, Lane 1 – 1kb ladder (NEB), Lane 2 –wild type, Lane 3 – mutant. Gel 2 – Southern Blot, Lane 1 – wild type, Lane 2 – mutant. (b) PCR and Southern Blot of deletion mutant. Gel 1 - PCR analysis, Lane 1 – 1kb ladder (NEB), Lane 2 –wild type, Lane 3 – mutant. Gel 2 – Southern Blot, Lane 1 – wild type, Lane 2 – mutant.

period of orange coloration with a concomitant increase in color intensity as pigments accumulated for longer periods of time.

Initial attempts at knockout construction failed to identify the proper mutant as a direct result of using the loss of coloration as an indication of successful gene disruption. The carotene pathway targeted for disruption is located near the end of the linear *Streptomyces* chromosome. Chromosomal ends of these bacteria contain non-vital sequences, frequently secondary metabolite pathways, and have been observed to be unstable when propagated in culture (83, 96). Several times, colonies were isolated upon transformation with the disruption plasmid and culture at high temperatures, which later showed antibiotic sensitivity and white coloration after non-selective propagation. These colonies did not give any band at all when subjected to the PCR screen or Southern blot (Fig. 27A), indicating that the entire region was missing from these isolates. Only after abandoning this screen and performing the experiment "blind" as described above were true disruptants identified.

Carotene Expression in Response to Culture Media. Disruption mutants did not show the colorless phenotype expected when grown on SPRM or MS solid medium. These media were chosen on the basis of their white appearance and their ability to induce useful attributes in *Streptomyces* cultures. SPRM is mechanically more stable than R2YE and gave higher transformation and protoplast recovery efficiencies and delayed sporulation. MS plates

gave rapid culture growth and sporulation, which is useful for replica plating and protoplast and spore stock production.

Many of the desirable properties of these media, particularly the coloration, are a function of their high soy flour content, but the soy content may have been causing suppression of the mutant phenotype. Soybean oil is the main source of Vitamin E used in nutritional supplements (92, 198). Vitamin E belongs to the family of natural products known as the

TABLE 7. Tabulated data for minimal media.

Media (Abbreviation)	Minimal Salts and Amino Acids	Glucose (G)	Cornstarch (CS)	Soluble Starch (SS)	Knockout Pigmentation	Contrast with Wild Type
Minimal Media + Glucose (MMG)	+	+	-	-	-	++
Minimal Media + Cornstarch (MMCS)	+	-	+	-	+	-
Minimal Media + Soluble Starch (MMSS)	+	-	-	+	+	-
Minimal Media + Glucose + Cornstarch (MMGCS)	+	+	+	-	-	+
Minimal Media + Glucose + Soluble Starch (MMGSS)	+	+	-	+	-	+

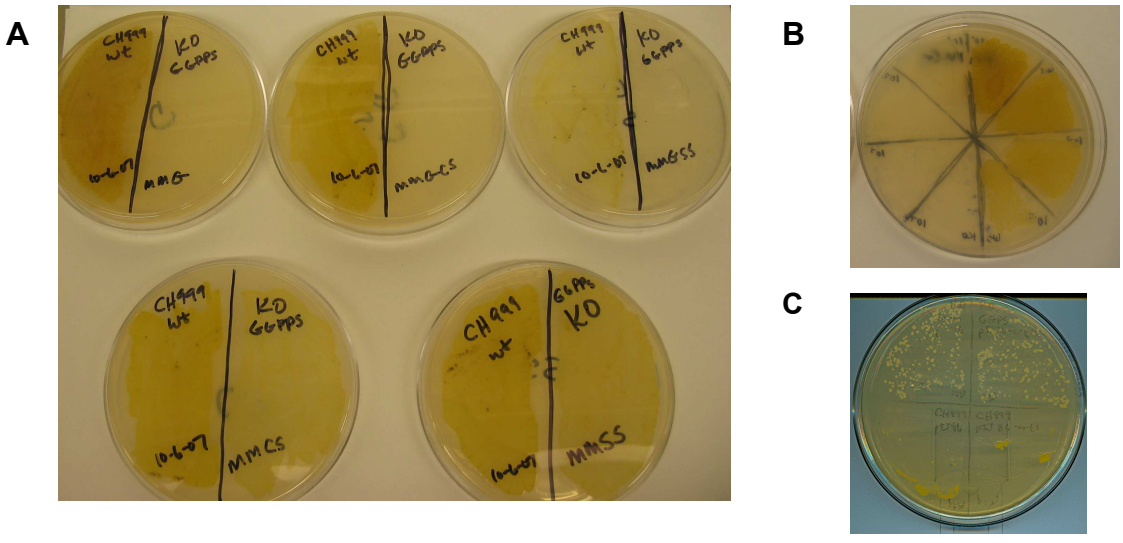


FIG. 28. Effect of minimal media composition on phenotype. (a) Coloration of cultures of wild-type CH999 and deletion mutant on various formulations of minimal media. From upper left to lower right, media are MMG, MMGCS, MMGSS, MMCS and MMSS. Wild-type is plated on the left of each plate and the deletion mutant is on the right. (b) Serial dilutions of wild-type and deletion mutant grown on MMG. Dilution increases from top to bottom, mutant strain on the left and wild-type on the right. (c) Protoplasts recovered on MPRM after transformation. Upper half is knockout strain, lower half is wild type.

tocopherols. These compounds are condensation products of GGPP and homogentesic acid. If biosynthesis of the tocopherols is occurring in the soybean, it is highly likely that upstream intermediates, like GGPP, or other byproducts, such as geranylgeraniol, may be present in significant amounts in soy flour that could complement the gene disruption in the mutant strain. It is also possible that a homolog of this gene is complementing the loss of activity created by the gene disruption. *Streptomyces coelicolor* A3(2) contains five other genes with high sequence homology to GGPPS.

Growth of the wild-type and mutant strains on the various minimal media compositions gave a range of contrast in production characteristics between the two strains (Fig. 28, Table 7). On all media, the wild-type strain was consistently dark orange, although the glucose/soluble starch mixture was lighter than the others and the glucose-only plate was slightly darker. The mutant, however, ranged in coloration from indistinguishable with the wild-type strain to nearly white. In addition, the mutant showed markedly different growth rates, with the least pigmented colonies approximately one-fifth the size of the most pigmented colonies after 3 days of growth. Media which contained starch alone as a carbon source resulted in levels of pigmentation which were not significantly different from the wild type, while glucose alone gave mutant colonies which were only pale yellow and very small. Mixtures of the two carbon sources gave colonies which, while slightly darker and larger than those which used glucose as the sole carbon source, were still much lighter and smaller than those with starch as the only carbon source. Overall, it appeared that the presence of glucose suppressed both pigmentation and the colony growth rate, and MMG media, which contained only glucose as a carbon source, gave the best contrast between phenotypes. Colony size increased proportionately with pigmentation, probably due to increased protection from oxidative damage induced by the fluorescent light as pigmentation increased. This information was used for the formulation of MPRM. Protoplasts recovered on MPRM showed good color contrast between the wild-type and mutant strains and delayed sporulation, although there was less colony size contrast.

DISCUSSION

Although large insert genomic library construction was quite challenging in these studies, library screening proved to be the major confounding hurdle. Genetic complementation is a robust screening technique that has been used by molecular biologists for many years. Some of the earliest microbial genetic experiments, such as the famous Lederburg experiments elucidating the fundamental genetics of *E. coli*, made use of auxotrophic strains and their complementation to detect the transfer of genetic information (109). In the case of auxotrophy,

where a nutrient required for survival can no longer be made by the organism, the genetic screen is converted into a genetic selection since the observable characteristic of complementation is survival on minimal media. In our lab, Eun Jin Kim recently demonstrated this approach in the elucidation of biotin biosynthesis in a strain of the marine microorganism *Chromohalobacter* (98).

A variety of expected and unexpected difficulties arose which made the construction of a GGPS complementation system in *Streptomyces coelicolor* challenging. First, the location of the pathway in an unstable region of the chromosome resulted in frequent loss of the entire pathway. This mass chromosomal deletion was highly unexpected and was finally detected as the absence of an observable band in the Southern blot of the isolated "knockout." Although the gene was technically knocked out in these mutants, the pathway was no longer capable of complementation since the remainder of the enzymatic activities was not present.

A much more foreseeable difficulty was the observation of internal complementation, a phenomenon commonly referred to as functional redundancy. This occurs when another gene within the same organism performs the function of the deleted sequence, though usually at a lower efficiency. It is particularly common in eukaryotes, which frequently exhibit a diploid genetic structure and contain a large amount of repetitive and "unused" sequence. It is also observed in prokaryotes, but in this case there is usually a much more severe phenotypic consequence to gene knockout due to prokaryotes' more efficient use of sequence and lack of sequence redundancy. In a closely related case, Fujisaki et al describe the deletion of the FPP synthase in *E. coli* (58). FPP is a vital metabolite for respiration, and although the mutants were viable, they grew at slower rates under aerobic conditions and showed drastically reduced production of quinones necessary for electron transfer. An assay showed that the cells had 5% of the wild type activity of the enzyme, even with the deletion, which they attribute to octaprenyl and undecaprenyl synthases present in other regions of the genome. So while it could have been anticipated that internal complementation might occur as a result of residual activity supplied by functional analogs elsewhere in the genome, and further that this activity might interfere with the application of the screen since heterologous expression from a non-native promoter would in many cases provide only low levels of activity, it could reasonably be assumed that there would at least be a noticeable difference in the phenotype of the mutant. This was not the case.

Fortunately, the activity appears to be manageable by changes in medium composition. The observation that soy flour, which, unlike most processed flours, is not defatted during processing, might lead to complementation by providing an external supply of geranylgeraniol

led to experimentation with minimal media. These experiments show that media conditions can be used to differentiate between the mutant and wild type, and further suggest that the presence of glucose suppresses carotene production in comparison to starch. This observation makes sense in light of current understanding of *Actinomycete* biology. It is believed that *Actinomycetes* produce antibiotics at certain points during their life cycle as a means of sequestering nutrients at a particularly vital point (34, 83, 96). During the mycelial growth phase, the cells multiply and use the nutrients of their surroundings for nourishment and growth. At this time, they typically do not produce antibiotics. It is when nutrients become depleted that the phase of the growth cycle shifts into a spore production phase and antibiotic production begins. The mycelial mat becomes senescent and releases its nutrients back into the media. Aerial hyphae grow and feed off this rich nutrient source, in effect cannibalizing the formerly productive biomass and using the nutrients to produce spores, which are used to spread the bacteria to new nutrient sources. It is believed that the purpose of antibiotic production is to ensure that other species are held at bay while the *Actinomycete* mobilizes these nutrients for its own purposes. An important observation which supports this idea is the behavior of *Streptomyces* strains which contain a family of mutations called "bald" mutations, *bld* (34, 83, 96). *Bld* mutants are incapable of forming aerial hyphae and spore bodies, so colonies with these mutations look smooth relative to their wild-type counterparts. This inability to sporulate is associated with a loss of antibiotic production (34, 83, 85, 96).

The media used for cultivation can therefore have profound impacts on gene regulation, especially for genes involved in cellular differentiation and secondary metabolite production. The results for growth of these strains on SPRM and MS illustrate this well. SPRM probably causes delayed sporulation because it is extremely rich in simple sugars and other easily metabolizable nutrients, while MS has most of its nutrients stored in complex biopolymers which must be broken down. *Actinomycetes* appeared on earth approximately 450 million years ago, and appear to have attained their evolutionary success by being able to metabolize the less metabolically accessible components of plants, which had appeared 100 million years earlier (34). They typically contain a wide array of enzymes for catabolizing complex biopolymers, such as amylases, cellulases, xylanases, even chitinases and agarases, and are perfectly capable of metabolizing the nutrients present in complex media like MS, as evidenced by their ability to grow to high cell densities in these types of media. MS media, and even the minimal media used in these experiments, is certainly not nutrient poor from a *Streptomyces* point of view, but if the simple sugars are tied up in biopolymers and must be mobilized by excreted enzymes, their

ambient concentrations will be low, which could serve as the trigger to induce the sporulation phase of growth. Assuming that the GGPPS complementing activity came from another secondary metabolite pathway, this could be the reason that a minimal medium actually led to suppression of secondary metabolite production in the case at hand, simply because it contained glucose as opposed to starch or cellulose. It may not be nutrient deficiency or starvation *per se* that induces secondary metabolite production, but the capacity of the media to induce the senescence phase of growth in mycelia by activating specific regulatory triggers. Glucose concentration appears to be one of those triggers, at least in this strain.

This would not explain the orange phenotype of the mutant when grown on SPRM, however. SPRM contains both glucose and starch and suppresses sporulation, and should by extension suppress mycelial senescence and secondary metabolite production. There appear to be two possibilities which might account for this apparent contradiction. Either there is complementing GGPPS activity arising from a pathway within the cell as a result of a separate set of regulatory mechanisms, or an advanced intermediate such as geranylgeraniol is being supplied externally by the medium. These experiments cannot distinguish between the two, although the second possibility appears to be more probable simply because it is the less complex explanation and the only substantial difference between MPRM and SPRM is the presence of soy flour and yeast extract in the latter. Further experimentation would be needed to provide more insight.

The recovery of protoplasts on MRPM indicates that this strain could be successfully transformed in an environment in which differential pigmentation could be observed. It does not necessarily indicate that the screen would work. A particularly important implication of the preceeding discussion with respect to potential screening applications of this strain is that useful media conditions for screening will be quite limited. This is unfortunate, because media variation has in the past been used as a powerful approach for causing recalcitrant pathways to express their products. It might be possible to expand the available media by performing additional knockouts of the remaining GGPPS homologs, but this would be a laborious task, especially if it is not known which homolog is performing complementation in a given situation. On the brighter side, it is also quite possible that the expression of a heterologous pathway may be enhanced simply because the non-native environment will not contain many of the regulatory elements, especially transcriptional repressors, found in the original host. Even a limited range of media conditions could reveal the presence of some pathways, if not the majority of them.

In conclusion, these experiments show that it is possible to selectively knock out the GGPPS activity of the light-inducible carotene biosynthetic pathway of *Streptomyces coelicolor* CH999 while maintaining the integrity of the remainder of the pathway, as well as providing guidance for choosing culture conditions for the manipulation of the pigmentation phenotype of this strain. This mutant strain may prove to be an important tool for the screening of metagenomic libraries for the presence of terpene secondary metabolite pathways. These experiments also provide clues about the regulatory roles of carbohydrate sources on growth, sporulation, and the induction of other isoprenoid pathways in this species.

CHAPTER VII

CONCLUSIONS

In these experiments, it was found that natural product research beginning from the genetic end of the production machinery and working towards active compounds is considerably more challenging than anticipated. DNA purification and manipulation posed serious hurdles for the generation of large-insert genomic libraries, particularly for the production of BAC libraries in the 100 kb size range and for mixed-culture samples. Conditions were far too specific to particular samples to be generally applicable in a more high-throughput setting. In addition, the design of degenerate primers general enough to use on a variety of samples, yet specific enough to give high-yielding reactions with little background was nearly impossible. As a result, these approaches are probably not practical for large-scale compound isolation schemes at this time. However, there are many other approaches, and indeed on an almost daily basis new technologies are added to the molecular biologist's toolbox which only adds to the possibilities. It could be that future techniques will have little resemblance to current practices, and the entire paradigm of library construction, probe design, and screening to find clones will be entirely displaced by more efficient processes.

NEW DEVELOPMENTS

Several recent developments point in this direction. In particular, sequencing methods have improved manifold, reducing the cost of this type of information dramatically (186). The impetus for cheaper, faster, and more accurate sequencing technology has primarily been driven by perceived benefits to the practice of medicine and human health. The Human Genome Project is one of the most visible and important manifestations of this drive for the application of molecular biology to medicine (101, 186, 201). Since its completion it has been followed by the $1000 Genome Project, which seeks to develop the technology to allow a human genome to be sequenced for $1000, presumably within the reach of the average citizen of a developed country (186).

The pursuit of these large scale sequencing projects and others has lead to many advances in sequencing technology. Notable among these was the approach of whole-genome shotgun sequencing, as opposed to directed sequencing, and high-throughput advances to Sanger sequencing technology (201, 202). The $1000 Genome Project has itself spawned a new round of advances. Several groups, including 454 Life Sciences and George Church's group at

Harvard Medical School, have produced solid-phase approaches which allow unprecedented levels of speed and throughput (119, 187). "454-sequencing" is capable of sequencing a microbial genome for $60,000. While this is still a few orders of magnitude away from the $1000 for ~30 Gb of data needed to satisfy the $1000 Genome Project's goals, it is a vast improvement over currently available Sanger technology. It is so powerful that it has allowed recovery of large quantities of sequence information of the wooly mammoth and Neanderthal genomes (68, 157), which were so fragmented and heavily contaminated with exogenous DNA that Sanger sequencing was untenable, and is capable of sequencing entire microbial genomes at high coverage levels within a single day. It has already been used for the sequencing of several microbial genomes only 2 years after its invention (66, 81, 142). Yet as impressive as this is, other researchers are already pondering the feasibility of single molecule sequencing (9). It is already possible to thread a single molecule through a nanopore embedded in a membrane. It remains to be seen whether this type of approach or other similar approaches can top the performance of the new solid phase methods developed so far, but if history is any guide, something will.

A second great development has been the availability of large quantities of microbial sequences, primarily a result of the reduction in the cost of acquiring sequence. Since 1990, the number of fully sequenced microbial genomes available publicly on NCBI has increased from 6 to 520 (Fig. 29). The number of new genomes available per year has increased from only a handful at the beginning of that time period to 174 last year. The growth over this time period is almost perfectly exponential. This is significant for several reasons. First, the sheer build up of genomic data on these organisms reveals new pathways for study whether purposely, in the case of those organisms selected for their reputation as natural product factories such as the *Actinomycetes* and *Streptomycetes*, or inadvertently, as in the case of those medically interesting organisms which also happen to contain uncharacterized biosynthetic gene clusters. More importantly, the availability of this sequence data gives important insight into the undiscovered pathways of new isolates and metagenomic samples. With information such as codon usage, GC content, amino acid homology, genetic organization, and promoter structure, cloning and manipulating pathways from novel organisms becomes a simpler task.

