

**Construction and Phenotypic Screening of Mid-size Insert Marine Microbial
Environmental Genomic Libraries**

by

Jennifer C. Braff

B.A., New York University, 2001

Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Science
at the
Massachusetts Institute of Technology
and the
Woods Hole Oceanographic Institution

February 2008

© 2008 Massachusetts Institute of Technology
All rights reserved.

Signature of Author

Joint Program in Oceanography/Applied Ocean Science and Engineering
Massachusetts Institute of Technology
and Woods Hole Oceanographic Institution
December 6, 2007

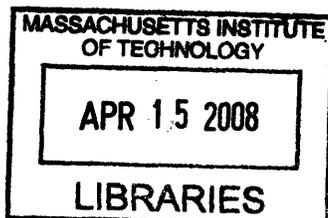
Certified by



Edward DeLong
Professor of Biological Engineering and Civil and Environmental Engineering
Chair, Joint Committee on Biological Oceanography
Thesis Supervisor

Accepted by

Daniele Veneziano
Professor of Civil and Environmental Engineering
Chairman, Departmental Committee on Graduate Students



ARCHIVES

**Construction and Phenotypic Screening of Mid-size Insert Marine Microbial
Environmental Genomic Libraries**

by

Jennifer C. Braff

Submitted to the Department of Civil and Environmental Engineering on December 6,
2007 in Partial Fulfillment of the Requirements for the Degree of Master of Science in
Biological Oceanography at the Massachusetts Institute of Technology and the Woods
Hole Oceanographic Institution

ABSTRACT

Functional screening of environmental genomic libraries permits the identification of clones expressing activities of interest without requiring prior knowledge of the genes responsible. In this study, protocols were optimized for the construction of mid-size DNA insert, inducible expression environmental genomic plasmid libraries for this purpose. A library with a mean insert size of 5.2 kilobases was constructed with environmental DNA isolated from surface ocean water collected at Hawaii Ocean Time-series station ALOHA in plasmid cloning vector pMCL200 under the inducible control of the P_{LAC} promoter. To begin to evaluate the utility of such libraries for gene expression-based screens, this library was screened phenotypically for clones expressing genes that confer fluorescence or distinctive coloration on colonies of host *Escherichia coli* cells, and results were compared to those for a fosmid library constructed from the same marine microbial DNA sample. Ecologically relevant sequences were identified in both libraries, and each was observed to offer both advantages and disadvantages. Results of this study suggest that mid-size insert plasmid libraries under the control of inducible promoters can provide a useful and complementary approach for both functional screening and shotgun sequencing of environmental genomic libraries.

Thesis Supervisor: Edward DeLong

Title: Professor, Division of Biological Engineering and Department of Civil and Environmental Engineering

Background

Since the development of culture-independent methods in environmental microbiology (Pace et al. 1985), environmental genomics has emerged as a powerful tool in assessing the taxonomic composition and genetic potential of marine planktonic microbial populations. The first culture-independent environmental genomic studies were aimed primarily at investigating phylogenetic diversity. Phylogeny was assigned by determining 5S rRNA (Stahl et al. 1985) or 16S rRNA (Lane et al. 1985) sequences directly from isolated microbial RNA. These studies offered the first glimpse at the phylogenetic composition of natural microbial populations, as the vast majority of environmental microbes have resisted cultivation (Staley and Konopka 1985), greatly limiting our ability to study them by the use of traditional microbiological methods. 16S small subunit (SSU) rRNA sequence clone libraries were later constructed and used to examine the phylogenetic makeup of marine bacterioplankton (Giovannoni et al. 1990), as well as hot spring microbial communities (Ward et al. 1990). Results of these early culture-independent phylogenetic studies revealed the presence of many previously unknown groups and demonstrated the usefulness of this approach for studying natural microbial populations. However, functional capabilities associated with many groups are largely unexplored, and recent studies have demonstrated that the distribution of metabolic characteristics do not necessarily track phylogenies based on SSU rRNA sequences (Frigaard et al. 2006), (Nakagawa et al. 2005).

By cloning genomic DNA isolated from marine picoplankton communities directly into bacteriophage lambda vectors and subsequently identifying and sequencing 16S rRNA sequences carried on individual clones (Schmidt et al. 1991), pre-cloning sequence amplification or reverse transcriptase steps could be avoided. More importantly, direct cloning of environmental DNA fragments made it possible to move

beyond assessment of phylogenetic diversity and examine the genomic context associated with SSU rRNA sequences. The development of low copy number, F factor based cloning vectors, including bacterial artificial chromosomes (BACs) which can hold 300 kilobases of DNA or more (Shizuya et al. 1992) and low copy cosmids known as *fosmids* (Kim et al. 1992), allowed stable, large insert clone libraries to be conveniently constructed and manipulated. Full sequencing of large-insert clones carrying SSU rRNA genes or other phylogenetic markers made possible analysis of genome fragments derived from abundant, yet poorly understood, marine *Chrenarchaeota* and yielded insight into their functional and ecological characteristics (Stein et al. 1996), (Hallam et al. 2006a). Analysis of uncultivated microbial genome fragments began to reveal the extent to which microbial diversity was largely unexplored and made apparent the full potential of environmental genomics. These techniques offered a means to gain insight into the functional and metabolic potential of various groups and to develop testable hypotheses linking phylogeny and function within natural microbial populations.

The discovery of proteorhodopsin, a gene with homology to previously characterized archaeal rhodopsins, in the sequence of a marine BAC clone derived from an uncultivated SAR86 type Gammaproteobacteria is one such example. Since *Escherichia coli* cells harboring this BAC expressed the proteorhodopsin gene, it was possible to demonstrate that the protein functioned as a light-driven proton pump in the presence of retinal (Beja et al. 2000). Functional characterization of a large-insert environmental genomic clone containing both proteorhodopsin and retinal biosynthetic genes has since demonstrated that the proteorhodopsin photosystem is sufficient to allow photophosphorylation in host *Escherichia coli* cells (Martinez et al. 2007). Sequence analysis of numerous large-insert environmental genomic clones carrying proteorhodopsin genes has provided insight into the evolutionary relationships between these uncultivated organisms, as well as their metabolic characteristics and ecological adaptations (Sabehi et al. 2005), (Frigaard et al. 2006), (McCarren and DeLong 2007). The combination of functional characterization and environmental genomic sequence

analysis has greatly advanced our understanding of the role that proteorhodopsin plays in planktonic marine microbes.

In recent years, this approach has been adapted to a variety of applications, and there have been important developments in library construction methodologies, sequencing technology, and bioinformatic analytic capabilities. The adaptation of shotgun genome sequencing methods to environmental samples has proven to be an effective means of characterizing natural microbial populations. Shotgun sequencing of plasmid libraries constructed from acid mine drainage biofilm samples allowed reconstruction of composite genomes for dominant organisms found in this specialized and relatively simple microbial community (Tyson et al. 2004). A similar approach making use of overlapping fosmid sequences was used to construct a composite genome of the sponge symbiont *Cenarchaeum symbiosum* and to examine genomic variability between two distinct subpopulations (Hallam et al. 2006b). The power of high-throughput shotgun sequencing to characterize complex planktonic microbial populations was illustrated by the sequencing of more than a billion base pairs of DNA isolated from Sargasso Sea seawater (Venter et al. 2004) and, more recently, approximately 6 billion base pairs of marine microbial DNA (Rusch et al. 2007). Environmental genomic surveys of this scale provide researchers with a wealth of data that will provide crucial information relating phylogenetic composition and functional properties of marine microbial populations. Recent studies have used environmental genomic sequence data to compare microbial populations between environments, with an emphasis on functional characteristics of the community rather than phylogenetic composition. Comparison of gene content, as assessed by analysis of environmental shotgun sequence data from disparate environments and similar yet geographically distant sites, revealed numerous habitat-specific characteristics consistent with our understanding of these environments (Tringe et al. 2005). End-sequencing and subsequent analysis of fosmid libraries constructed from microbial DNA sampled from discrete depths at Hawaii Ocean Time-series (HOT) station ALOHA resulted in gene profiles that reflect known distributions of metabolic properties with depth, as well as hinting at intriguing and previously

unobserved community characteristics such as patterns in viral abundance across the water column and overrepresentation of transposase sequences in deep water samples (DeLong et al. 2006).