It is not just the culturable organisms and bulk environmental samples which are being sequenced. Technology now exists to generate complete genome sequences from individual cells. Dean and co-workers have produced a method of amplification of single genomes using a

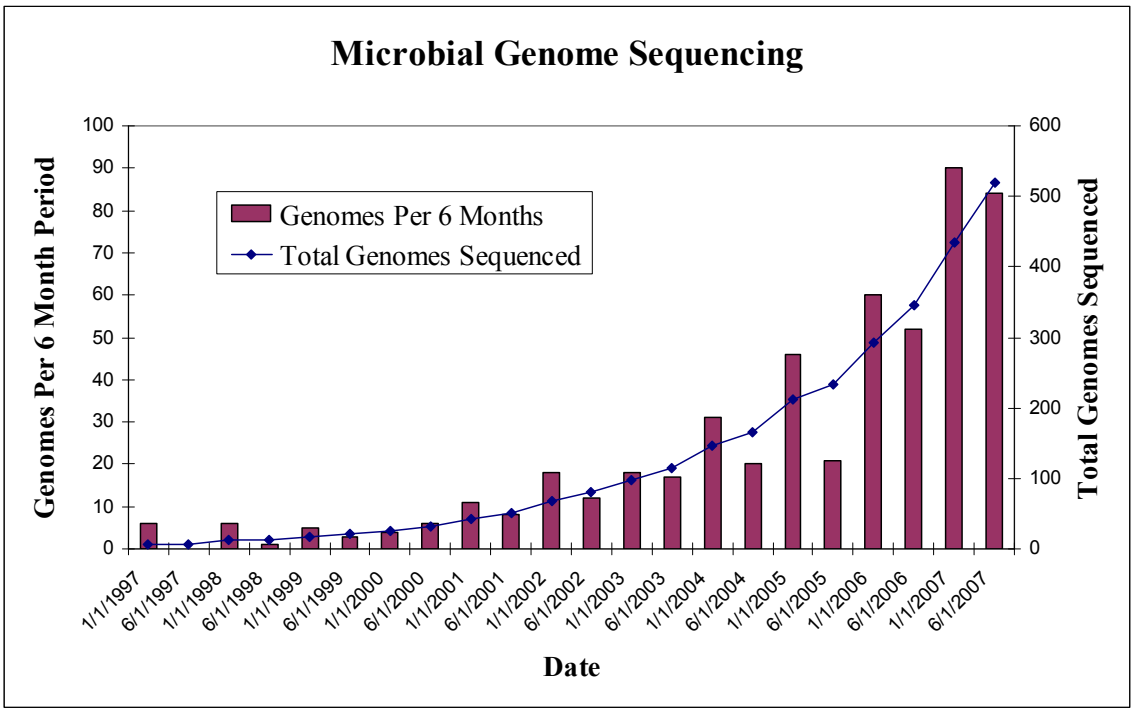


FIG. 29. Progress of microbial genome sequencing efforts available on NCBI.

high-fidelity, high-processivity phage DNA polymerase to generate quantities of DNA large enough and unbiased enough for sequencing projects called multiple displacement amplification (MDA) (46, 146). The MDA procedure is capable of amplifying the genome 10,000-fold, yet the reaction has so little bias that most sequences are over or underrepresented by no more than 3-fold from the average, a vast improvement over standard random priming PCR, which typically gives such bias on the order of multiple orders of magnitude (146). George Church's group has developed this and other amplification techniques for whole genome sequencing projects from single cells (127, 187, 223). The amplified DNA pools are called "polonies" or "plones," depending on their origin and function, which is short for "polymerase derived" colonies or clones, whatever the case may be. Once amplified, the DNA is sequenced by standard methods. They applied it to several species, showing that genome sequencing by this method was sufficient for sequencing environmentally derived organisms available in only single-cell quantities (223). Church proposes that metagenomic sequencing be performed by this route, employing a precise dilution step prior to polony formation to allow individual genomes to form each polony, then sequencing the amplified DNA from each separate reaction. Such a system would produce large quantities of sequence data very quickly and efficiently, and would

have the added advantage of being able to assemble contigs more easily and produce whole genomes of even rare organisms. This would be an improvement over bulk sampling which produces all sequence data in one giant pool and necessarily oversamples organisms present in higher numbers, therefore increasing the sample size needed to observe rarer sequences and making their study more difficult.

Manipulation of large DNA constructs has also been revolutionized by new recombination tools. For the last 25 years or so, the primary means of manipulating DNA has been the use of restriction enzymes and DNA ligases coupled with the use of chemical oligonucleotide synthesis and mutagenic PCR to generate custom synthetic sequences. While very useful for DNA fragments in the 100 bp to 10 kb size range, larger fragments become unwieldy due to the difficulty of finding unique restriction sites, and at sizes greater than 50 kb shearing becomes a problem. Recently, however, many researchers have generated powerful new tools for highly specific DNA rearrangements based on DNA recombinase systems which overcome these difficulties. Several systems, including the yeast FLP recombinase and the Cre-Lox system, are capable of performing inversions and deletions between two conserved sites that are inserted synthetically (97, 135). These sites are large enough and specific enough that the probability of unwanted rearrangements due to chance is virtually zero, even in a complex genome. Even more promising, however, is the development of the λ-recombinase and the closely related Rac prophage recombinase systems for use in *E. coli*. Several groups, including Francis Stewart's at the University of Technology, Dresden, Barry Wanner's at Purdue University, Keenan Murphy at the University of Massachusetts Medical School, and Donald Court's at the National Cancer Institute, have developed transient expression systems for the recombinase system of bacteriophage λ (43, 131, 132, 134, 218, 225). The recombinase system consists of three components with three activities: an exonuclease, a recognition and strand invasion component, and a component which inhibits the *E. coli* RecBCD complex (159). The RecBCD complex inhibits the desired recombination event, probably because it has an exonuclease activity which degrades intermediates in the recombination pathway (133). DNA ends are first bound and trimmed by the exonuclease to reveal a 3' single-stranded region. The single strand generated then displaces the base pairing partner in its homolog, and the final resolution of this structure effectively splices the linear DNA into the site of homology. When both ends of a linear DNA are so treated, the effect is analogous to a double-crossover event. These systems are capable of recognizing homology as low as 50 nt at the ends of linear DNA and integrating with very high efficiency. Because the regions are so small and because they are

not specific to the recombinase, they can be incorporated into oligonucleotides and leave a "seamless" recombination site, unlike their FLP and Cre counterparts.

Francis Stewart's group in particular has adapted this system to perform a wide range of useful tasks, which he terms recombineering, such as modifying BAC vectors (134), assembling large constructs (149, 210), and even direct cloning of target sequences (226). These types of manipulations are particularly promising, because they open the door to modification of pathways for engineering, of assembly of large constructs from smaller library clones, and, if appropriately developed, may eliminate the necessity of library construction altogether. This could greatly accelerate pathway discovery and move pathways rapidly towards expression in heterologous hosts.

This ability to manipulate such large constructs in ways never before available was recently exploited to heterologously express the myxochromide S pathway from *Stigmatella aurantiaca* in *Pseudomonas putida* (210). The pathway was originally located on two separate cosmids from a genomic library of the producing strain. The cosmids were "recombineered" to generate a new plasmid which contained the entire pathway, equipped with a tuoluic acid inducible promoter which is active in *Pseudomonas*, and a site for integration into the new host. The new strain actually gave higher yields than the original producer and was simpler to grow and harvest. This experiment clearly demonstrates that these techniques eliminate the need to clone pathways intact because reconstruction and modification are relatively straightforward and screening of the library was based on sequence homology and was not dependent on detection of the product.

That this technique is amenable for cloning from a high-complexity DNA sample was demonstrated in another experiment in which linearized plasmid was co-electroporated with the sample containing the gene of interest, such that recombination resulted simultaneously in cloning and plasmid rescue (226). The linearized plasmid contained homology arms that flanked the region of interest in a genomic DNA sample. Upon mixing the two samples and electroporation, antibiotic resistant transformants were observed, indicating that a recombination event had circularized the plasmid. Only a fraction of the population of transformants contained the correct insert, while the remainder was the result of undesired, spurious recombinants. This undesired recombination was shown to occur at very short sites of homology within the plasmid, as low as 6 bp. Still, screening from among these would be a simple task, and with improvement such a strategy could become a routine method for cloning large inserts from complex mixtures without the need for library production. From that point, careful construction of the cloning

plasmid might just make transfer to heterologous hosts and expression a trivial issue. The technique was shown to work on several genes from several genomes, but the example from the mouse genome was most significant because it is complex enough to mimic cloning from mixed microbial samples.

Another technology called transformation-associated recombination (TAR) cloning looks even more promising for directed cloning of large DNA fragments (14, 17, 101, 105, 139, 184). This method is not particularly new, the earlier experiments having been performed in the mid-1990's (105), but it has not seen widespread application presumably due to a lack of sequence information. As more sequence data becomes available, TAR cloning will doubtless have more utility. Other researchers have used the technique to produce some very interesting plasmid manipulation strategies which could also find application in metagenomic cloning procedures (143, 161, 163). The process is very similar to the previously described methods, except that *Saccharomyces cerevisiae* is the host organism and endogenous yeast recombinases are used instead of λ-recombinase. A linearized vector containing homology arms that flank the sequence of interest is constructed and co-transformed with the genomic DNA of interest into yeast spheroblasts. Double-crossover events cause circularization and rescue of the plasmid. In this case, counterselection markers are particularly important because the vector has a strong tendency to recircularize through non-homologous recombination, generating an empty vector (139, 162). Placing a counterselection marker at the cloning site reduces this undesired background. DNA fragments as large as 250 kb have been cloned this way, indicating that the method is more than adequate to handle even very large pathways. *E. coli* has some difficulty accepting DNA of this size, and transformation efficiencies decrease drastically with increasing size even at much lower size ranges (188). This small-insert bias is one of the major drawbacks of BAC library construction. YAC's, on the other hand, are routinely constructed with multi-megabase sized inserts.

THE FUTURE OF GENOMIC AND METAGENOMIC CLONING

What does the future hold for this field? Clearly the price of sequence data is getting lower and lower with each passing year, and at some point in the near future it may become economical to investigate genomic and environmental samples through shotgun sequencing as a complement to large insert library construction, probing and screening. It also seems clear that the new precision recombination based tools of molecular biology are undermining the case for the construction of large insert libraries based on the random methods of partial digestion and shearing to obtain large DNA fragments. Having little utility on either the upstream

investigative end or the downstream tailoring end of the process, it appears that such methods may no longer be used.

The issue of costs not withstanding, the simplest strategy going forward might be one of "sequence first, ask questions later" (Fig. 30). A typical experiment will probably closely resemble something like the following. Samples may first be sequenced, the data assembled into contigs by computer, then searched for pathways of interest. An added bonus of getting the sequence information up front is that many more types of pathways will be amenable to cloning, not just the ones for which degenerate primers are easily constructed. Promising pathways will be selected and cloned using custom vectors constructed by PCR complete with homology arms, selection and counterselection markers, host chromosomal insertion sites, and species specific inducible promoters. The approaches depicted in Fig. 31 and Fig. 32 represent two possible examples that would utilize only currently existing, proven technology for specific cloning of pre-sequenced pathways. Because only the homology arms will vary for cloning each pathway, vectors may be modified rapidly via casette insertion prior to each cloning attempt. Each vector will be matched with culturable hosts sampled from across the entire range of bacterial taxonomy, so that pathway homology, codon usage, and cofactor requirements can be matched with an appropriate host. The pathways will be cloned by recombination from bulk DNA or MDA amplifications into yeast or *E. coli*, tailored if necessary using the methods of recombineering, then transferred into host strains. Extracts from these strains can be thoroughly screened for metabolite production easily, since only one or a few will be produced for each pathway sought.

CONCLUSION

If these advances are any indication, the future of both metagenomics and the field of molecular biology will look very different from the past. Currently, genomic sequencing is still prohibitively expensive for species that are not intrinsically interesting, so a pure sequencing approach for natural product biosynthetic pathway discovery would not make sense just yet. However, this situation is rapidly changing, and such a change could be anticipated in the near future. In the meantime, there is an abundance of sequence data being generated now through committed genome sequencing laboratories that would allow such research on strains which have been deemed worthy of genomic sequencing efforts for other reasons. There is already an abundance of interesting targets and the tools for integrating directed cloning and heterologous expression in a manner similar to that previously described could and should be developed now.

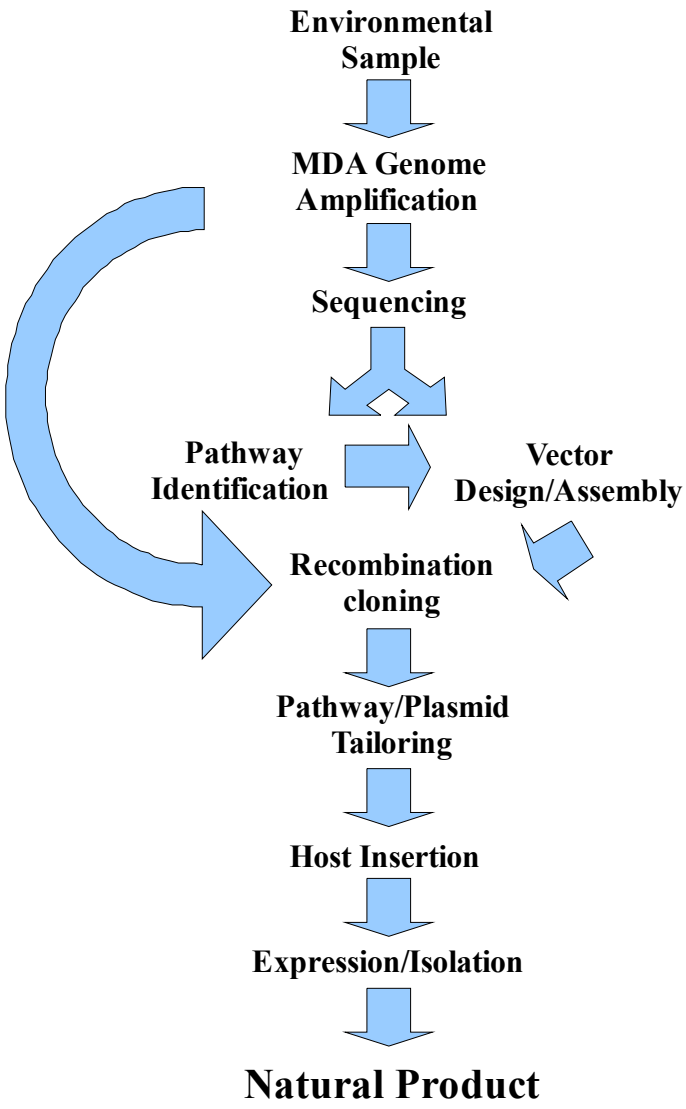
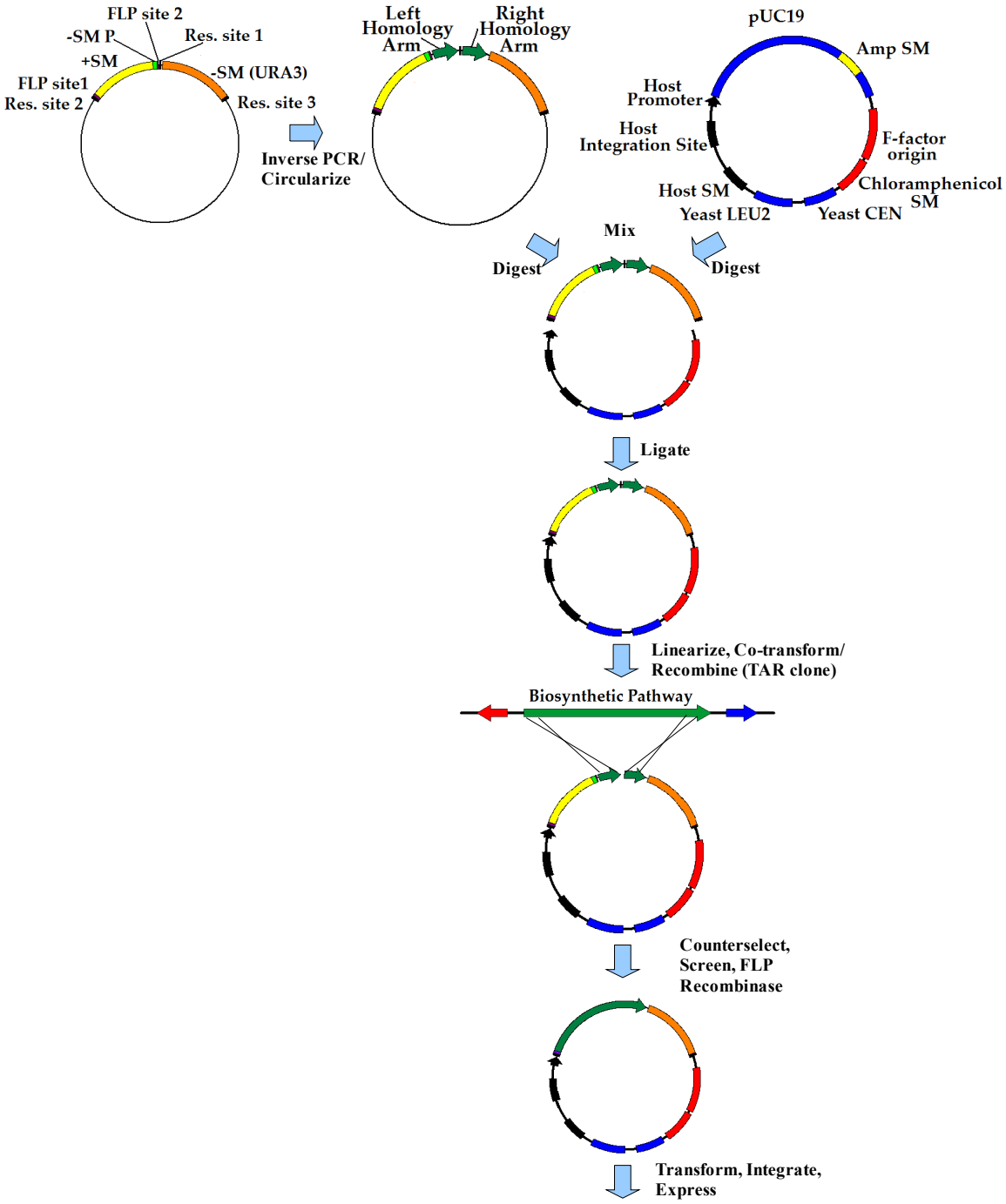


FIG. 30. Flowchart for future biosynthetic pathway cloning experiments. Individual environmentally isolated cells are MDA amplified and polony sequenced. Sequence data are used to identify pathways of interest, genetic architecture, and strain identification for host matching. Recombination targeting vectors are designed and constructed on the basis of the pathway sequence. The desired sequence is cloned by TAR cloning or another recombination method, then the expression vector tailored for optimal expression. The new construct is transformed into a new host as an integrating plasmid and the compound isolated from this producing strain.

Future work in this area should focus on developing these techniques, further reducing sequencing costs, and especially in developing new hosts and genetic systems for expression. With time and any luck, this exciting and rapidly developing field will finally tap into the full potential of the metagenome and result in a torrent of discoveries of new and promising molecules.

FIG. 31. Suggested method for routine construction of directed recombinational cloning vectors. Possible method for routine vector construction and TAR cloning. The starting vector should contain an *E. coli* F-factor or P1 origin plus the chloramphenicol or other suitable selectable marker for single-copy propagation in *E. coli*. It should also contain a *S. cerevesiae* centromeric sequence and a selectable marker such as LEU2 for leucing auxotrophic strains for TAR cloning and propagation in *S. cerevesiae*, as well as an inducible promoter, selectable marker, and integration site for eventual integration into the final host. These sites would vary from host to host. In addition, it is useful to contain a copy of pUC19 inserted into the cloning site, preferably a site suitable for TA cloning, so that high-copy propagation is possible in preparation for cloning. The pUC stuffer fragment is digested out to generate an acceptor vector for the recombination cassette. Simultaneously, the recombination cassette is prepared by inverse PCR amplification of plasmid containing an *E. coli* positive selectable marker (+SM) flanked by FLP recombination sites adjacent to a yeast negative selectable marker (-SM), such as URA3, with a unique restriction site located between them to allow for linearization in later steps. The site of amplification should be between the promoter of the yeast negative selectable marker so that the pathway cloning step will inactivate gene expression. The primers contain 50 bp flanking homologous sequence to the desired pathway, such that the PCR product generates the homology arms in the proper orientation between the yeast negative selection marker and its promoter. This distance of separation will not interfere with gene expression. Appropriate restriction sites should be included for ligation into the vector. This insert is ligated into the vector to generate a pathway specific recombination targeting vector. The vector is prepped, linearized and co-transformed into yeast spheroblasts with the MDA amplified or bulk genomic DNA sample. Recombinants are selected by plating on media containing 5-fluoroorotic acid and lacking leucine. The final plasmid should be sequenced, and is then ready for FLP removal of the *E. coli* selectable marker, any necessary tailoring, and eventual transfer into the new host for expression.