As the number of fully sequenced microbial genomes increases and databases of known genes sequences and proteins become larger and more sophisticated, so does our ability to annotate environmental sequence data and thus draw environmentally meaningful conclusions from these data sets. However, a large fraction of open reading frames identified in environmental genomic data have no known function assigned, and in many cases genes responsible for key processes have not been identified. In cases where the host cell strain and cloning vector permit the relevant genes to be expressed by host cells, functional screening of environmental genomic libraries is one approach with which to probe this unknown sequence space. By screening libraries for functions rather than sequences of interest, researchers may identify previously unknown genes or sequences of known genes that are too divergent to be observed by sequence-dependent methods such as hybridization or polymerase chain reaction (PCR).

An early application of this approach involved screening environmental genomic libraries for catabolic genes of interest, such as genes that confer on host *Escherichia coli* cells the ability to utilize 4-hydroxybutyrate as a carbon source. Growth assays with environmental genomic libraries constructed from soil samples with 4-hydroxybutyrate as a test substrate resulted in the identification of new variants of known 4-hydroxybutyrate dehydrogenases, as well as clones which exhibited 4-hydroxybutyrate dehydrogenase activity but whose sequences had no homology to previously characterized enzymes (Henne et al. 1999). Novel lipolytic enzymes were also detected by visually screening soil libraries plated out on supplemented solid media (Henne et al. 2000), and environmental $\text{Na}^+(\text{Li}^+)/\text{H}^+$ antiporter genes were identified by plating these libraries on agar plates containing LiCl in antiporter-deficient host cells (Majernik et al. 2001). Variations on this approach involving the substrate induced expression of catabolic genes of interest cloned into green fluorescent protein (GFP) fusion vectors allow higher throughput screens and easy isolation via fluorescence activated cell sorting

(FACS) of GFP expressing clones (Uchiyama et al. 2004). Functional screening of environmental genomic libraries has also proven to be a promising technique in the search for novel antibiotics (Courtois et al. 2003) and molecules involved in quorum sensing (Williamson et al. 2005).

Since functional screening of environmental libraries depends on successful expression of cloned genes in library host cells, in many cases screens may have a better outcome if multiple host cell types can be employed. New methodologies that allow BAC libraries to be shuttled between host species have proven useful in optimizing gene expression-based screens, as some genes that fail to express well in *Escherichia coli* were successfully expressed in *Streptomyces lividans* or *Pseudomonas putida* host cells (Martinez et al. 2004). Multiple host-range libraries have also been employed in isolating novel alcohol dehydrogenase genes (Wexler et al. 2005) and tryptophan synthesis (*trp*) operons from waste-water treatment libraries (Li et al. 2005).

Screens of this type often require that genes encoded by cloned environmental DNA be expressed at relatively high levels in the host cell, as low-level expression may not result in a sufficiently distinct phenotype. However, high-level expression of foreign DNA places a significant burden on host cells (Guzman et al. 1995) and may constitute an additional source of cloning bias in library construction, as well as reducing clone stability (Kim et al. 1992). A number of cloning technologies have been developed to address such problems, some of which have already been successfully employed in environmental genomic research. Dual origin BACs that allow BAC copy number to be elevated by induction of the higher copy *oriV* origin greatly increase yield on vector DNA extraction so facilitate sequencing and other downstream applications. The host cell strain for these vectors, *E. coli* strain Epi300, carries a chromosomally integrated *trfA* replication gene, which is under the control of the arabinose inducible P_{BAD} promoter and is required for high copy number replication from the *oriV* origin (Wild et al. 2002).

Adaptation of this methodology to fosmid vectors, such as the Epicentre CopyControl Fosmid Library Production Kit (Epicentre, Madison, WI), offers a number of advantages for environmental genomic library construction. Phage packaging allows

high quality libraries to be produced with limited quantities of environmental insert DNA, as phage packaging size selection virtually eliminates the problem of background non-insert containing clones, and transduction allows 40 kilobase chunks of insert DNA to be transferred to host cells with high efficiency (Ausubel et al. 1992). As with variable copy number BACs, copy number can be elevated prior to DNA preparation. In addition, the copy-up feature of these vectors has facilitated phenotypic screening of surface water libraries for orange-tinted clones carrying proteorhodopsin and retinal biosynthetic genes (Martinez et al. 2007). However, insert instability has been observed under high-copy conditions, so such vectors may be of limited utility in screens that require high-level expression for extended periods of time.