FLP site 2
-SM P
Res. site 1
+SM
-SM (URA3)
FLP site1
Res. site 2
Res. site 3
Inverse PCR/
Circularize
Left
Homology
Arm
Right
Homology
Arm
pUC19
Amp SM
Host
Promoter
Host
Integration Site
F-factor
origin
Chloramphenicol
SM
Host SM
Yeast LEU2
Yeast CEN
Digest
Mix
Digest
Ligate
Linearize, Co-transform/
Recombine (TAR clone)
Biosynthetic Pathway
Counterselect,
Screen, FLP
Recombinase
Transform, Integrate,
Express
Natural Product

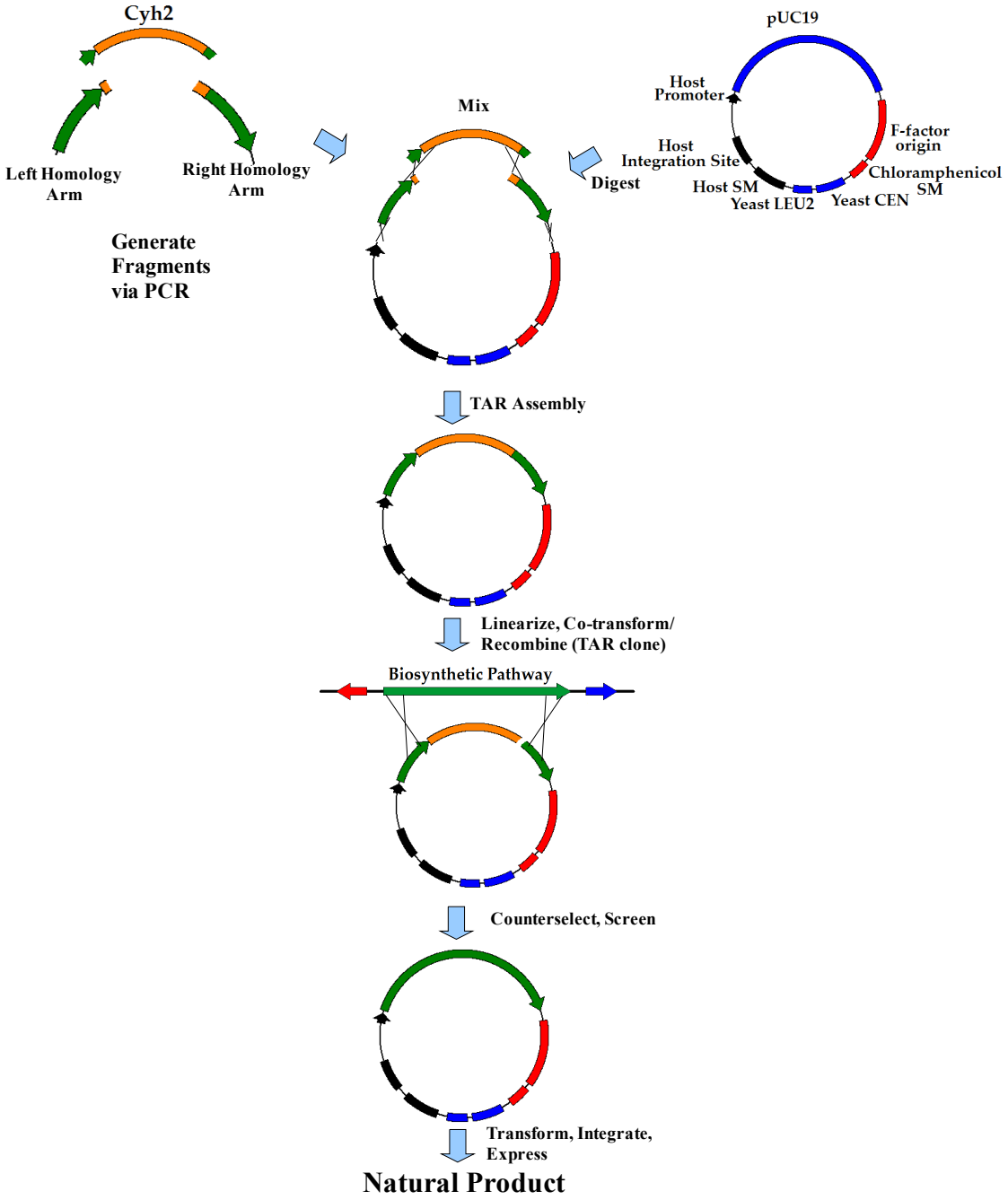


FIG. 32. Second method for routine construction of directed recombinational cloning vectors. Second possible method for routine vector construction and TAR cloning. The starting vector is similar to the previous example and is prepared similarly. The recombination cassette is prepared by the methods of yeast recombinational plasmid assembly (143, 161, 163). Homology arms and counterselectable marker (cyclohexamide sensitivity conferred by CYH2) are amplified with primers that contain regions of homology to one another and the vector, such that co-transformation results in the assembly of the cassette. The directed cloning vector produced contains ~500 bp homology arms flanking the CYH2 locus, with a unique restriction site between the CYH2 locus and one of the arms. The plasmid is linearized at this site and co-transformed again into yeast with the source DNA, resulting in pathway cloning as before.

REFERENCES

1. **Abramova, N. E., B. D. Cohen, O. Sertil, R. Kapoor, K. J. A. Davies, and C. V. Lowry.** 2001. Regulatory mechanisms controlling expression of the DAN/TIR mannoprotein genes during anaerobic remodeling of the cell wall in *Saccharomyces cerevisiae*. Genetics **157:**1169-1177.
2. **Agnew, W. S.** 1985. Squalene synthetase. Methods Enzymol. **110:**359-373.
3. **Akselband, Y., C. Cabral, T. P. Castor, H. M. Chikarmane, and P. McGrath.** 2006. Enrichment of slow-growing marine microorganisms from mixed cultures using gel microdrop (GMD) growth assay and fluorescence-activated cell sorting. J. Exp. Mar. Biol. Eco. **329:**196-205.
4. **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification and in-situ detection of individual microbial-cells without cultivation. Microbiol. Rev. **59:**143-169.
5. **Angell, S., B. J. Bench, H. Williams, and C. M. Watanabe.** 2006. Pyocyanin isolated from a marine microbial population: synergistic production between two distinct bacterial species and mode of action. Chem. Biol. **13:**1349-1359.
6. **Austin, M. B., and A. J. P. Noel.** 2003. The chalcone synthase superfamily of type III polyketide synthases. Nat. Prod. Rep. **20:**79-110.
7. **Ausubel, F. M., R. Brent, R.E. Kingston, D.D. Moore. J.G. Seidman, J.A. Smith, K. Struhl, eds.** 1998. Current protocols in molecular biology. John Wiley & Sons, New York.
8. **Avery, A. M., and S. V. Avery.** 2001. *Saccharomyces cerevisiae* expresses three phospholipid hydroperoxide glutathione peroxidases. J. Biol. Chem. **276:**33730-33735.
9. **Bayley, H.** 2006. Sequencing single molecules of DNA. Curr. Opin. Chem. Biol. **10:**628-637.
10. **Beja, O., M. T. Suzuki, E. V. Koonin, L. Aravind, A. Hadd, L. P. Nguyen, R. Villacorta, M. Amjadi, C. Garrigues, S. B. Jovanovich, R. A. Feldman, and E. F. DeLong.** 2000. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. Environ. Microbiol. **2:**516-529.

11. **Bennett, M. D., and I. J. Leitch.** 1995. Nuclear-DNA amounts in angiosperms. Ann. Bot. **76:**113-176.

12. **Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, and D. A. Hopwood.** 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature **417:**141-147.

13. **Berdy, J.** 2005. Bioactive microbial metabolites - a personal view. J. Antibiot. **58:**1-26.

14. **Bhargava, L., C. S. Shashikant, J. L. Carr, H. Juan, K. L. Bentley, and F. H. Ruddle.** 1999. Direct cloning of genomic DNA by recombinogenic targeting method using a yeast-bacterial shuttle vector, pClasper. Genomics **62:**285-288.

15. **Blagg, B. S. J., M. B. Jarstfer, D. H. Rogers, and C. D. Poulter.** 2002. Recombinant squalene synthase. A mechanism for the rearrangement of presqualene diphosphate to squalene. J. Am. Chem. Soc. **124:**8846-8853.

16. **Blunt, J. W., B. R. Copp, M. H. G. Munro, P. T. Northcote, and M. R. Prinsep.** 2006. Marine natural products. Nat. Prod. Rep. **23:**26-78.

17. **Bradshaw, M. S., J. A. Bollekens, and F. H. Ruddle.** 1995. A new vector for recombination-based cloning of large DNA fragments from yeast artificial chromosomes. Nucleic Acids Res. **23:**4850-4856.

18. **Brady, S. F., C. J. Chao, and J. Clardy.** 2004. Long-chain N-acyltyrosine synthases from environmental DNA. Appl. Environ. Microbiol. **70:**6865-6870.

19. **Brady, S. F., C. J. Chao, and J. Clardy.** 2002. New natural product families from an environmental DNA (eDNA) gene cluster. J. Am. Chem. Soc. **124:**9968-9969.

20. **Brady, S. F., C. J. Chao, J. Handelsman, and J. Clardy.** 2001. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. Org. Lett. **3:**1981-1984.

21. **Brady, S. F., and J. Clardy.** 2005. Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA. Angew. Chem. Int. Edit. **44:**7063-7065.

22. **Brady, S. F., and J. Clardy.** 2000. Long-chain N-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. J. Am. Chem. Soc. **122:**12903-12904.

23. **Brady, S. F., and J. Clardy.** 2005. N-acyl derivatives of arginine and tryptophan isolated from environmental DNA expressed in *Escherichia coli*. Org. Lett. **7:**3613-3616.

24. **Brady, S. F., and J. Clardy.** 2003. Synthesis of long-chain fatty acid enol esters isolated from an environmental DNA clone. Org. Lett. **5:**121-124.

25. **Brady, S. F., and J. Clardy.** 2005. Systematic investigation of the *Escherichia coli* metabolome for the biosynthetic origin of an isocyanide carbon atom. Angew. Chem. Int. Edit. **44:**7045-7048.

26. **Brown, G. D.** 1998. The biosynthesis of steroids and triterpenoids. Nat. Prod. Rep. **15:**653-696.

27. **Burke, P. V., D. C. Raitt, L. A. Allen, E. A. Kellogg, and R. O. Poyton.** 1997. Effects of oxygen concentration on the expression of cytochrome c and cytochrome c oxidase genes in yeast. J. Biol. Chem. **272:**14705-14712.

28. **Button, D. K., F. Schut, P. Quang, R. Martin, and B. R. Robertson.** 1993. Viability and isolation of marine-bacteria by dilution culture - theory, procedures, and initial results. Appl. Environ. Microbiol. **59:**881-891.

29. **Camilli, A., and B. L. Bassler.** 2006. Bacterial small-molecule signaling pathways. Science **311:**1113-1116.

30. **Cane, D. E.** 1985. Isoprenoid biosynthesis - stereochemistry of the cyclization of allylic pyrophosphates. Acc. Chem. Res. **18:**220-226.

31. **Cane, D. E.** 1980. The stereochemistry of allylic pyrophosphate metabolism. Tetrahedron **36:**1109-1159.

32. **Carreras, C. W., and C. Khosla.** 1998. Purification and in vitro reconstitution of the essential protein components of an aromatic polyketide synthase. Biochemistry **37:**2084-2088.

33. **Carreras, C. W., R. Pieper, and C. Khosla.** 1997. The chemistry and biology of fatty acid, polyketide, and nonribosomal peptide biosynthesis. Top. Curr. Chem. **188:**85-126.

34. **Chater, K. F.** 2006. Streptomyces inside-out: a new perspective on the bacteria that provide us with antibiotics. Philos. Trans. R. Soc. Lond. B Biol. Sci. **361:**761-768.

35. **Chin-A-Woeng, T.** 2000. Molecular basis of biocontrol of tomato foot and root rot by *Pseudomonas chloroaphis* strain PCL1391. University of Leiden, Leiden, The Netherlands.

36. **Choi, J. H., W. Lou, and A. Vancura.** 1998. A novel membrane-bound glutathione S-transferase functions in the stationary phase of the yeast *Saccharomyces cerevisiae*. J. Biol. Chem. **273:**29915-29922.

37. **Chugani, S. A., M. Whiteley, K. M. Lee, D. D'Argenio, C. Manoil, and E. P. Greenberg.** 2001. QscR, a modulator of quorum-sensing signal synthesis and virulence in *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. U S A **98:**2752-2757.

38. **Clarke, L., and J. Carbon.** 1976. Colony bank containing synthetic Col El hybrid plasmids representative of entire *Escherichia coli* genome. Cell **9:**91-99.

39. **Clarkson, C., D. Staerk, S. H. Hansen, and J. W. Jaroszewski.** 2005. Hyphenation of solid-phase extraction with liquid chromatography and nuclear magnetic resonance: Application of HPLC-DAD-SPE-NMR to identification of constituents of *Kanahia laniflora*. Anal. Chem. **77:**3547-3553.

40. **Connon, S. A., and S. J. Giovannoni.** 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Appl. Environ. Microbiol. **68:**3878-3885.

41. **Courtois, S., C. M. Cappellano, M. Ball, F. X. Francou, P. Normand, G. Helynck, A. Martinez, S. J. Kolvek, J. Hopke, M. S. Osburne, P. R. August, R. Nalin, M. Guerineau, P. Jeannin, P. Simonet, and J. L. Pernodet.** 2003. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl. Environ. Microbiol. **69:**49-55.

42. **Cueto, M., P. R. Jensen, and W. Fenical.** 2002. Aspergilloxide, a novel sesterterpene epoxide from a marine-derived fungus of the genus Aspergillus. Org. Lett. **4:**1583-1585.

43. **Datsenko, K. A., and B. L. Wanner.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. U S A **97:**6640-6645.

44. **Davis, E. M., and R. Croteau.** 2000. Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. Top. Curr. Chem. **209:**53-95.

45. **Davis, R. H., F. J. De Serres.** 1970. Genetic and microbiological research techniques for *Neurospora crassa*. Methods Enzymol. **17A:**79-143.

46. **Dean, F. B., S. Hosono, L. H. Fang, X. H. Wu, A. F. Faruqi, P. Bray-Ward, Z. Y. Sun, Q. L. Zong, Y. F. Du, J. Du, M. Driscoll, W. M. Song, S. F. Kingsmore, M. Egholm, and R. S. Lasken.** 2002. Comprehensive human genome amplification using multiple displacement amplification. Proc. Natl. Acad. Sci. U S A **99:**5261-5266.

47. **Dewar, M. J. S., and J. M. Ruiz.** 1987. Mechanism of the biosynthesis of squalene from farnesyl pyrophosphate. Tetrahedron **43:**2661-2674.

48. **Dewick, P. M.** 2001. Medicinal natural products: a biosynthetic approach, 2nd ed. Wiley, Chichester, West Sussex, England.

49. **Diggle, S. P., P. Cornelis, P. Williams, and M. Camara.** 2006. 4-Quinolone signalling in *Pseudomonas aeruginosa*: old molecules, new perspectives. Int. J. Med. Microbiol. **296:**83-91.

50. **Donadio, S., M. J. Staver, J. B. Mcalpine, S. J. Swanson, and L. Katz.** 1991. Modular organization of genes required for complex polyketide biosynthesis. Science **252:**675-679.

51. **EndoIchikawa, Y., H. Kohno, T. Furukawa, T. Ueda, Y. Ogawa, R. Tokunaga, and S. Taketani.** 1996. Requirement of multiple DNA-protein interactions for inducible expression of RNR3 gene in Saccharomyces cerevisiae in response to DNA damage. Biochemical and Biophysical Research Communications **222:**280-286.

52. **Felder, T., M. Knunbauer, E. Bogengruber, and P. Briza.** 2001. Identification of a sporulation-specific yeast chitinase. Yeast **18:**(S1)47.

53. **Finking, R., and M. A. Marahiel.** 2004. Biosynthesis of nonribosomal peptides. Annu. Rev. Microbiol. **58:**453-488.

54. **Fitch, W. M., and E. Margoliash.** 1967. Construction of phylogenetic trees. Science **155:**279-284.

55. **Floyd, M. M., J. Tang, M. Kane, and D. Emerson.** 2005. Captured diversity in a culture collection: case study of the geographic and habitat distributions of environmental isolates held at the American Type Culture Collection. Appl. Environ. Microbiol. **71:**2813-2823.

56. **Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen, and C. R. Woese.** 1980. The phylogeny of prokaryotes. Science **209:**457-463.

57. **Friedrich, M. W.** 2006. Stable-isotope probing of DNA: insights into the function of uncultivated microorganisms from isotopically labeled metagenomes. Curr. Opin. Biotechnol. **17:**59-66.

58. **Fujisaki, S., I. Takahashi, H. Hara, K. Horiuchi, T. Nishino, and Y. Nishimura.** 2005. Disruption of the structural gene for farnesyl diphosphate synthase in *Escherichia coli*. J. Biochem. **137:**395-400.

59. **Funayama, S., S. Eda, K. Komiyama, S. Omura, and T. Tokunaga.** 1989. Structure of phenazinomycin, a novel antitumor antibiotic. Tetrahedron Lett. **30:**3151-3154.

60. **Fuqua, C., S. C. Winans, and E. P. Greenberg.** 1996. Census and consensus in bacterial ecosystems: The LuxR-LuxI family of quorum-sensing transcriptional regulators. Annu. Rev. Microbiol. **50:**727-751.

61. **Fyrst, H., B. Oskouian, F. A. Kuypers, and J. D. Saba.** 1999. The PLB2 gene of *Saccharomyces cerevisiae* confers resistance to lysophosphatidylcholine and encodes a phospholipase B lysophospholipase. Biochemistry **38:**5864-5871.

62. **Gallagher, L. A., S. L. McKnight, M. S. Kuznetsova, E. C. Pesci, and C. Manoil.** 2002. Functions required for extracellular quinolone signaling by *Pseudomonas aeruginosa*. J. Bacteriol. **184:**6472-6480.

63. **Gillespie, D. E., S. F. Brady, A. D. Bettermann, N. P. Cianciotto, M. R. Liles, M. R. Rondon, J. Clardy, R. M. Goodman, and J. Handelsman.** 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Appl. Environ. Microbiol. **68:**4301-4306.

64. **Glansdorp, F. G., G. L. Thomas, J. J. K. Lee, J. M. Dutton, G. P. C. Salmond, M. Welch, and D. R. Spring.** 2004. Synthesis and stability of small molecule probes for *Pseudomonas aeruginosa* quorum sensing modulation. Org. Biomol. Chem. **2:**3329-3336.

65. **Godon, C., G. Lagniel, J. Lee, J. M. Buhler, S. Kieffer, R. Perrot, H. Boucherie, M. B. Toledano, and J. Labarre.** 1998. The H_2O_2 stimulon in *Saccharomyces cerevisiae*. J. Biol. Chem. **273:**22480-22489.

66. **Goldberg, S. M. D., J. Johnson, D. Busam, T. Feldblyum, S. Ferriera, R. Friedman, A. Halpern, H. Khouri, S. A. Kravitz, F. M. Lauro, K. Li, Y. H. Rogers, R. Strausberg, G. Sutton, L. Tallon, T. Thomas, E. Venter, M. Frazier, and J. C. Venter.** 2006. A Sanger/pyrosequencing hybrid approach tor the generation of high-

quality draft assemblies of marine microbial genomes. Proc. Natl. Acad. Sci. U S A **103:**11240-11245.