Plasmid vectors with inducible promoters, which have long been used in gene expression studies and for heterologous protein production, offer an alternative to variable copy-number cloning vectors for functional or phenotypic screening of environmental genomic libraries. A number of plasmid vectors with various origins of replication, antibiotic resistance markers, and transcription promoter combinations have been constructed that allow cloning into low or mid-range copy number plasmids and tight regulation of transcription into the multi-cloning site. Sets of plasmid vectors with compatible origins of replication and promoters engineered for both independent regulation and enhanced gene expression range are also available (Lutz and Bujard 1997), (Registry of Standard Biological Parts, <http://parts.mit.edu>). These may provide combinations suitable for functional or phenotypic screens that require multiple genes to be co-expressed at specified levels (Williamson et al. 2005).

In this study, medium copy number plasmids incorporating the P_{BAD} and P_{LAC} inducible expression systems were investigated for the purpose of constructing environmental genomic libraries to be used in gene expression-based screening. Plasmid vectors that place cloned DNA under the control of the arabinose operon P_{BAD} promoter and its regulatory gene *araC* offer both positive and negative regulation of insert gene expression. Transcription is tightly repressed in the presence of glucose and can be induced to moderately high levels by the addition of L-arabinose. The P_{BAD} system

offers an additional advantage in that intermediate levels of induction can be reliably produced under the correct conditions (Guzman et al. 1995). This capability may prove valuable in developing functional screens as multiple expression levels can be tested without requiring additional library construction or transfer of the library into alternate host strains. Medium to low copy plasmids carrying the stronger isopropyl- β -D-thiogalactopyranoside (IPTG) inducible P_{LAC} regulatory element have also been developed with the similar aim of allowing researchers to clone genes that are unstable or difficult to clone at high copy number (Nakano et al. 1995).

Methods

Construction of Mid-Size Insert Plasmid Libraries

pMCL200 Vector Preparation

Plasmid pMCL200 was prepared according to protocols obtained from the Joint Genome Institute (JGI) (Walnut Creek, CA). Briefly, plasmid DNA was isolated from liquid cultures of *E. coli* cells using the Qiagen HiSpeed Midi Kit (Qiagen, Valencia, CA) according to manufacturer's directions. Plasmid DNA was digested for 4 hours at 37° C with EcoRV restriction endonuclease (New England Biolabs, Beverly, MA) yielding blunt ends, and enzyme was subsequently removed using the Qiagen QIAQuick PCR Purification kit (Qiagen, Valencia, CA). 5' phosphate groups were removed from vector DNA using 100 units shrimp alkaline phosphatase (Roche Applied Science) per 200 μ L reaction containing up to 2 μ g of linear vector DNA. Reactions were run for 2 hours at 37° C and then heat inactivated for 15 minutes at 65° C before being run out on a 1% low-melting-point (LMP) SeaPlaque agarose (FMC BioProducts, Rockland, ME) gel. Gels were stained with, SYBR Gold nucleic acid stain (Molecular Probes, Eugene OR) and visualized on a non-UV Dark Reader transilluminator (Clare Chemical Research,

Dolores, CO) for band excision. Linear, blunt-end, dephosphorated vector DNA was then purified using Qiagen QIAQuick Gel Extraction kit (Qiagen, Valencia, CA).