67. **Gonzalez, J. M., M. C. Portillo, and C. Saiz-Jimenez.** 2005. Multiple displacement amplification as a pre-polymerase chain reaction (pre-PCR) to process difficult to amplify samples and low copy number sequences from natural environments. Environ. Microbiol. **7:**1024-1028.

68. **Green, R. E., J. Krause, S. E. Ptak, A. W. Briggs, M. T. Ronan, J. F. Simons, L. Du, M. Egholm, J. M. Rothberg, M. Paunovic, and S. Paabo.** 2006. Analysis of one million base pairs of Neanderthal DNA. Nature **444:**330-336.

69. **Gregonis, D. E., and H. C. Rilling.** 1974. Stereochemistry of *trans*-phytoene synthesis - some observations on lycopersene as a carotene precursor and a mechanism for synthesis of *cis*-phytoene and *trans*-phytoene. Biochemistry **13:**1538-1542.

70. **Gross, H., V. O. Stockwell, M. D. Henkels, B. Nowak-Thompson, J. E. Loper, and W. H. Gerwick.** 2007. The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters. Chem. Biol. **14:**53-63.

71. **Grossart, H. P., A. Schlingloff, M. Bernhard, M. Simon, and T. Brinkhoff.** 2004. Antagonistic activity of bacteria isolated from organic aggregates of the German Wadden Sea. FEMS Microbiol. Ecol. **47:**387-396.

72. **Hallam, S. J., K. T. Konstantinidis, N. Putnam, C. Schleper, Y. Watanabe, J. Sugahara, C. Preston, J. de la Torre, P. M. Richardson, and E. F. DeLong.** 2006. Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. Proc. Natl. Acad. Sci. U S A **103:**18296-18301.

73. **Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman.** 1998. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. Chem. Biol. **5:**R245-R249.

74. **Hardie, D. C., T. R. Gregory, and P. D. N. Hebert.** 2002. From pixels to picograms: a beginners' guide to genome quantification by Feulgen image analysis densitometry. J. Histochem. Cytochem. **50:**735-749.

75. **Hartig, A., and H. Ruis.** 1986. Nucleotide sequence of the *Saccharomyces cerevisiae* cttl gene and deduced amino acid sequence of yeast catalase T. Eur. J. Biochem. **160:**487-490.

76. **Hassan, H. M., and I. Fridovich.** 1980. Mechanism of the antibiotic action pyocyanine. J. Bacteriol. **141:**156-163.

77. **Helmann, J. D., M. F. W. Wu, A. Gaballa, P. A. Kobel, M. M. Morshedi, P. Fawcett, and C. Paddon.** 2003. The global transcriptional response of Bacillus subtilis to peroxide stress is coordinated by three transcription factors. J. Bacteriol. **185:**243-253.

78. **Hernandez, M. E., and D. K. Newman.** 2001. Extracellular electron transfer. Cell. Mol. Life Sci. **58:**1562-1571.

79. **Higgins, V. J., N. Alic, G. W. Thorpe, M. Breitenbach, V. Larsson, and I. W. Dawes.** 2002. Phenotypic analysis of gene deletant strains for sensitivity to oxidative stress. Yeast **19:**203-214.

80. **Hoffman, R. A.** 2001. Standardization and quantitation in flow cytometry. Methods Cell. Bio. **63:**299-340.

81. **Hofreuter, D., J. Tsai, R. O. Watson, V. Novik, B. Altman, M. Benitez, C. Clark, C. Perbost, T. Jarvie, L. Du, and J. E. Galan.** 2006. Unique features of a highly pathogenic *Campylobacter jejuni* strain. Infec. Immun. **74:**4694-4707.

82. **Holden, M. T. G., S. R. Chhabra, R. de Nys, P. Stead, N. J. Bainton, P. J. Hill, M. Manefield, N. Kumar, M. Labatte, D. England, S. Rice, M. Givskov, G. P. C. Salmond, G. S. A. B. Stewart, B. W. Bycroft, S. A. Kjelleberg, and P. Williams.** 1999. Quorum-sensing cross talk: isolation and chemical characterization of cyclic dipeptides from *Pseudomonas aeruginosa* and other Gram-negative bacteria. Mol. Microbiol. **33:**1254-1266.

83. **Hopwood, D. A.** 2007. Streptomyces in nature and medicine. Oxford University Press, Oxford.

84. **Horak, C. E., N. M. Luscombe, J. A. Qian, P. Bertone, S. Piccirrillo, M. Gerstein, and M. Snyder.** 2002. Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. Genes Dev. **16:**3017-3033.

85. **Horinouchi, S.** 2007. Mining and polishing of the treasure trove in the bacterial genus Streptomyces. Biosci., Biotechnol., Biochem. **71:**283-299.

86. **Hu, J. Y., Y. Fan, Y. H. Lin, H. B. Zhang, S. L. Ong, N. Dong, J. L. Xu, W. J. Ng, and L. H. Zhang.** 2003. Microbial diversity and prevalence of virulent pathogens in biofilms developed in a water reclamation system. Res. Microbiol. **154:**623-629.

87. **Hurt, R. A., X. Y. Qiu, L. Y. Wu, Y. Roh, A. V. Palumbo, J. M. Tiedje, and J. H. Zhou.** 2001. Simultaneous recovery of RNA and DNA from soils and sediments. Appl. Environ. Microbiol. **67:**4495-4503.

88. **Ichinose, K., D. J. Bedford, M. J. Bibb, W. P. Revill, and D. A. Hopwood.** 1998. The granaticin biosynthetic gene cluster of *Streptomyces violaceoruber* Tu22: sequence analysis and expression in a heterologous host. Chem. Biol. **5:**647-659.

89. **Imhoff, J. F., and R. Stohr.** 2003. Sponge-associated bacteria: general overview and special aspects of bacteria associated with *Halichondria panicea*. Prog. Mol. Subcell. Bio. **37:**35-57.

90. **Johnston, J. S., M. D. Bennett, A. L. Rayburn, D. W. Galbraith, and H. J. Price.** 1999. Reference standards for determination of DNA content of plant nuclei. Am. J. Bot. **86:**609-613.

91. **Jordan, A., and P. Reichard.** 1998. Ribonucleotide reductases. Annu. Rev. Biochem. **67:**71-98.

92. **Karunanandaa, B., Q. G. Qi, M. Hao, S. R. Baszis, P. K. Jensen, Y. H. H. Wong, J. Jiang, M. Venkatramesh, K. J. Gruys, F. Moshiri, D. Post-Beittermiller, J. D. Weiss, and H. E. Valentin.** 2005. Metabolically engineered oilseed crops with enhanced seed tocopherol. Metab. Eng. **7:**384-400.

93. **Keller, M., and K. Zengler.** 2004. Tapping into microbial diversity. Nature Rev. Microbiol. **2:**141-150.

94. **Kelly, G. T., C. M. Liu, R. Smith, R. S. Coleman, and C. M. H. Watanabe.** 2006. Cellular effects induced by the antitumor agent azinomycin B. Chem. Biol. **13:**485-492.

95. **Keyer, K., and J. A. Imlay.** 1996. Superoxide accelerates DNA damage by elevating free-iron levels. Proc. Natl. Acad. Sci. U S A **93:**13635-13640.

96. **Kieser, T., M. J. Bibb, M. J. Buttner, K. F. Chater, and D. A. Hopwood.** 2000. Practical streptomyces genetics. Crowes, Norwich, England.

97. **Kilby, N. J., M. R. Snaith, and J. A. H. Murray.** 1993. Site-specific recombinases - tools for genome engineering. Trends Genet. **9:**413-421.

98. **Kim, E. J., S. Angell, and C. M. Watanabe.** 2007. Estimating P-coverage of biosynthetic pathways in DNA libraries and screening by genetic selection: biotin biosynthesis in the marine microorganism *Chromohalobacter*. Mol. Biosyst. **In press.**

99. **Kim, U. J., H. Shizuya, P. J. Dejong, B. Birren, and M. I. Simon.** 1992. Stable propagation of cosmid sized human DNA inserts in an F-factor based vector. Nucleic Acids Res. **20:**1083-1085.

100. **Kingston, D. G. I.** 2001. Taxol, a molecule for all seasons. Chem. Commun. **2001:**867-880.

101. **Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al.** 2001. Initial sequencing and analysis of the human genome. Nature **409:**860-921.

102. **Lander, E. S., and M. S. Waterman.** 1988. Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics **2:**231-239.

103. **Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace.** 1985. Rapid-determination of 16S ribosomal-RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. U S A **82:**6955-6959.

104. **Lanoil, B. D., L. M. Ciuffetti, and S. J. Giovannoni.** 1996. The marine bacterium *Pseudoalteromonas haloplanktis* has a complex genome structure composed of two separate genetic units. Genome Res. **6:**1160-1169.

105. **Larionov, V., N. Kouprina, J. Graves, X. N. Chen, J. R. Korenberg, and M. A. Resnick.** 1996. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. Proc. Natl. Acad. Sci. U S A **93:**491-496.

106. **Laursen, J. B., and J. Nielsen.** 2004. Phenazine natural products: biosynthesis, synthetic analogues, and biological activity. Chem. Rev. **104:**1663-1685.

107. **Lautru, S., R. J. Deeth, L. M. Bailey, and G. L. Challis.** 2005. Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. Nat. Chem. Biol. **1:**265-269.

108. **Law, D. T. S., and J. Segall.** 1988. The Sps100 gene of *Saccharomyces cerevisiae* is activated late in the sporulation process and contributes to spore wall maturation. Mol. Cell. Biol. **8:**912-922.

109. **Lederberg, J.** 1947. Gene recombination and linked segregations in *Escherichia coli*. Genetics **32:**505-525.

110. **Ledgham, F., I. Ventre, C. Soscia, M. Foglino, J. N. Sturgis, and A. Lazdunski.** 2003. Interactions of the quorum sensing regulator QscR: interaction with itself and the other regulators of *Pseudomonas aeruginosa* LasR and RhlR. Mol. Microbiol. **48:**199-210.

111. **Lesburg, C. A., J. M. Caruthers, C. M. Paschall, and D. W. Christianson.** 1998. Managing and manipulating carbocations in biology: terpenoid cyclase structure and mechanism. Current Opinion in Structural Biology **8:**695-703.

112. **Li, H., A. K. Singh, L. M. McIntyre, and L. A. Sherman.** 2004. Differential gene expression in response to hydrogen peroxide and the putative PerR regulon of *Synechocystis* sp strain PCC 6803. J. Bacteriol. **186:**3331-3345.

113. **Li, Y. G., M. Wexler, D. J. Richardson, P. L. Bond, and A. W. B. Johnston.** 2005. Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned trp genes. Environ. Microbiol. **7:**1927-1936.

114. **Liang, P. H., T. P. Ko, and A. H. J. Wang.** 2002. Structure, mechanism and function of prenyltransferases. Eur. J. Biochem. **269:**3339-3354.

115. **Long, R. A., and F. Azam.** 2001. Antagonistic interactions among marine pelagic bacteria. Appl. Environ. Microbiol. **67:**4975-4983.

116. **Lucier, T. S., and R. R. Brubaker.** 1992. Determination of genome size, macrorestriction pattern polymorphism, and nonpigmentation-specific deletion in *Yersinia pestis* by pulsed-field gel-electrophoresis. J. Bacteriol. **174:**2078-2086.

117. **Lussier, M., A. M. White, J. Sheraton, T. diPaolo, J. Treadwell, S. B. Southard, C. I. Horenstein, J. ChenWeiner, A. F. J. Ram, J. C. Kapteyn, T. W. Roemer, D. H. Vo, D. C. Bondoc, J. Hall, W. W. Zhong, A. M. Sdicu, J. Davies, F. M. Klis, P. W. Robbins, and H. Bussey.** 1997. Large scale identification of genes involved in cell surface biosynthesis and architecture in *Saccharomyces cerevisiae*. Genetics **147:**435-450.

118. **Makrides, S. C.** 1996. Strategies for achieving high-level expression of genes in *Escherichia coli*. Microbiol. Rev. **60:**512-538.

119. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. T. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. G. Yu, R. F. Begley, and J. M. Rothberg.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376-380.

120. **Martinez, A., S. J. Kolvek, C. L. T. Yip, J. Hopke, K. A. Brown, I. A. MacNeil, and M. S. Osburne.** 2004. Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. Appl. Environ. Microbiol. **70:**2452-2463.

121. **Mavrodi, D. V., R. F. Bonsall, S. M. Delaney, M. J. Soule, G. Phillips, and L. S. Thomashow.** 2001. Functional analysis of genes for biosynthesis of pyocyanin and phenazine-1-carboxamide from *Pseudomonas aeruginosa* PAO1. J. Bacteriol. **183:**6454-6465.

122. **Mavrodi, D. V., V. N. Ksenzenko, R. F. Bonsall, R. J. Cook, A. M. Boronin, and L. S. Thomashow.** 1998. A seven-gene locus for synthesis is of phenazine-1-carboxylic acid by *Pseudomonas fluorescens* 2-79. J. Bacteriol. **180:**2541-2548.

123. **McDaniel, R., S. Ebert-Khosla, D. A. Hopwood, and C. Khosla.** 1993. Engineered biosynthesis of novel polyketides. Science **262:**1546-1550.

124. **Meister, A., and M. E. Anderson.** 1983. Glutathione. Annu. Rev. Biochem. **52:**711-760.

125. **Miljanich, G. P.** 2004. Ziconotide: neuronal calcium channel blocker for treating severe chronic pain. Curr. Med. Chem. **11:**3029-3040.

126. **Misener, S., and S. A. Krawetz.** 2000. Bioinformatics methods and protocols. Humana Press, Totowa, NJ.

127. **Mitra, R. D., J. Shendure, J. Olejnik, Edyta-Krzymanska-Olejnik, and G. M. Church.** 2003. Fluorescent in situ sequencing on polymerase colonies. Anal. Biochem. **320:**55-65.

128. **Moradas-Ferreira, P., V. Costa, P. Piper, and W. Mager.** 1996. The molecular defences against reactive oxygen species in yeast. Mol. Microbiol. **19:**651-658.

129. **Moskovitz, J., B. S. Berlett, J. M. Poston, and E. R. Stadtman.** 1997. The yeast peptide-methionine sulfoxide reductase functions as an antioxidant in vivo. Proc. Natl. Acad. Sci. U S A **94:**9585-9589.

130. **Muller, M.** 2002. Pyocyanin induces oxidative stress in human endothelial cells and modulates the glutathione redox cycle. Free Radic. Biol. Med. **33:**1527-1533.

131. **Murphy, K. C.** 1998. Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. J. Bacteriol. **180:**2063-2071.

132. **Murphy, K. C., K. G. Campellone, and A. R. Poteete.** 2000. PCR-mediated gene replacement in *Escherichia coli*. Gene **246:**321-330.

133. **Muyrers, J. P. P., Y. M. Zhang, and A. F. Stewart.** 2000. ET-Cloning: think recombination first. Genet. Eng. **22:**77-98.

134. **Muyrers, J. P. P., Y. M. Zhang, G. Testa, and A. F. Stewart.** 1999. Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res. **27:**1555-1557.

135. **Nagy, A.** 2000. Cre recombinase: the universal reagent for genome tailoring. Genesis **26:**99-109.

136. **Newman, D. J., and G. M. Cragg.** 2007. Natural products as sources of new drugs over the last 25 years. J. Nat. Prod. **70:**461-477.

137. **Newman, D. J., and G. M. Cragg.** 2006. Natural products from marine invertebrates and microbes as modulators of antitumor targets. Curr. Drug Targets **7:**279-304.

138. **Newman, D. J., G. M. Cragg, and K. M. Snader.** 2003. Natural products as sources of new drugs over the period 1981-2002. J. Nat. Prod. **66:**1022-1037.

139. **Noskov, V. N., N. Kouprina, S. H. Leem, I. Ouspenski, J. C. Barrett, and V. Larionov.** 2003. A general cloning system to selectively isolate any eukaryotic or prokaryotic genomic region in yeast. BMC Genom. **4(16)**.

140. **O'Malley, Y. Q., K. J. Reszka, G. T. Rasmussen, M. Y. Abdalla, G. M. Denning, and B. E. Britigan.** 2003. The *Pseudomonas* secretory product pyocyanin inhibits

catalase activity in human lung epithelial cells. American Journal of Physiology - Lung Cellular and Molecular Physiology **285:**L1077-L1086.

141. **O'Malley, Y. Q., K. J. Reszka, D. R. Spitz, G. M. Denning, and B. E. Britigan.** 2004. *Pseudomonas aeruginosa* pyocyanin directly oxidizes glutathione and decreases its levels in airway epithelial cells. American Journal of Physiology - Lung Cellular and Molecular Physiology **287:**94-103.

142. **Oh, J. D., H. Kling-Backhed, M. Giannakis, J. Xu, R. S. Fulton, L. A. Fulton, H. S. Cordum, C. Y. Wang, G. Elliott, J. Edwards, E. R. Mardis, L. G. Engstrand, and J. I. Gordon.** 2006. The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression. Proc. Natl. Acad. Sci. U S A **103:**9999-10004.

143. **Oldenburg, K. R., K. T. Vo, S. Michaelis, and C. Paddon.** 1997. Recombination-mediated PCR-directed plasmid construction in vivo in yeast. Nucleic Acids Res. **25:**451-452.

144. **Olsen, G. J., D. J. Lane, S. J. Giovannoni, N. R. Pace, and D. A. Stahl.** 1986. Microbial ecology and evolution - a ribosomal-RNA approach. Annu. Rev. Microbiol. **40:**337-365.

145. **Osoegawa, K., P. Y. Woon, B. H. Zhao, E. Frengen, M. Tateno, J. J. Catanese, and P. J. de Jong.** 1998. An improved approach for construction of bacterial artificial chromosome libraries. Genomics **52:**1-8.

146. **Paez, J. G., M. Lin, R. Beroukhim, J. C. Lee, X. J. Zhao, D. J. Richter, S. Gabriel, P. Herman, H. Sasaki, D. Altshuler, C. Li, M. Meyerson, and W. R. Sellers.** 2004. Genome coverage and sequence fidelity of phi 29 polymerase-based multiple strand displacement whole genome amplification. Nucleic Acids Res. **32(e71)**.

147. **Palma, M., D. DeLuca, S. Worgall, and L. E. N. Quadri.** 2004. Transcriptome analysis of the response of *Pseudomonas aeruginosa* to hydrogen peroxide. J. Bacteriol. **186:**248-252.

148. **Park, S., and J. A. Imlay.** 2003. High levels of intracellular cysteine promote oxidative DNA damage by driving the Fenton reaction. J. Bacteriol. **185:**1942-1950.

149. **Perlova, O., J. Fu, S. Kuhlmann, D. Krug, A. F. Stewart, Y. M. Zhang, and R. Muller.** 2006. Reconstitution of the myxothiazol biosynthetic gene cluster by Red/ET recombination and heterologous expression in *Myxococcus xanthus*. Appl. Environ. Microbiol. **72:**7485-7494.

150. **Pesci, E. C., J. B. J. Milbank, J. P. Pearson, S. McKnight, A. S. Kende, E. P. Greenberg, and B. H. Iglewski.** 1999. Quinolone signaling in the cell-to-cell communication system of *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. U S A **96:**11229-11234.

151. **Pesci, E. C., J. P. Pearson, P. C. Seed, and B. H. Iglewski.** 1997. Regulation of las and rhl quorum sensing in *Pseudomonas aeruginosa*. J. Bacteriol. **179:**3127-3132.

152. **Peterson, D. G., J. P. Tomkins, D. A. Frisch, R. A. Wing, A. H. Patterson.** 2000. Construction of plant bacterial artificial chromosomal libraries: an illustrated guide. J. Agric. Gen. **5:**1-100.