pBAD33 Vector Preparation

pBAD33 plasmid DNA was isolated from liquid cultures of *E. coli* cells by alkaline lysis and isopropanol precipitation of plasmid DNA according to standard protocols. Plasmid DNA was further purified by two rounds of CsCl equilibrium density gradient centrifugation (Sambrook and Russell 2001), (CHS Protocols; 2006; doi:10.1101/pdb.prot3927). Purified plasmid DNA was digested with SmaI restriction endonuclease (New England Biolabs, Beverly, MA) overnight at room temperature. The reaction was heat inactivated by incubation for 20 minutes at 65° C. Vector dephosphorylation was performed at 37° C for 20 minutes using 5 units of Apex-Heat-Labile Alkaline Phosphatase (Epicentre, Madison, WI) per µg vector DNA, followed by a 5 minute heat inactivation at 70° C. To reduce the potential for vector religation in downstream applications, phosphatase-treated vector DNA was subjected to a self-ligation step prior to gel purification to religate any incompletely dephosphorylated vector. The sample was EtOH precipitated according to standard protocols and resuspended in sterile water at a concentration of 100 ng/µL. Self-ligation reactions were run overnight at 16° C with 2000 units of T4 DNA ligase (New England Biolabs, Beverly, MA) for each µg of plasmid DNA in a volume of 100 µL. After heat inactivation, blunt-end linear vector DNA was run out on a 1% LMP SeaPlaque agarose gel (FMC BioProducts, Rockland, ME) to separate linear vector DNA from uncut or religated pBAD33 plasmid. Gels were stained with SYBR Gold stain (Molecular Probes, Eugene, OR) and viewed on a non-UV transilluminator (Dark Reader, Clare Chemical Research, Dolores, CO); vector was isolated using the Qiagen QIAEX II kit (Qiagen, Valencia, CA) according to manufacturer's directions.

Insert Preparation

Marine microbial DNA samples were isolated from the 0.2-2 μm size fraction of filtered seawater as described previously (DeLong et al. 2006), (Suzuki et al. 2004). Library construction began with extracted and concentrated crude DNA, which had not been further purified by CsCl density gradient centrifugation. Random shearing of insert DNA was performed using a GeneMachines Hydroshear (Genomic Solutions, Ann Arbor, MI). The Hydroshear was calibrated with sample DNA prior to preparation of insert DNA to determine the syringe speed necessary to produce fragments of the desired size. Crude DNA samples were filtered through a 0.2 μm filter and diluted to <100 ng/ μL in filter sterilized Tris-EDTA (TE) before shearing to avoid clogging the shearing orifice. Sample volume was generally between 50 and 150 μL . After this filtration step, sample preparation proceeded according to Hydroshear manufacturer's directions (Genomic Solutions, Ann Arbor, MI). Samples were incubated at 37° C for 30 minutes, with vortexing every 10 minutes, then spun at 13,000 rpm on a bench top microcentrifuge. No pellet was ever observed at this step, and samples were transferred to clean 1.5 mL tubes and kept on ice. After pre-washing the Hydroshear (Genomic Solutions, Ann Arbor, MI) with four washes of 0.2 M HCl, four washes of 0.2 M NaOH, and four washes of TE, samples were sheared at speed code 15 for 6-10 kilobase fragments or 16 for 8-12 kilobase fragments for 25 cycles and immediately placed on ice.

To produce blunt-end fragments with 5' phosphate groups for cloning, sheared insert DNA was treated using End-It DNA End-Repair Kit (Epicentre, Madison, WI). Reactions were set up according to the manufacturer's directions and scaled up as needed. After a 60 minute incubation at room temperature, 125 mM EDTA was added to a concentration of 10 mM in order to prevent exonuclease activity by T4 DNA polymerase, and reactions were heat inactivated for 10 minutes at 70° C. Samples were extracted once with 25:24:1 (v/v/v) phenol/chloroform/isoamyl alcohol, followed by extraction with 24:1 (v/v) chloroform/isoamyl alcohol to remove residual phenol. Finally, insert DNA was ethanol precipitated with Pellet Paint Co-Precipitant (Novagen)

according to the manufacturer's instructions and resuspended in 30 μ L sterile 10 mM Tris-Cl.

Sheared, end-repaired insert DNA was size-selected by running samples on a 1% LMP SeaPlaque agarose (FMC BioProducts, Rockland, ME) gel in tris-acetate-EDTA (TAE) buffer alongside DNA size standards. Gels were cast with sufficient volume, when possible, to allow the entire sample to be run in a single lane. Before use, gels were chilled at 4° C then run at approximately 100 volts for three to four hours. TAE running buffer was changed every two hours to prevent overheating. Gels were stained with freshly prepared SYBR Gold nucleic acid gel stain (Molecular Probes, Eugene, OR) in TAE buffer for 15 minutes shielded from light with gentle shaking. A non-UV safe-light transilluminator (Dark Reader, Clare Chemical Research, Dolores, CO) was used to view the gel, and two size fractions of insert DNA, a longer one containing the majority of the sample and a shorter secondary fraction, were excised from the gel.