153. **Piel, J.** 2006. Bacterial symbionts: prospects for the sustainable production of invertebrate-derived pharmaceuticals. Curr. Med. Chem. **13:**39-50.

154. **Piel, J., D. Butzke, N. Fusetani, D. Q. Hui, M. Platzer, G. P. Wen, and S. Matsunaga.** 2005. Exploring the chemistry of uncultivated bacterial symbionts: antitumor polyketides of the pederin family. J. Nat. Prod. **68:**472-479.

155. **Piel, J., D. Q. Hui, N. Fusetani, and S. Matsunaga.** 2004. Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. Environ. Microbiol. **6:**921-927.

156. **Pierson, L. S., T. Gaffney, S. Lam, and F. C. Gong.** 1995. Molecular analysis of genes encoding phenazine biosynthesis in the biological control bacterium *Pseudomonas aureofaciens* 30-84. FEMS Microbiol. Lett. **134:**299-307.

157. **Poinar, H. N., C. Schwarz, J. Qi, B. Shapiro, R. D. E. MacPhee, B. Buigues, A. Tikhonov, D. H. Huson, L. P. Tomsho, A. Auch, M. Rampp, W. Miller, and S. C. Schuster.** 2006. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science **311:**392-394.

158. **Porwollik, S., J. Frye, L. D. Florea, F. Blackmer, and M. McClelland.** 2003. A non-redundant microarray of genes for two related bacteria. Nucleic Acids Res. **31:**1869-1876.

159. **Poteete, A. R.** 2001. What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. FEMS Microbiol. Lett. **201:**9-14.

160. **Ran, H., D. J. Hassett, and G. W. Lau.** 2003. Human targets of *Pseudomonas aeruginosa* pyocyanin. Proc. Natl. Acad. Sci. U S A **100:**14315-14320.

161. **Raymond, C. K., T. A. Pownder, and S. L. Sexson.** 1999. General method for plasmid construction using homologous recombination. Biotechniques **26:**134-141.

162. **Raymond, C. K., E. H. Sims, A. Kas, D. H. Spencer, T. V. Kutyavin, R. G. Ivey, Y. Zhou, R. Kaul, J. B. Clendenning, and M. V. Olson.** 2002. Genetic variation at the O-antigen biosynthetic locus in *Pseudomonas aeruginosa*. J. Bacteriol. **184:**3614-3622.

163. **Raymond, C. K., E. H. Sims, and M. V. Olson.** 2002. Linker-mediated recombinational subcloning of large DNA fragments using yeast. Genome Res. **12:**190-197.

164. **Reading, N. C., and V. Sperandio.** 2006. Quorum sensing: the many languages of bacteria. FEMS Microbiol. Lett. **254:**1-11.

165. **Roach, J. C., C. Boysen, K. Wang, and L. Hood.** 1995. Pairwise end sequencing: a unified approach to genomic mapping and sequencing. Genomics **26:**345-53.

166. **Roh, C., F. Villatte, B. G. Kim, and R. D. Schmid.** 2005. "In-gel patch electrophoresis": A new method for environmental DNA purification. Electrophoresis **26:**3055-3061.

167. **Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman.** 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. **66:**2541-2547.

168. **Rowinsky, E. K., and R. C. Donehower.** 1995. Paclitaxel (taxol). The New England journal of medicine **332:**1004-1014.

169. **Sambrook, J., and D. W. Russell.** 2001. Molecular cloning : a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

170. **Samoilenko, I. I., E. I. Vasilyeva, I. B. Pavlova, and M. A. Tumanian.** 1983. The mechanisms of bactericidal action of hydrogen peroxide. Zh. Mikrob. Epid. Immun. **12:**30-33.

171. **Sandmann, G.** 1994. Carotenoid biosynthesis in microorganisms and plants. Eur. J. Biochem. **223:**7-24.

172. **Santi, D. V., M. A. Siani, B. Julien, D. Kupfer, and B. Roe.** 2000. An approach for obtaining perfect hybridization probes for unknown polyketide synthase genes: a search for the epothilone gene cluster. Gene **247:**97-102.

173. **Sarker, S. D., Z. Latif, and A. I. Gray.** 2006. Natural products isolation, 2nd ed. Humana Press, Totowa, NJ.

174. **Schirmer, A., R. Gadkari, C. D. Reeves, F. Ibrahim, E. F. DeLong, and C. R. Hutchinson.** 2005. Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. Appl. Environ. Microbiol. **71:**4840-4849.

175. **Schloss, P. D., and J. Handelsman.** 2004. Status of the microbial census. Microbiol. Mol. Biol. Rev. **68:**686-691.

176. **Schmidt, E. W., J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. G. Haygood, and J. Ravel.** 2005. Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of Lissoclinum patella. Proc. Natl. Acad. Sci. U S A **102:**7315-7320.

177. **Schmidt, T. M., E. F. Delong, and N. R. Pace.** 1991. Analysis of a marine picoplankton community by 16s ribosomal-RNA gene cloning and sequencing. J. Bacteriol. **173:**4371-4378.

178. **Scholz-Schroeder, B. K., J. D. Soule, and D. C. Gross.** 2003. The sypA, sypB and sypC synthetase genes encode twenty-two modules involved in the nonribosomal peptide synthesis of syringopeptin by *Pseudomonas syringae* pv. *syringae* B301D. Mol. Plant-Microbe Inter. **16:**271-280.

179. **Schuster, M., C. P. Lostroh, T. Ogi, and E. P. Greenberg.** 2003. Identification, timing, and signal specificity of *Pseudomonas aeruginosa* quorum-controlled genes: a transcriptome analysis. J. Bacteriol. **185:**2066-2079.

180. **Schwarzer, D., R. Finking, and M. A. Marahiel.** 2003. Nonribosomal peptides: from genes to products. Nat. Prod. Rep. **20:**275-287.

181. **Schweikert, C., A. Liszkay, and P. Schopfer.** 2002. Polysaccharide degradation by Fenton reaction- or peroxidase-generated hydroxyl radicals in isolated plant cell walls. Phytochem. **61:**31-35.

182. **Seed, B.** 1982. Theoretical study of the fraction of a long chain DNA that can be incorporated in a recombinant DNA partial digest library. Biopolymers **21:**1793-1810.

183. **Seed, B., R. C. Parker, and N. Davidson.** 1982. Representation of DNA sequences in recombinant DNA libraries prepared by restriction enzyme partial digestion. Gene **19:**201-209.

184. **Shashikant, C. S., J. L. Carr, J. Bhargava, K. L. Bentley, and F. H. Ruddle.** 1998. Recombinogenic targeting: a new approach to genomic analysis - a review. Gene **223:**9-20.

185. **Shen, B.** 2003. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr. Opin. Chem. Biol. **7:**285-295.

186. **Shendure, J., R. D. Mitra, C. Varma, and G. M. Church.** 2004. Advanced sequencing technologies: methods and goals. Nat. Rev. Genet. **5:**335-344.

187. **Shendure, J., G. J. Porreca, N. B. Reppas, X. X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra, and G. M. Church.** 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. Science **309:**1728-1732.

188. **Sheng, Y., V. Mancino, and B. Birren.** 1995. Transformation of *Escherichia coli* with large DNA molecules by electroporation. Nucleic Acids Res. **23:**1990-1996.

189. **Smalla, K., N. Cresswell, L. C. Mendoncahagler, A. Wolters, and J. D. Vanelsas.** 1993. Rapid DNA extraction protocol from soil for polymerase chain reaction-mediated amplification. J. Appl. Bacteriol. **74:**78-85.

190. **Sorensen, H. P., and K. K. Mortensen.** 2005. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J. Biotechnol. **115:**113-128.

191. **Stahl, D. A., D. J. Lane, G. J. Olsen, and N. R. Pace.** 1985. Characterization of a Yellowstone hot-spring microbial community by 5S ribosomal-RNS sequences. Appl. Environ. Microbiol. **49:**1379-1384.

192. **Staunton, J., and K. J. Weissman.** 2001. Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. **18:**380-416.

193. **Suga, H., and K. M. Smith.** 2003. Molecular mechanisms of bacterial quorum sensing as a new drug target. Curr. Opin. Chem. Biol. **7:**586-591.

194. **Takano, H., S. Obitsu, T. Beppu, and K. Ueda.** 2005. Light-induced carotenogenesis in *Streptomyces coelicolor* A3(2): identification of an extracytoplasmic function sigma factor that directs photodependent transcription of the carotenoid biosynthesis gene cluster. J. Bacteriol. **187:**1825-1832.

195. **Torsvik, V., L. Ovreas, and T. F. Thingstad.** 2002. Prokaryotic diversity - magnitude, dynamics, and controlling factors. Science **296:**1064-1066.

196. **Tringe, S. G., C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P. Hugenholtz, and E.**

M. Rubin. 2005. Comparative metagenomics of microbial communities. Science **308:**554-557.

197. **Uchiyama, T., T. Abe, T. Ikemura, and K. Watanabe.** 2005. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nat. Biotechnol. **23:**88-93.

198. **Van Eenennaam, A. L., K. Lincoln, T. P. Durrett, H. E. Valentin, C. K. Shewmaker, G. M. Thorne, J. Jiang, S. R. Baszis, C. K. Levering, E. D. Aasen, M. Hao, J. C. Stein, S. R. Norris, and R. L. Last.** 2003. Engineering vitamin E content: From Arabidopsis mutant to soy oil. Plant Cell **15:**3007-3019.

199. **Van Lanen, S. G., and B. Shen.** 2006. Microbial genomics for the improvement of natural product discovery. Current Opin. Microbiol. **9:**252-260.

200. **Veiga, M., F. Gutierrez, M. Lopez, and C. Fernandez.** 1992. Computer assisted theoretical calculation of the probability of a gene being intact in a DNA library. J. Microbiol. Methods. **16:**23-32.

201. **Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. Q. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. H. Zhang, G. L. G. Miklos, C. Nelson, S. Broder, A. G. Clark, C. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. M. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. M. Ge, F. C. Gong, Z. P. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. X. Ke, K. A. Ketchum, Z. W. Lai, Y. D. Lei, Z. Y. Li, J. Y. Li, Y. Liang, X. Y. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. X. Shue, J. T. Sun, Z. Y. Wang, A. H. Wang, X. Wang, J. Wang, M. H. Wei, R. Wides, C. L. Xiao, C. H. Yan, et al.** 2001. The sequence of the human genome. Science **291:**1304-1351.

202. **Venter, J. C., M. D. Adams, G. G. Sutton, A. R. Kerlavage, H. O. Smith, and M. Hunkapiller.** 1998. Shotgun sequencing of the human genome. Science **280:**1540-1542.

203. **Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Y. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith.** 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304:**66-74.

204. **Wade, D. S., M. W. Calfee, E. R. Rocha, E. A. Ling, E. Engstrom, J. P. Coleman, and E. C. Pesci.** 2005. Regulation of *Pseudomonas* quinolone signal synthesis in *Pseudomonas aeruginosa*. J. Bacteriol. **187:**4372-4380.

205. **Wagner, V. E., D. Bushnell, L. Passador, A. I. Brooks, and B. H. Iglewski.** 2003. Microarray analysis of *Pseudomonas aeruginosa* quorum-sensing regulons: effects of growth phase and environment. J. Bacteriol. **185:**2080-2095.

206. **Wang, C. X., D. J. Meek, P. Panchal, N. Boruvka, F. S. Archibald, B. T. Driscoll, and T. C. Charles.** 2006. Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. Appl. Environ. Microbiol. **72:**384-391.

207. **Wang, G. Y. S., E. Graziani, B. Waters, W. B. Pan, X. Li, J. McDermott, G. Meurer, G. Saxena, R. J. Andersen, and J. Davies.** 2000. Novel natural products from soil DNA libraries in a streptomycete host. Org. Lett. **2:**2401-2404.

208. **Watanabe, C. M. H., L. Supekova, and P. G. Schultz.** 2002. Transcriptional effects of the potent enediyne anti-cancer agent calicheamicin gamma. Chem. Biol. **9:**245-251.

209. **Wendl, M. C., M. A. Marra, L. W. Hillier, A. T. Chinwalla, R. K. Wilson, and R. H. Waterston.** 2001. Theories and applications for sequencing randomly selected clones. Genome Res. **11:**274-280.

210. **Wenzel, S. C., F. Gross, Y. M. Zhang, J. Fu, A. F. Stewart, and R. Muller.** 2005. Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via red/ET recombineering. Chem. Biol. **12:**349-356.

211. **Wexler, M., P. L. Bond, D. J. Richardson, and A. W. B. Johnston.** 2005. A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. Environ. Microbiol. **7:**1917-1926.

212. **Wheeler, G. L., and C. M. Grant.** 2004. Regulation of redox homeostasis in the yeast *Saccharomyces cerevisiae*. Physiol. Plant. **120:**12-20.

213. **Whiteley, M., K. M. Lee, and E. P. Greenberg.** 1999. Identification of genes controlled by quorum sensing in *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. U S A **96:**13904-13909.

214. **Williams, D. H., and B. Bardsley.** 1999. The vancomycin group of antibiotics and the fight against resistant bacteria. Angew. Chem. Int. Edit. **38:**1173-1193.

215. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221-271.

216. **Woese, C. R., and G. E. Fox.** 1977. Phylogenetic structure of prokaryotic domain - primary kingdoms. Proc. Natl. Acad. Sci. U S A **74:**5088-5090.

217. **Woese, C. R., O. Kandler, and M. L. Wheelis.** 1990. Towards a natural system of organisms - proposal for the domains archaea, bacteria, and eucarya. Proc. Natl. Acad. Sci. U S A **87:**4576-4579.

218. **Yu, D. G., H. M. Ellis, E. C. Lee, N. A. Jenkins, N. G. Copeland, and D. L. Court.** 2000. An efficient recombination system for chromosome engineering in *Escherichia coli*. Proc. Natl. Acad. Sci. U S A **97:**5978-5983.

219. **Yu, J. J., P. K. Chang, K. C. Ehrlich, J. W. Cary, D. Bhatnagar, T. E. Cleveland, G. A. Payne, J. E. Linz, C. P. Woloshuk, and J. W. Bennett.** 2004. Clustered pathway genes in aflatoxin biosynthesis. Appl. Environ. Microbiol. **70:**1253-1262.

220. **Zazopoulos, E., K. X. Huang, A. Staffa, W. Liu, B. O. Bachmann, K. Nonaka, J. Ahlert, J. S. Thorson, B. Shen, and C. M. Farnet.** 2003. A genomics-guided approach for discovering and expressing cryptic metabolic pathways. Nat. Biotechnol. **21:**187-190.

221. **Zeller, T., O. V. Moskvin, K. Y. Li, G. Klug, and M. Gomelsky.** 2005. Transcriptome and physiological responses to hydrogen peroxide of the facultatively phototrophic bacterium *Rhodobacter sphaeroides*. J. Bacteriol. **187:**7232-7242.

222. **Zhang, H. B.** 2000. Construction and manipulation of large-insert bacterial clone libraries-[manual]. Texas A&M University, College Station, TX.

223. **Zhang, K., A. C. Martiny, N. B. Reppas, K. W. Barry, J. Malek, S. W. Chisholm, and G. M. Church.** 2006. Sequencing genomes from single cells by polymerase cloning. Nat. Biotechnol. **24:**680-686.

224. **Zhang, L. X., R. An, J. P. Wang, N. Sun, S. Zhang, J. C. Hu, and J. Kuai.** 2005. Exploring novel bioactive compounds from marine microbes. Current Opin. Microbiol. **8:**276-281.

225. **Zhang, Y. M., F. Buchholz, J. P. P. Muyrers, and A. F. Stewart.** 1998. A new logic for DNA engineering using recombination in *Escherichia coli*. Nat. Genet. **20:**123-128.

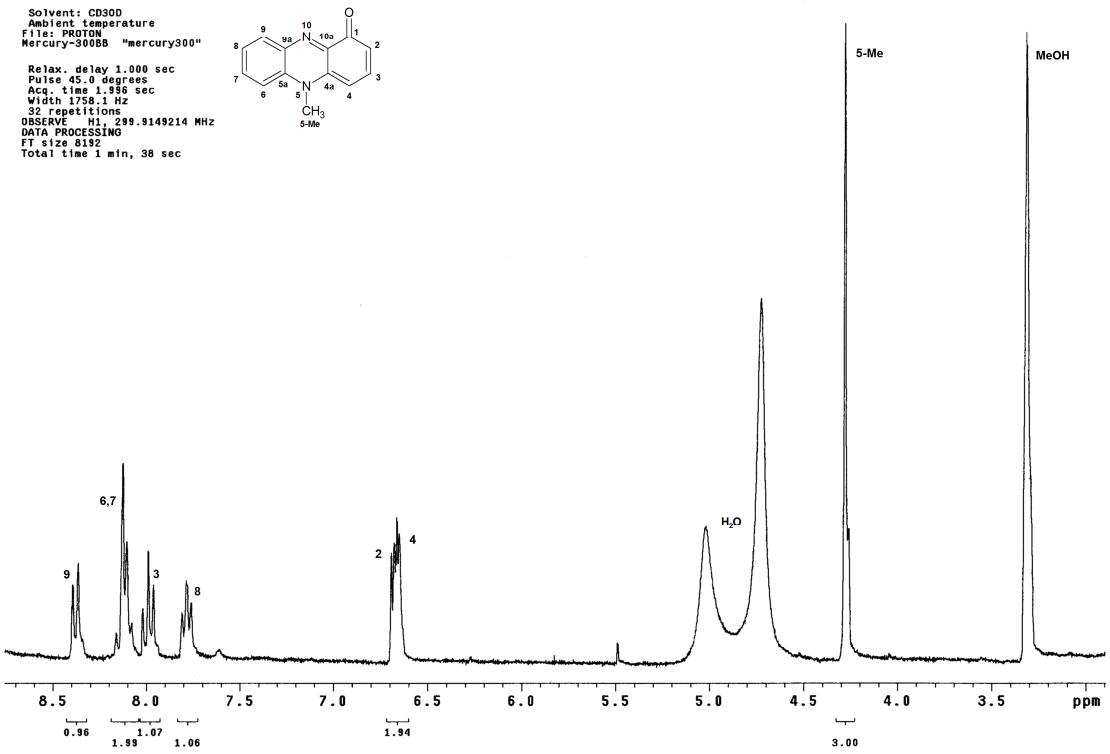
226. **Zhang, Y. M., J. P. P. Muyrers, G. Testa, and A. F. Stewart.** 2000. DNA cloning by homologous recombination in *Escherichia coli*. Nat. Biotechnol. **18:**1314-1317.

227. **Zheng, M., X. Wang, L. J. Templeton, D. R. Smulski, R. A. LaRossa, and G. Storz.** 2001. DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide. J. Bacteriol. **183:**4562-4570.

228. **Zhou, Z., and S. J. Elledge.** 1992. Isolation of crt mutants constitutive for transcription of the DNA damage inducible gene RNR3 in *Saccharomyces cerevisiae*. Genetics **131:**851-866.

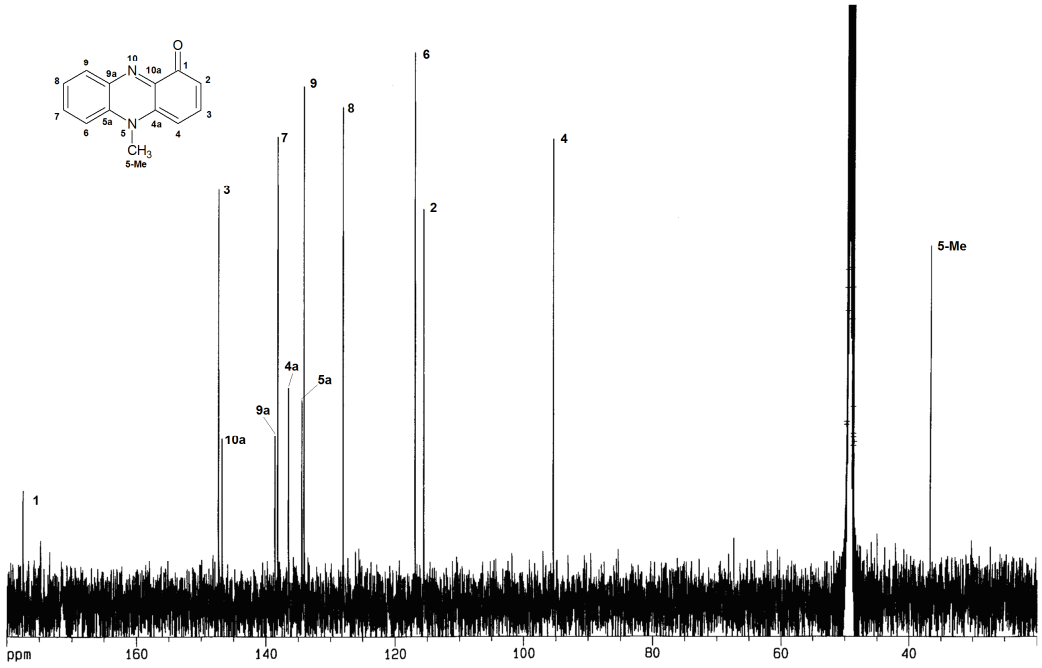
229. **Zuckerkandl, E., and L. Pauling.** 1965. Molecules as documents of evolutionary history. J. Theor. Biol. **8:**357-366.

APPENDIX A
SUPPLEMENTARY DATA FOR CHAPTER II



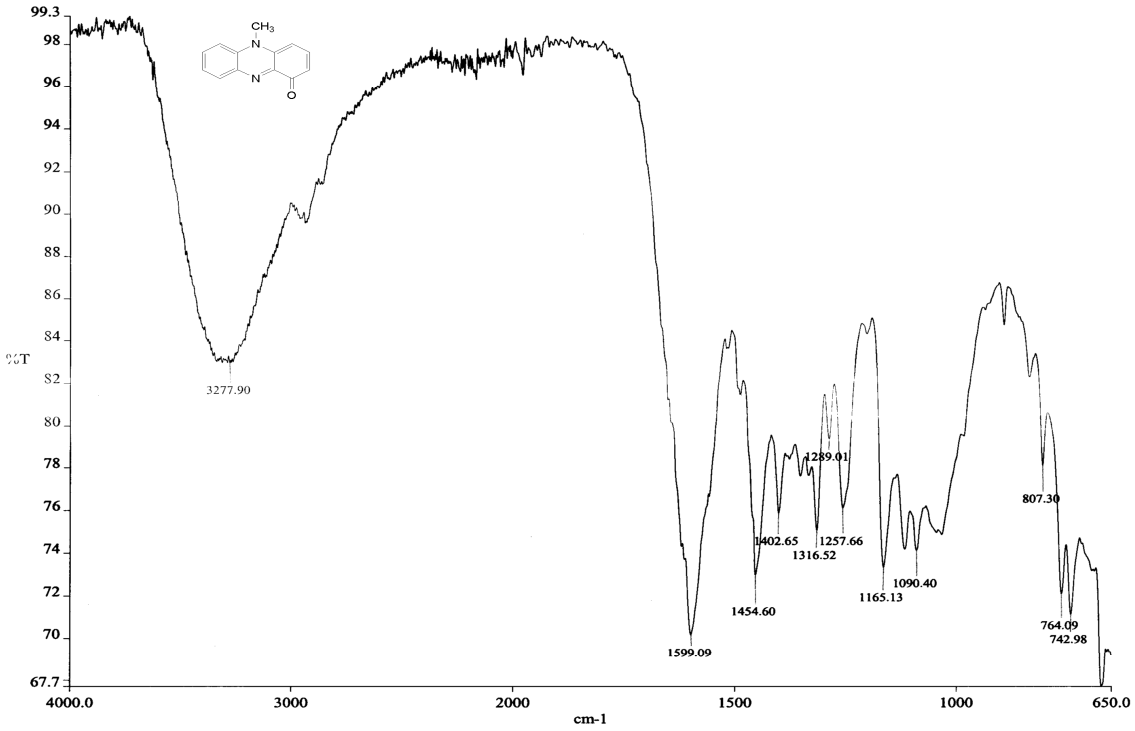
^{1}H NMR of pyocyanin.

^{1}H-NMR: (300 MHz, methanol-d_4) 8.38 (d, J = 8.4 Hz, 1H), 8.12 (m, 2H), 7.99 (t, J = 8.4Hz, 1H), 7.79 (t, J = 7.0 Hz, 1H), 6.67 (dd, J = 8.4, 4.1 Hz, 2H), 3.95 (s, 3H)

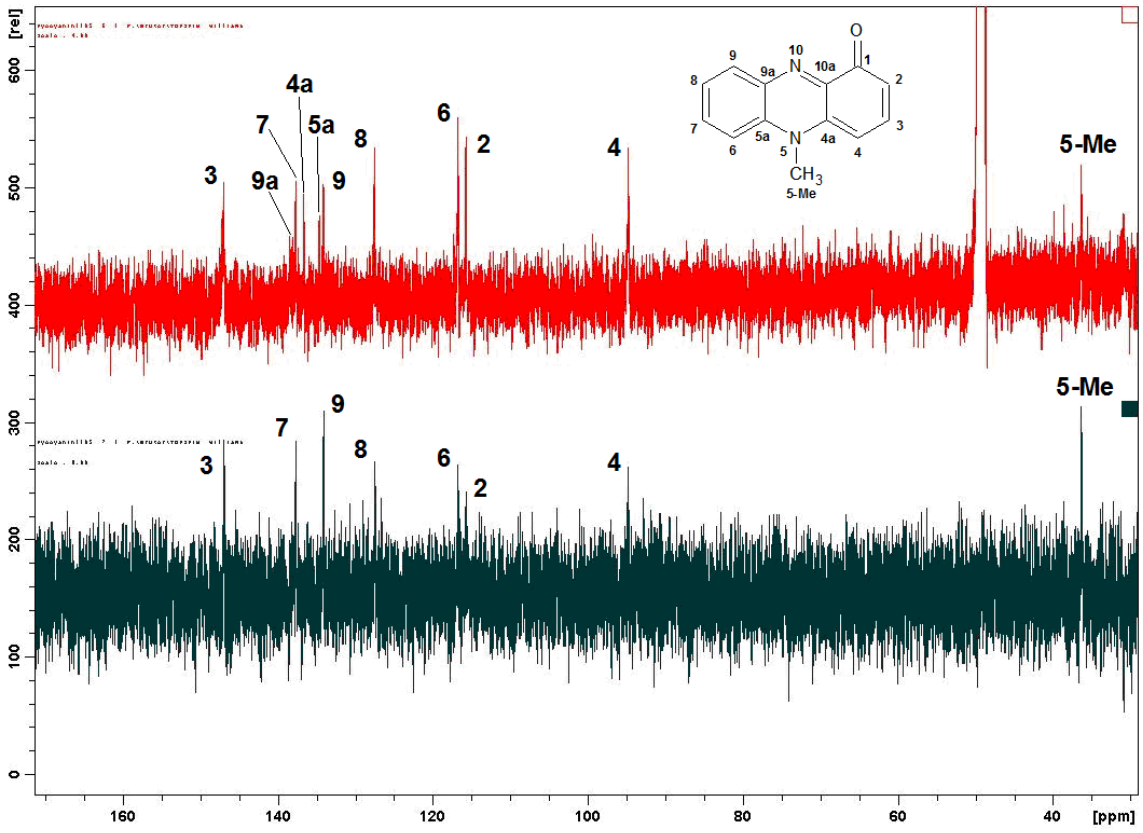


^{13}C NMR spectrum of pyocyanin.

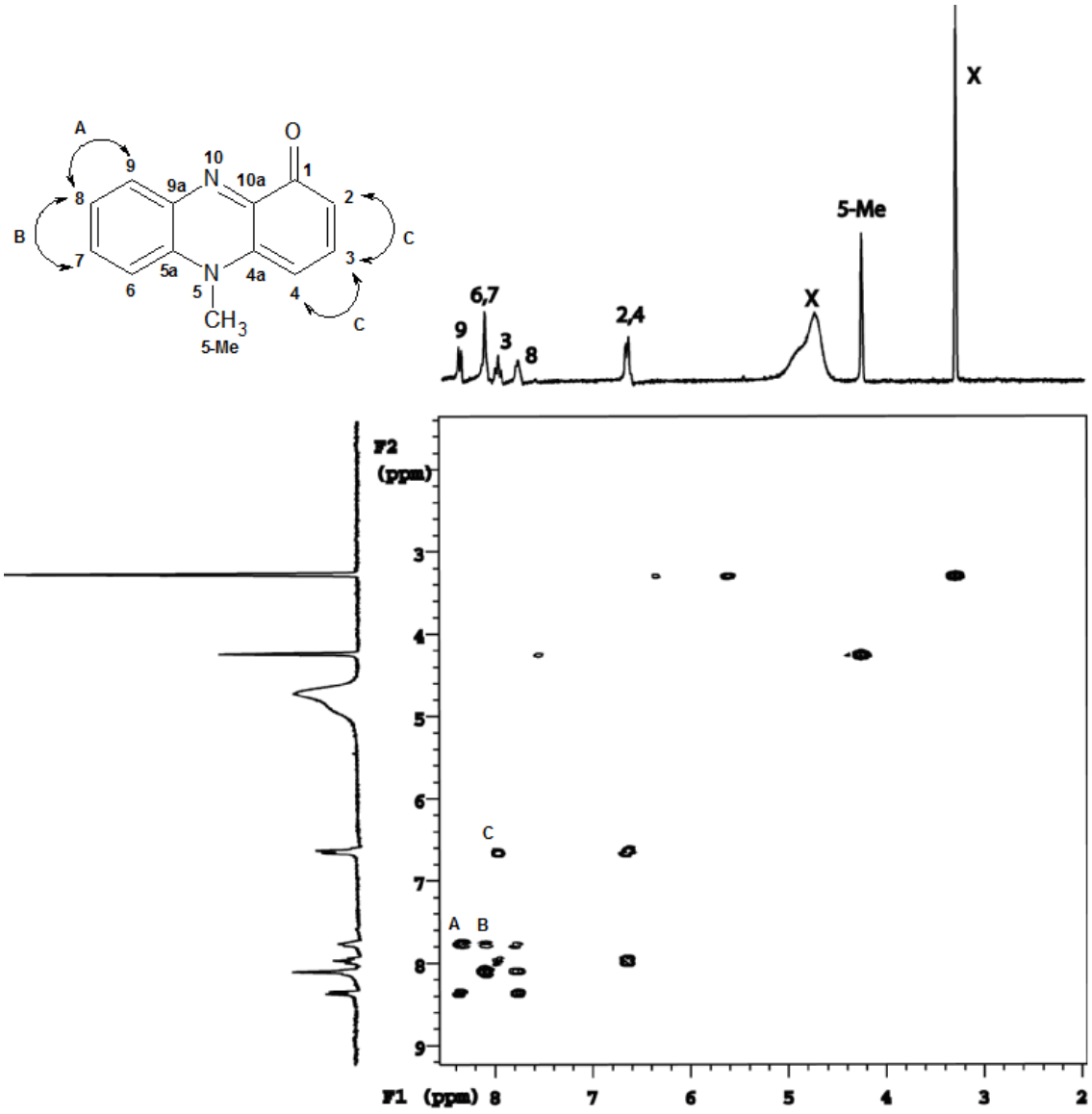
^{13}C NMR: (75 MHz, Methanol-d_4) 177.5, 147.3, 146.7, 138.5, 138.1, 136.4, 134.4, 134.0, 128.0, 116.8, 115.5, 95.3, 36.7



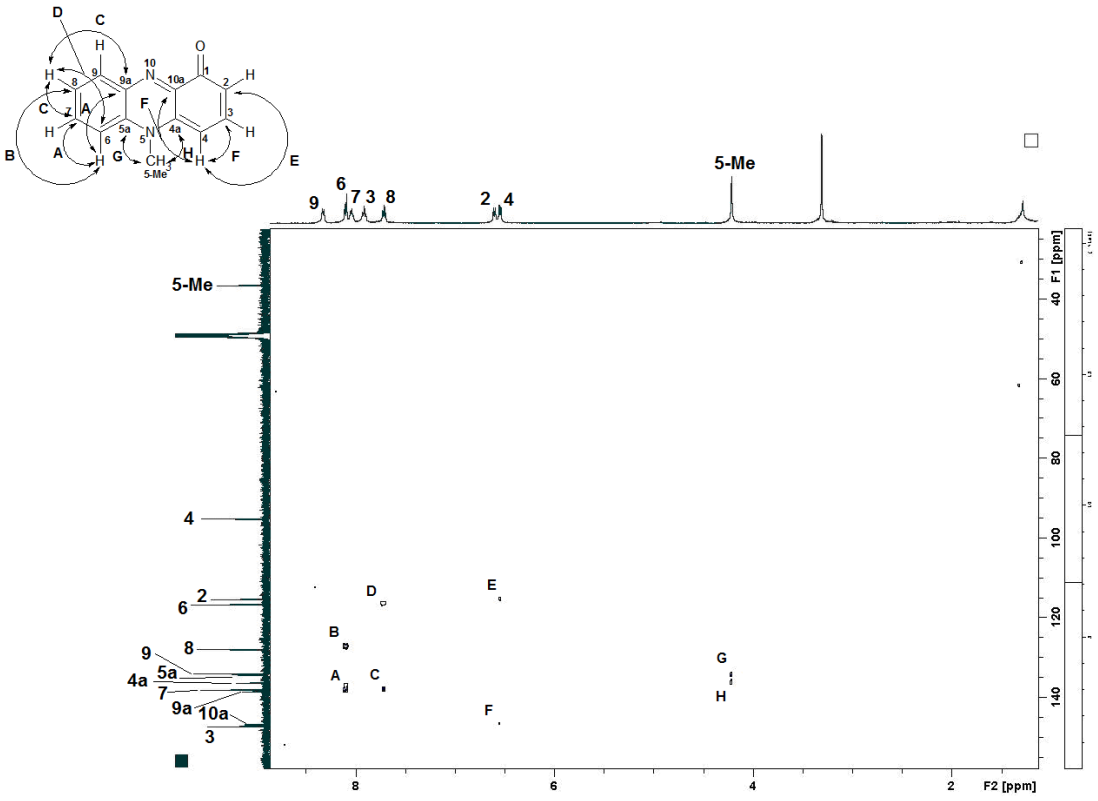
ATR-FTIR (Attenuated Total Reflection, Fourier-Transform Infrared) spectrum of pyocyanin.



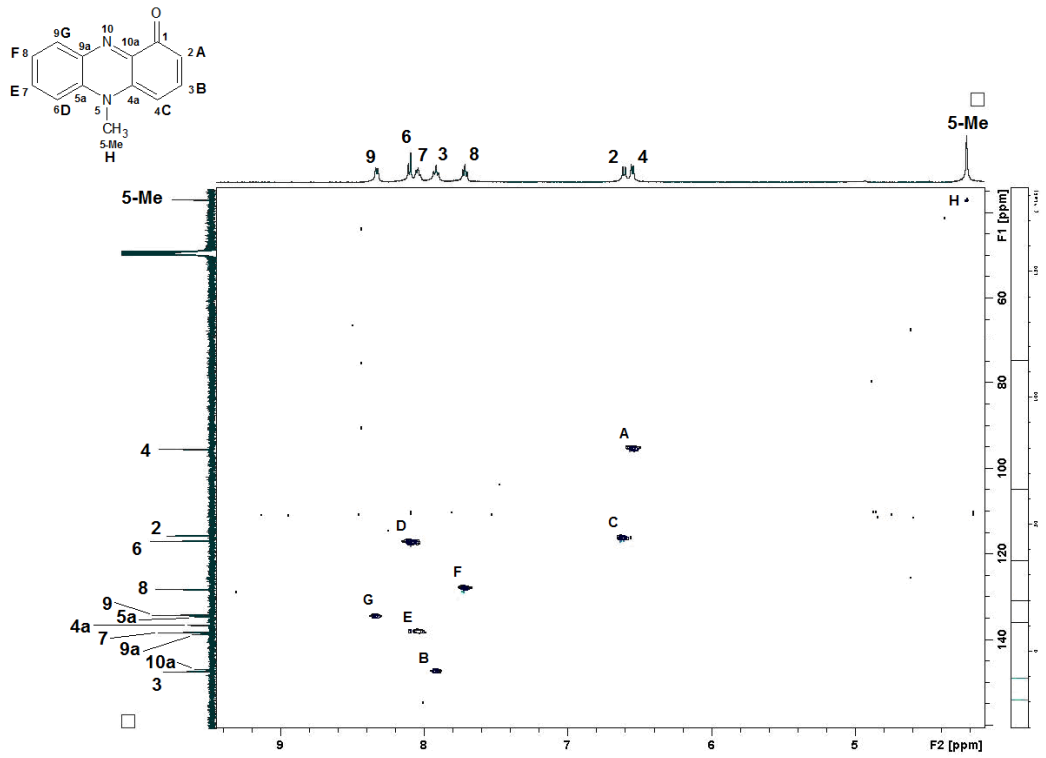
DEPT spectrum of pyocyanin. Upper spectrum is a normal ^{13}C spectrum, lower spectrum is a DEPT-45 spectrum, showing only carbons which have an attached hydrogen.



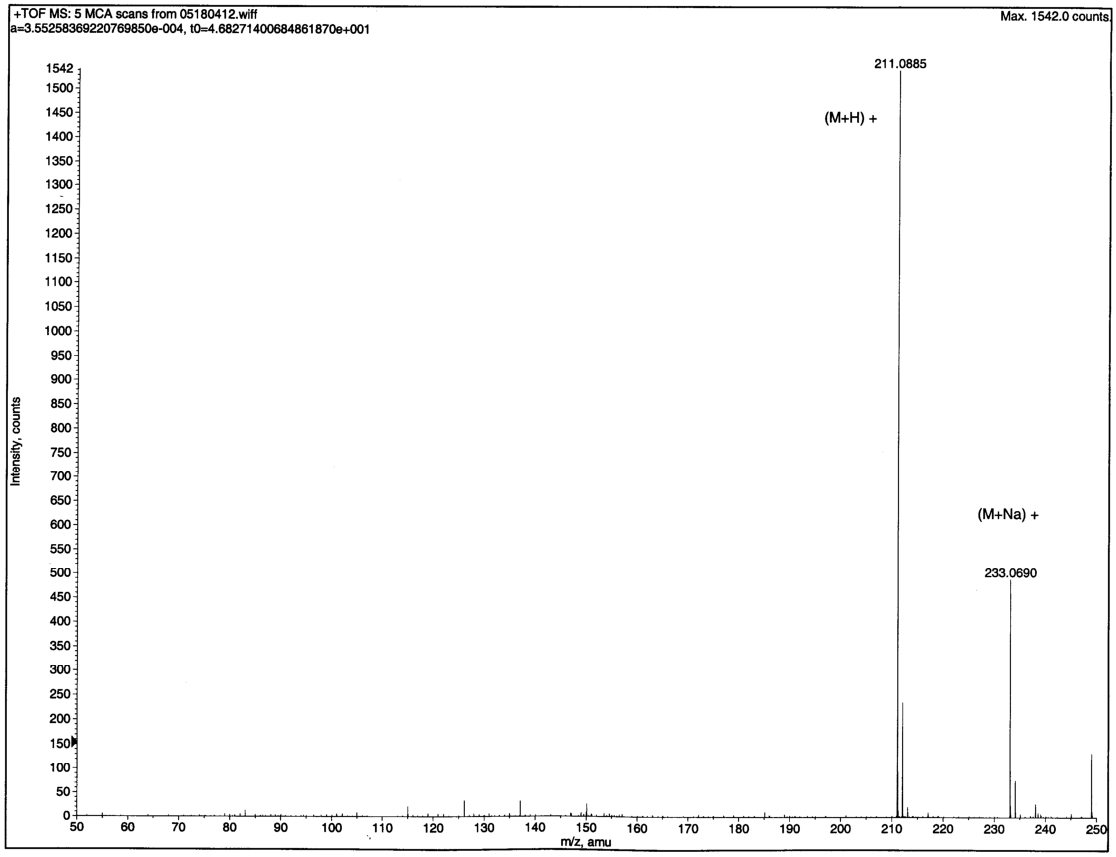
^{1}H – ^{1}H COSY (Correlation Spectroscopy) spectrum of pyocyanin. Cross peaks indicate H-H coupling.



HMBC (Heteronuclear Multi-Bond Correlation) spectrum of pyocyanin. Cross peaks indicate C-H coupling through multi-bond networks.



HSQC (Heteronuclear Single-Quantum Correlation) spectrum of pyocyanin. Cross peaks indicate C-H coupling through a single bond.



ESI/TOF MS (Electrospray Ionization/Time-of-Flight Mass Spectrometry) spectrum of pyocyanin. Peaks for protonated and sodiated molecular ions are indicated.

Candidate Molecular Formula	Deviation from calc. mass (in ppm)
$C_{13}H_{11}O_1N_2$	0.04897
$C_8H_{11}O_3N_4$	19.1061

Within 20 ppm of the calculated mass, only two molecular formulas are possible. X-ray data and ^{13}C-NMR data supports a molecular formula of $C_{13}H_{11}O_1N_2$.

Sequence Analysis of the Promoter Regions of Pseudomonas aeruginosa Phenazine Biosynthesis Regions

These sequence alignments were performed against the NCBI database genomic sequence for *Pseudomonas aeruginosa* PAO1. Sequence was obtained by PCR amplification of the desired region, TA cloning, and Sanger sequencing as described in the Methods section. Results indicate minimal genetic differences between our strain and the model strain, suggesting that regulation is more complex than simple promoter mutation. The las box and QscR locus are known to be involved in quorum sensing regulation, but do not appear to be altered in this strain.

Promoter Region Sequence Data for PHZ1 Pathway
gi|110227054|gb|AE004091.2| Pseudomonas aeruginosa PAO1, complete genome
Length=6264404

Features flanking this part of subject sequence:
54 bp at 5' side: probable phenazine-specific methyltransferase
7 bp at 3' side: probable phenazine biosynthesis protein

Score = 1239 bits (625), Expect = 0.0
Identities = 634/637 (99%), Gaps = 0/637 (0%)
Strand=Plus/Minus

```
Query 1        CCTCCGAGAGGGCTCTCCAGGTATGCCGGAGAAACTTTTCCCTCGCGGCATCGGTTATCC 60
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct 4713789  CCTCCGAGAGGGCTCTCCAGGTATGCCGGAGAAACTTTTCCCTCGCGGCATCGGTTATTC 4713730

Query 61       GCTGGCGCGCAGGAGTTCGTCGCTCGGTAAGGGAGCGATATTCGCTGGGCGCATTGTTTA 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 4713729  GCTGGCGCGCAGGAGTTCGTCGCTCGGTAAGGGAGCGATATTCGCTGGGCGCATTGTTTA 4713670

Query 121      TACGTAGGGTGAAATTTCGCTTTACATATGTTCGCCTGTCGTTTCCAATTTTCGAAAGGC 180
               |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Sbjct 4713669 TACGTAGGGTGAAATTTCGCGTTACATATGTTCGCCTGTCGTTTCCAATTTTCGAAAGGC 4713610

Query 181      AATAGGAGTTTCATCCCGGGTTTCTTTTGGAAACCTGTCGGTAATGGATTCGCAATGCTT 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 4713609  AATAGGAGTTTCATCCCGGGTTTCTTTTGGAAACCTGTCGGTAATGGATTCGCAATGCTT 4713550

Query 241      TCGTTAAGGTGCGACAGACGAGGTCGCGAAGACTTTCAGCGTCATTCCGTGAAGTGTTTC 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 4713549  TCGTTAAGGTGCGACAGACGAGGTCGCGAAGACTTTCAGCGTCATTCCGTGAAGTGTTTC 4713490

Query 301      AAATAGCCAGCATCCCTGGTCGCTGATGTGGATTGCATAAAACACAGAACGCTCGTACCG 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 4713489  AAATAGCCAGCATCCCTGGTCGCTGATGTGGATTGCATAAAACACAGAACGCTCGTACCG 4713430

Query 361      GCTCAACTACAAGATCTGGTAGGTGCCAGACAGGGTATGCGGGATTGCTAAGCTGATGCT 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 4713429  GCTCAACTACAAGATCTGGTAGGTGCCAGACAGGGTATGCGGGATTGCTAAGCTGATGCT 4713370

Query 421      TCCTGCAATGCCGGAGGTTGTAGCCAAGTTGTAATTTTATTCTTTCGGTACGCAGGAAAA 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 4713369  TCCTGCAATGCCGGAGGTTGTAGCCAAGTTGTAATTTTATTCTTTCGGTACGCAGGAAAA 4713310

Query 481      GGCTCTGGAACAGGCAGTTGGAAAGTTCCAGGGTTGTTTCCGCAACGAGATCCAGGGAGC 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 4713309  GGCTCTGGAACAGGCAGTTGGAAAGTTCCAGGGTTGTTTCCGCAACGAGATCCAGGGAGC 4713250
```

```
Query 541      GCTGATCCGGGGGATGAGCGCATTTTTTCTCAAAGTTCCGAATGAACCGGGGTGTTCTGC 600
               |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Sbjct 4713249  GCTGATCCGGGGGATGAGCGTATTTTTTCTCAAAGTTCCGAATGAACCGGGGTGTTCTGC 4713190

Query 601      GGTATTTCTCGCCGAGCGGGCAAGAAGTGCGAAATAG 637
               |||||||||||||||||||||||||||||||||||||
Sbjct 4713189  GGTATTTCTCGCCGAGCGGGCAAGAAGTGCGAAATAG 4713153
```

Red = las box

Sequence Data for Promotor Region of PHZ2 Pathway and QscR

gi|110227054|gb|AE004091.2| Pseudomonas aeruginosa PAO1, complete genome
Length=6264404

Features in this part of subject sequence:
hypothetical protein
quorum-sensing control repressor

Score = 3943 bits (1989), Expect = 0.0
Identities = 2035/2042 (99%), Gaps = 1/2042 (0%)
Strand=Plus/Plus

```
Query  7        GGAACCAACTGTTCCAGCATGAAGATGAATAGCGCCACTCCGATCAAGTAATAAATCATA  66
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2068670  GGAACCAACTGTTCCAGCATGAAGATGAATAGCGCCACTCCGATCAAGTAATAAATCATA  2068729

Query  67       TTTATCTCCTCTTCCTTAGGGGCGGACGCCATGCGGAACTTTCATCGATACGAAAGGCCA  126
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2068730  TTTATCTCCTCTTCCTTAGGGGCGGACGCCATGCGGAACTTTCATCGATACGAAAGGCCA  2068789

Query  127      GCGTCCTGTTTATTGTCTGAAGCACTTCTCGGCAGGCTGCGGTTATCGCTTCGCTATATT  186
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2068790  GCGTCCTGTTTATTGTCTGAAGCACTTCTCGGCAGGCTGCGGTTATCGCTTCGCTATATT  2068849

Query  187      CGGAAGGCGATCTCGGTCAGGCGTGGATAGCTTGTCCATGGTCGAATGGAGCGCTCTTTT  246
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2068850  CGGAAGGCGATCTCGGTCAGGCGTGGATAGCTTGTCCATGGTCGAATGGAGCGCTCTTTT  2068909

Query  247      TAATTGCCTCAGAATGCGCTTCGTGAAAATCGGCACGGACAATGAAATGCCTGGTCGAAT  306
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2068910  TAATTGCCTCAGAATGCGCTTCGTGAAAATCGGCACGGACAATGAAATGCCTGGTCGAAT  2068969

Query  307      TAAATTGTCACAGCCCGGCAGGGACAACCTGCCCTTCCGGGCAGGTATCTGCGGAGAGAT  366
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2068970  TAAATTGTCACAGCCCGGCAGGGACAACCTGCCCTTCCGGGCAGGTATCTGCGGAGAGAT  2069029

Query  367      AggggggAAGTATGTAGTTGCCATTTCTGGTCCCCGGGGAGCCTCGTCGACGAGGCCCG  426
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069030  AGGGGGGGAAGTATGTAGTTGCCATTTCTGGTCCCCGGGGAGCCTCGTCGACGAGGCCCG  2069089

Query  427      CCATGGGCCAAGGTTTGTTGTCGGGAGGCTCTCCCGAACGACGATGGAGCGTGCGAGAAG  486
                |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Sbjct  2069090  CCATGGGCCAAGGTTTGTTGTCGGGAGGCTCTCCCGA-CGACGATGGAGCGTGCGAGAAG  2069148

Query  487      AACAATGAGAAAGACCGCCGTGAGGCCCATCGGAGAGCCGTTCTACGGTTTCCGCAAAGA  546
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069149  AACAATGAGAAAGACCGCCGTGAGGCCCATCGGAGAGCCGTTCTACGGTTTCCGCAAAGA  2069208

Query  547      TCCGGGGCGCCGTCCCCTCCAGCCCAGCGCAGTTCCTGCGCGGTGCCTCGTGTCCGTGCT  606
                |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct  2069209  TCCGGGGCGCCGTCCCCTCCAGCCCAGCGCAGTTCCTGCGCGGCGCCTCGTGTCCGTGCT  2069268
```

```
Query  607      CATCGAGAAGTTCTCTTCAGCCTCGTTTCGTCGTCGCCCGGCGGGCGGCGAATGGGCTCG  666
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069269  CATCGAGAAGTTCTCTTCAGCCTCGTTTCGTCGTCGCCCGGCGGGCGGCGAATGGGCTCG  2069328

Query  667      ACCTCGTCCGGAACACCCGCACAGGGCCGGTGGCGATATGTACTTCCAGGTCCGGCTTGA  726
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069329  ACCTCGTCCGGAACACCCGCACAGGGCCGGTGGCGATATGTACTTCCAGGTCCGGCTTGA  2069388

Query  727      TAAAGGGAATTGTCATGAGTGGATAAGACGGAAACAAAAAAGAATAAAAACGCTGAAGAA  786
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069389  TAAAGGGAATTGTCATGAGTGGATAAGACGGAAACAAAAAAGAATAAAAACGCTGAAGAA  2069448

Query  787      CCGAATCCTGCCGGGATCGATTGTTGACTGGTGAAGCTGGCATGCATGATGAGAGAGAGG  846
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069449  CCGAATCCTGCCGGGATCGATTGTTGACTGGTGAAGCTGGCATGCATGATGAGAGAGAGG  2069508

Query  847      GATATCTCGAGATTTTGTCAAGAATAACAACCGAGGAAGAGTTCTTCTCCCTGGTTCTCG  906
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069509  GATATCTCGAGATTTTGTCAAGAATAACAACCGAGGAAGAGTTCTTCTCCCTGGTTCTCG  2069568

Query  907      AGATATGCGGTAATTATGGATTCGAATTCTTTTCATTCGGTGCGCGGGCGCCTTTCCCGC  966
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069569  AGATATGCGGTAATTATGGATTCGAATTCTTTTCATTCGGTGCGCGGGCGCCTTTCCCGC  2069628

Query  967      TGACCGCACCTAAATATCATTTCCTGTCCAATTACCCAGGGGAATGGAAAAGCAGATATA  1026
                 ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069629  TGACCGCGCCTAAATATCATTTCCTGTCCAATTACCCAGGGGAATGGAAAAGCAGATATA  2069688

Query  1027     TCTCCGAAGACTACACATCCATCGACCCGATCGTGCGCCATGGTCTCCTGGAATACACCC  1086
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069689  TCTCCGAAGACTACACATCCATCGACCCGATCGTGCGCCATGGTCTCCTGGAATACACCC  2069748

Query  1087     CGCTGATCTGGAATGGCGAAGACTTCCAGGAGAACCGTTTCTTCTGGGAGGAAGCGCTGC  1146
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069749  CGCTGATCTGGAATGGCGAAGACTTCCAGGAGAACCGTTTCTTCTGGGAGGAAGCGCTGC  2069808

Query  1147     ATCACGGCATCCGTCACGGCTGGTCGATCCCGGTCCGCGGCAAGTACGGGCTGATCAGCA  1206
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069809  ATCACGGCATCCGTCACGGCTGGTCGATCCCGGTCCGCGGCAAGTACGGGCTGATCAGCA  2069868

Query  1207     TGCTGTCCCTGGTGCGTTCCAGCGAGAGCATCGCCGCCACGGAAATCCTGGAGAAGGAAT  1266
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069869  TGCTGTCCCTGGTGCGTTCCAGCGAGAGCATCGCCGCCACGGAAATCCTGGAGAAGGAAT  2069928

Query  1267     CCTTCCTGCTCTGGATCACCAGCATGCTGCAGGCTACCTTCGGCGACCTGCTGGCGCCGC  1326
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069929  CCTTCCTGCTCTGGATCACCAGCATGCTGCAGGCTACCTTCGGCGACCTGCTGGCGCCGC  2069988

Query  1327     GCATCGTCCCGGAAAGCAATGTGCGCCTGACCGCCAGGGAAACCGAGATGCTCAAGTGGA  1386
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2069989  GCATCGTCCCGGAAAGCAATGTGCGCCTGACCGCCAGGGAAACCGAGATGCTCAAGTGGA  2070048

Query  1387     CCGCGGTGGGCAAGACCTACGGCGAGATCGGCCTGATCCTGTCGATCGACCAGCGCACGG  1446
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2070049  CCGCGGTGGGCAAGACCTACGGCGAGATCGGCCTGATCCTGTCGATCGACCAGCGCACGG  2070108

Query  1447     TGAAGTTCCATATCGTCAATGCGATGCGCAAGCTCAACTCCAGCAACAAGGCGGAGGTCA  1506
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Sbjct  2070109  TGAAGTTCCATATCGTCAATGCGATGCGCAAGCTCAACTCCAGCAACAAGGCGGAGGCCA  2070168

Query  1507     CCATGAAGGCTTACGCCATCGGCCTGCTCAACTGAATCGACGCCTCGTCGCCTAGCGAGG  1566
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2070169  CCATGAAGGCTTACGCCATCGGCCTGCTCAACTGAATCGACGCCTCGTCGCCTAGCGAGG  2070228

Query  1567     CCGCCGCGCAAGCGTCCGGCCATTCACCGAATGGCCGGATAGCGTTTGCGCCGGTCGCCT  1626
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2070229  CCGCCGCGCAAGCGTCCGGCCATTCACCGAATGGCCGGATAGCGTTTGCGCCGGTCGCCT  2070288
```

```
Query  1627     GAGCGCAGCCTTCCCACCGGCAGCGTTTCCCCGCTGCCCCCTTCGCCATTGCGCCCGTCC  1686
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2070289  GAGCGCAGCCTTCCCACCGGCAGCGTTTCCCCGCTGCCCCCTTCGCCATTGCGCCCGTCC  2070348

Query  1687     TCCTGTTGTCCGGCACGCTAGTGCAACTTTCCGGGCGCTTGGCAAACCGGCCAAAGAATA  1746
                |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Sbjct  2070349  TCCTGTTGTCCGGCACGCTAGTGCAACTTTCCGGGCGCCTGGCAAACCGGCCAAAGAATA  2070408

Query  1747     GAACGGAATCGATGCCCCACACCTGTAATTTTTAAGGGGTTATGGCTATTGCAAAAAAGC  1806
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2070409  GAACGGAATCGATGCCCCACACCTGTAATTTTTAAGGGGTTATGGCTATTGCAAAAAAGC  2070468

Query  1807     GTTTATAAGTTTGTCCCCTGTCAAATCTGGTTACAACTGGGTTTCAGGCGAAACATTCGG  1866
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2070469  GTTTATAAGTTTGTCCCCTGTCAAATCTGGTTACAACTGGGTTTCAGGCGAAACATTCGG  2070528

Query  1867     TCATGGCAACTCGGCATTAGTTGAAACTTTGGAGACGCTCCGAAGCGGGCAACTTTTGCC  1926
                ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2070529  TCATGGCAATTCGGCATTAGTTGAAACTTTGGAGACGCTCCGAAGCGGGCAACTTTTGCC  2070588

Query  1927     CGGAAAAAGCTTCACGGCAATTTCTCCGGCCTGTCATCCCGATGTCTTCTTTCCGGTATG  1986
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2070589  CGGAAAAAGCTTCACGGCAATTTCTCCGGCCTGTCATCCCGATGTCTTCTTTCCGGTATG  2070648

Query  1987     GATGCCAGTCGATTCGAACTGGCGGAGATCCGCACCATGCGAGAGTACCAACGGTTGAAA  2046
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Sbjct  2070649  GATGCCAGTCGATTCGAACTGGCGGAGATTCGCACCATGCGAGAGTACCAACGGTTGAAA  2070708

Query  2047     GG  2048
                ||
Sbjct  2070709  GG  2070710
```

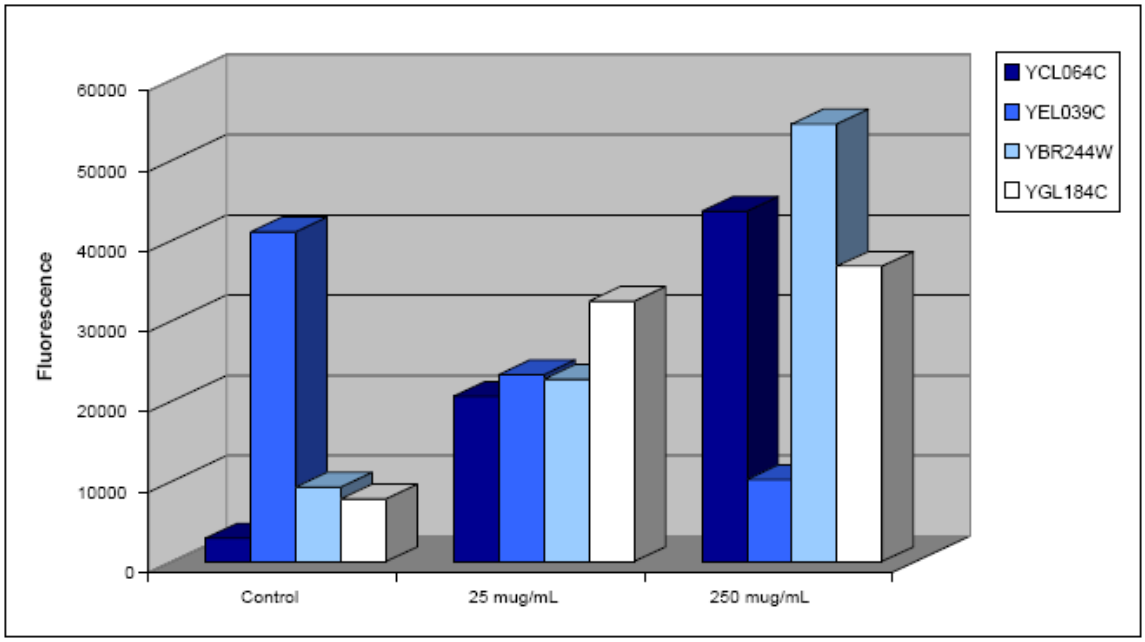
Green = unknown gene; locus PA1897

Blue = QscR

Red = phz2A

RT-PCR Confirmation of Selected Gene Changes

This graph shows the results of an RT-PCR analysis of several genes whose transcription level was altered by exposure to pyocyanin, as detected by the transcriptional array profiling experiments. These data confirm the results of the transcriptional profiling experiment for these genes and support the accuracy of the results for the remainder of the transcriptome.

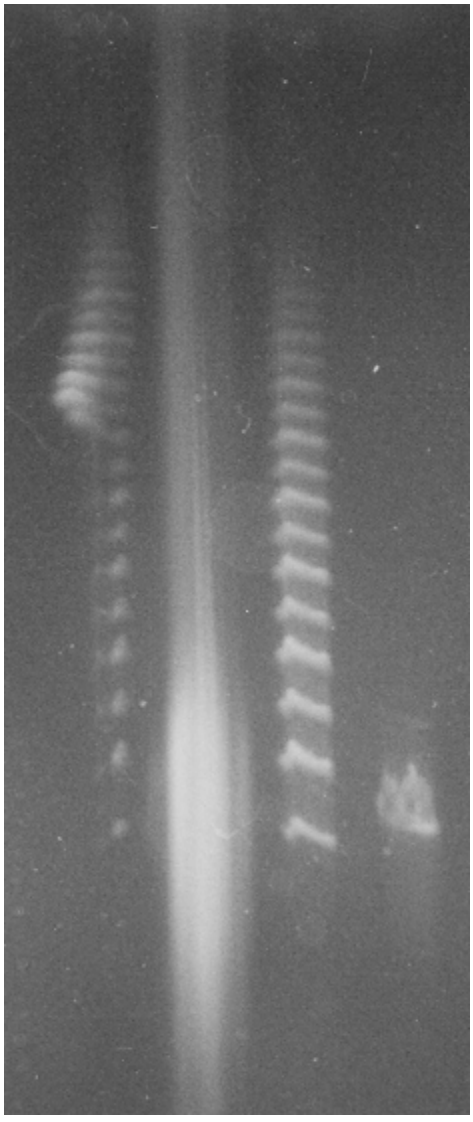


RT-PCR Analysis of Selected Gene Changes

YCL064C, catabolic serine/threonine dehydratase; YEL039C, iso-2-cytochrome C; YBR244W, glutathione peroxidase; YGL184C, cystathionine beta-lyase. The standard deviation between trials (T1 and T2) on all samples was + 0.2-0.3-fold.

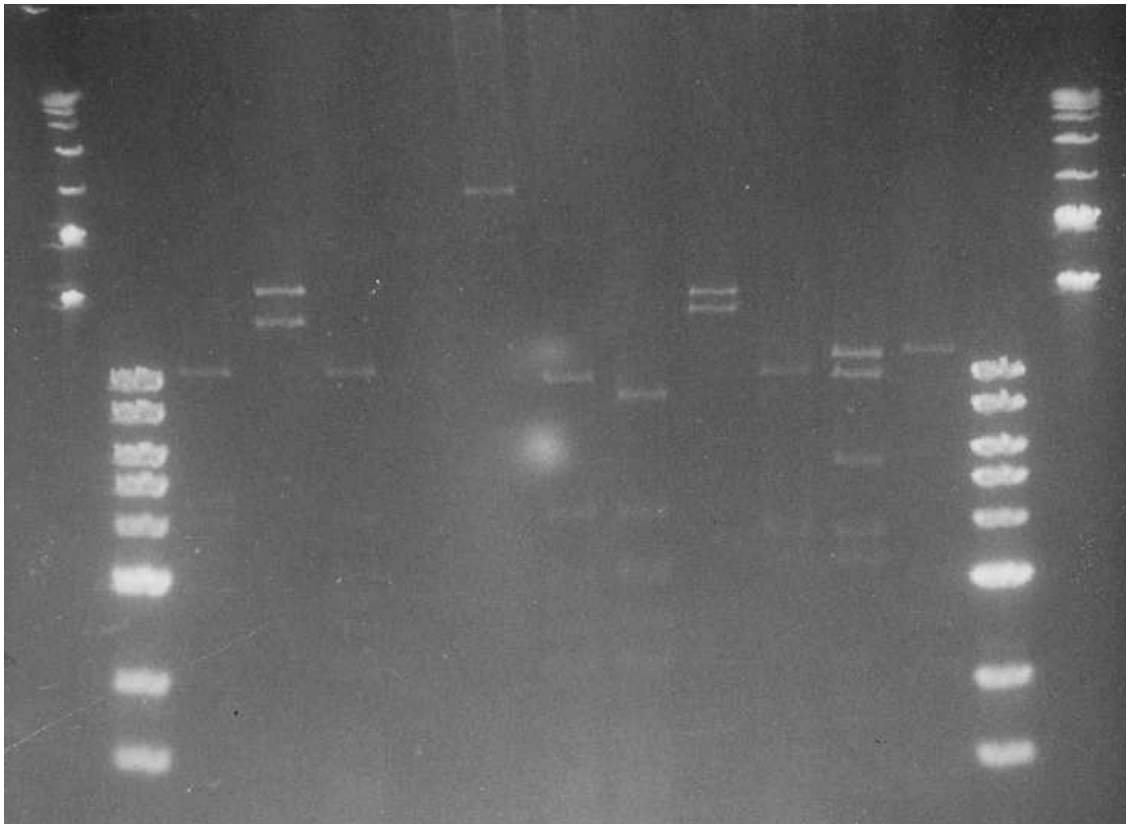
APPENDIX B

SUPPLEMENTARY DATA FOR CHAPTERS III AND IV



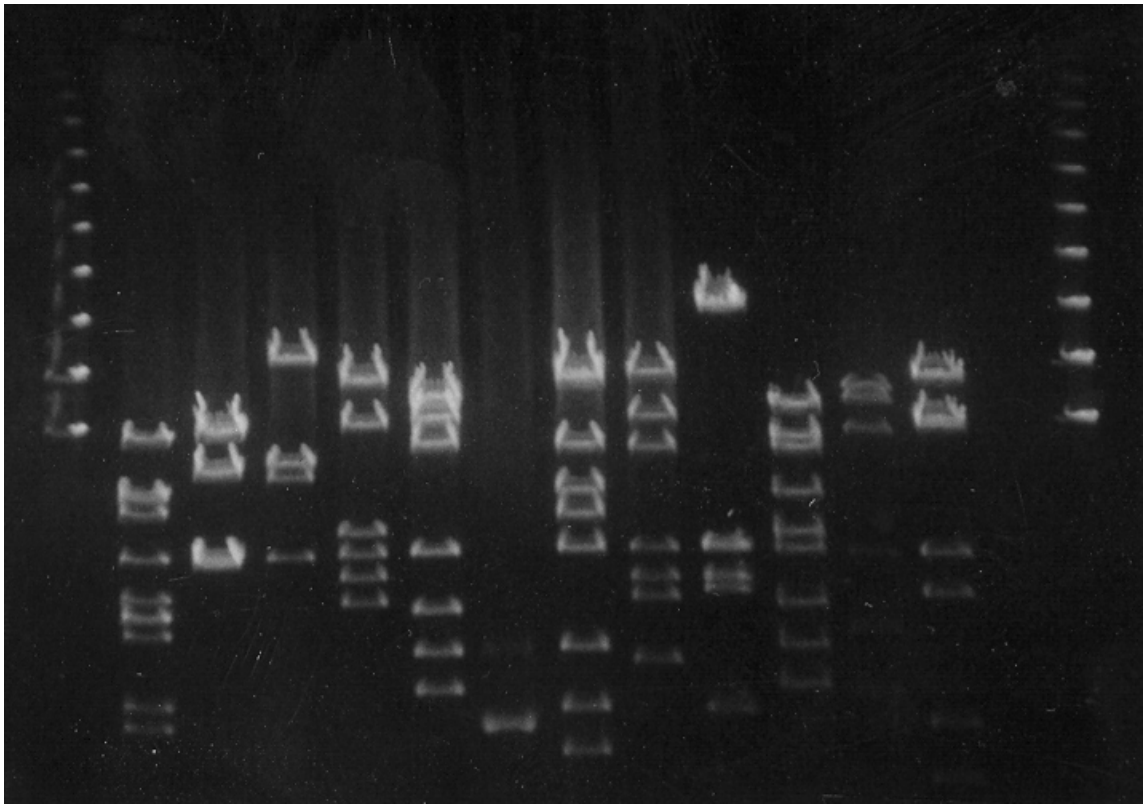
1 2 3 4

Isolation of High-Molecular Weight DNA For Fosmid Library Construction. Lanes 1 and 3 - NEB MidRange PFG Markers, Lane 4 - Bacteriophage λ DNA, Lane 2 – sheared DNA from filtered seawater sample.



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Analysis of Fosmid Inserts (Surf1). Lanes 1 and 15 - Lanes 1 and 15 – NEB MidRange II PFG Markers, Lanes 2 and 14 – NEB 1 kb ladder, Lanes 3 through 13 - NotI digests of randomly selected clones from Surf 1 fosmid Library.



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Analysis of BAC Inserts (Hon6) by HindIII Digestion. Lanes 1 and 15 – NEB MidRange II PFG Markers, Lanes 2 – 13 HindIII digests of randomly selected clones from Hon6 BAC Library.

APPENDIX C
SUPPLEMENTARY DATA FOR CHAPTER V

VBA Code for Genomic Library Simulator

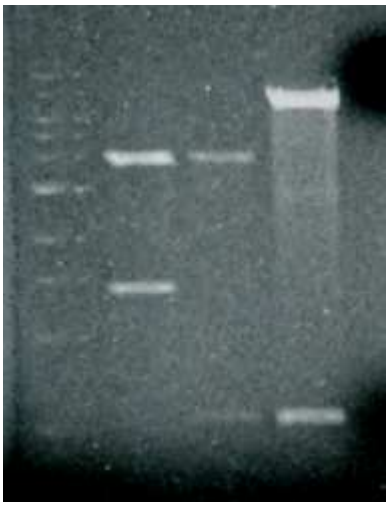
```
Sub StartButton_Click()
   'begin simulation
   Randomize
        PositiveLibraryCounter = 0
        Application.Workbooks("Library
                Simulator(works).xls").Worksheets("Data").Columns("A:H").Clear
        Application.Workbooks("Library Simulator (works).xls").Worksheets("Data").Select
        coef = (InsertSize - TargetPathSize) / InsertSize
        FractionIG = (InsertSize * 1000) / (GenomeSize * 1000000)
        ExpectedP = 1 - (1 - coef * FractionIG) ^ NumberClones

        For X = 1 To NumberSimulations
                ReDim Clone(1 To NumberClones)
                TargetSite = Rnd * GenomeSize * 1000000
                TargetLeft = Int(TargetSite)
                TargetRight = TargetLeft + TargetPathSize * 1000
                PositiveCloneCounter = 0
                For Y = 1 To NumberClones 'begin individual Library simulation
                        Clone(Y) = Int(Rnd * GenomeSize * 1000000)
                        InsertLeft = Clone(Y)
                        InsertRight = Clone(Y) + InsertSize * 1000
                        If InsertLeft < TargetLeft And InsertRight > TargetRight Then
                                PositiveCloneCounter = PositiveCloneCounter + 1
                        End If
                Next Y
                If PositiveCloneCounter > 0 Then
                        PositiveLibraryCounter = PositiveLibraryCounter + 1
                End If
                ' output results of individual library simulation
```

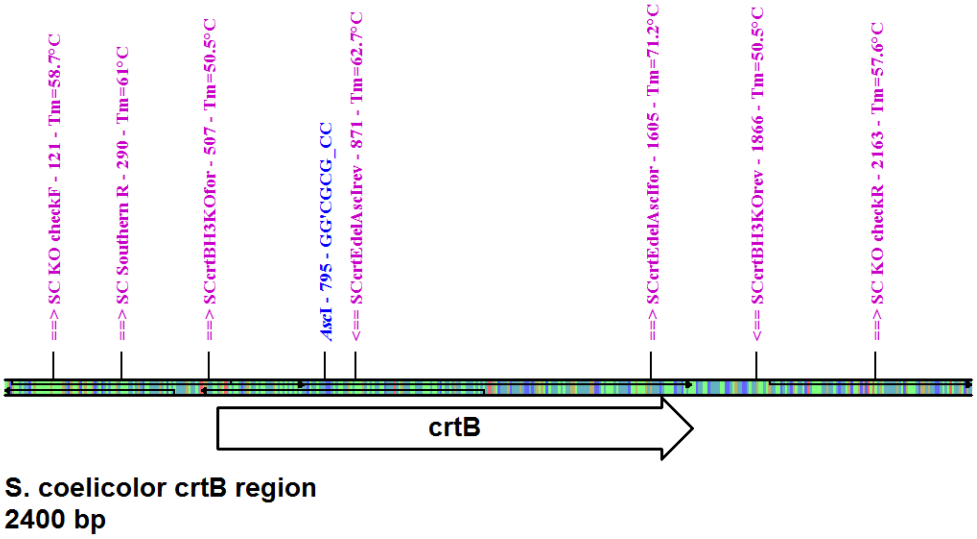
```
                Cells(4, 1) = "Sim. Lib. Number"
                Cells(4, 2) = "Pos. Clones"
                Cells(4, 3) = "Cum. Cloning P"
                Cells(4, 4) = "Calculated P"
                Cells(4 + X, 1) = X
                Cells(4 + X, 2) = PositiveCloneCounter
                Cells(4 + X, 3) = PositiveLibraryCounter / X
                Cells(4 + X, 4) = ExpectedP
        Next X
Unload SimOptionsBox
End Sub
```

APPENDIX D

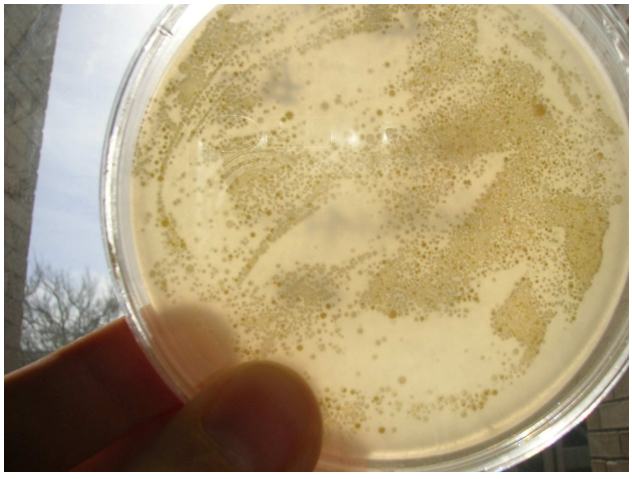
SUPPLEMENTARY DATA FOR CHAPTER VI



HindIII restriction digests of plasmids used for knockout. Lane 1 – NEB 1 kb ladder, Lane 2 – TA clone of the *Streptomyces coelicolor* crtB gene and surrounding region amplified by PCR using primers SCcrtBH3KOfor and SCcrtBKOrev. The band at 3.9 kb is the TA cloning vector. The band at 1.4 kb is the PCR generated insert. Lane 3 – plasmid constructed by inverse PCR of TA cloned crtB gene using primers SCcrtEdelAscIrev and SCcrtEdelAscIfor. The band at 3.9 kb is the TA cloning vector. The band at 600 bp is the crtB gene containing an 800 bp deletion. Lane 4 – pKC1139 carrying the deletion mutant crtB gene. The band at 6500 bp is pKC1139. The band at 600 bp is the deletion mutant crtB gene.

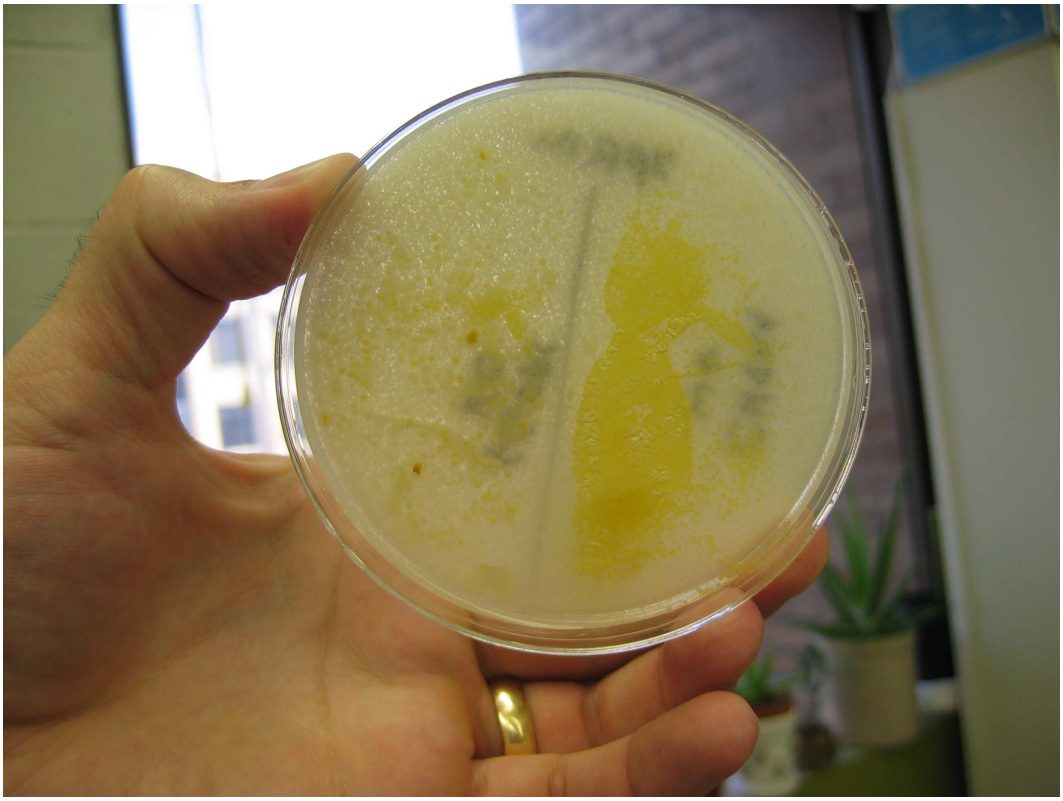


Map of the crtB region of *Streptomyces coelicolor*. Primer annealing sites and important restriction sites are indicated. Graphic was generated in pDRAW32.



Chromosomal deletion mutants growing on a MS plate, showing the white phenotype as a result of loss of the carotene pathway.

Chromosomal deletion mutants, left, growing on an MS plate next to the wild-type CH999 strain, right.



Desired gene disruption mutant, left, next to wild-type CH999, grown on SPRM over fluorescent light.

VITA

Name: Scott Edward Angell

Education: B.S., Genetics, Texas A&M University, 2001
Ph.D. Chemistry, Texas A&M University, 2008

Current Address: 3402 Preston Ave. #626
Pasadena, TX 77505

Publications:

Angell, S., B. J. Bench, H. Williams, and C. M. Watanabe. 2006. Pyocyanin isolated from a marine microbial population: synergistic production between two distinct bacterial species and mode of action. Chem. Biol. **13:**1349-59.

Kim, E. J., S. Angell, and C. M. Watanabe. 2007. Estimating P-Coverage of Biosynthetic Pathways in DNA Libraries and Screening by Genetic Selection: Biotin Biosynthesis in the Marine Microorganism *Chromohalobacter*. Mol. Biosyst. **In press.**