Agarose gel purification of DNA fragments shorter than 10 kilobases was performed using Qiagen QIAQuick Gel Extraction Kit (Qiagen, Valencia, CA) according to manufacturer's directions with slight modifications. Both the isopropanol and solubilization buffer (QG) column wash steps were included. DNA was eluted with buffer EB (10 mM Tris-Cl pH 8.5) pre-warmed to 50° C, and samples were allowed to sit ten minutes after buffer EB was applied to the columns to increase recovery of larger fragments. The Qiagen QIAEX II kit (Qiagen, Valencia, CA) was used according to manufacturer's directions in cases where the excised gel slice included fragments 10 kilobases or greater in length. No more than 150 mg of agarose gel was processed in a single reaction, and samples were flicked, rather than vortexed, to avoid shearing insert DNA.

Ligation, Transformation, and Plating of Libraries in pMCL200

Prior to ligation, aliquots of both the purified vector and insert DNA were run out on 1% agarose gels stained with ethidium bromide next to DNA size standards of known

concentration (1 kb DNA ladder, New England Biolabs, Beverly, MA) to determine the approximate concentration of each component. Ligations were conducted in 10-15 μL volumes with approximately 10 ng of vector DNA. An insert to vector molar ratio of approximately 3:1 was found to be a good target ratio for this protocol. Blunt-end ligation reactions were run overnight (~16 hours) at 16° C with 200 units of T4 DNA ligase (New England Biolabs, Beverly, MA). After ligation, the reaction was heat inactivated for 10 minutes at 65° C, followed by dialysis against MilliQ water for 30 minutes using a Millipore 0.025 μm filter to remove salts before electroporation.

After dialysis, approximately 10% of the ligation reaction (1-1.5 μL) was gently mixed with 50 μL of Top10-F' electrocompetent cells (Invitrogen, Carlsbad, CA). The remainder of the ligation was stored at -20° C, in case additional clones were required. Cells were transferred to a chilled 1 mm cuvette, and electroporation was performed at 25 μF , 2.0 kV, and 200 Ohms on a BTX ECM 630 electroporator (BTX Harvard Apparatus). Cells were immediately mixed with 1 mL room temperature SOC media and incubated at 37° C for one hour with shaking, according to standard protocols (Sambrook and Russell 2001), (CSH Protocols; 2006; doi:10.1101/pdb.prot3933). A 10 μL sample of transformation product was set aside to evaluate transformation efficiency and check library quality, and the remainder was mixed with an equal volume of 50% glycerol and stored at -80° C.

The 10 μL sample was then diluted 1:20 into fresh SOC to give a final volume of 200 μL . 100 μL was plated onto a small Luria-Bertani (LB) plate with 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol, 10 $\mu\text{g}/\text{mL}$ tetracycline, and 40 $\mu\text{g}/\text{mL}$ X-gal, and an equal quantity was plated similarly with the addition of 0.5 mM IPTG. Plates were incubated at 37° C for 16 hours and scored for colony forming units (cfu) per mL transformation product and percentage of blue, apparently non-insert containing colonies. Between 6 and 12 white clones were grown up in liquid culture, plasmid DNA was prepared, and approximately 200 ng was digested with BamHI and HindIII restriction endonucleases in NEB buffer 2 at 37° C for 4 hours (New England Biolabs, Beverly, MA). Digests were run out on a 1% agarose gel and stained with EtBr to evaluate mean insert size.

The remaining transformation product (stored as a 25% glycerol stock at -80° C), was plated at a cell density of <3000 cfu/plate, onto large rectangular LB plates with 12.5 µg/mL chloramphenicol, 10 µg/mL tetracycline, 40 µg/mL X-gal, and 1 mM IPTG for automated picking. Plates were incubated at 37° C for 16 hours, and colonies were picked using a Genetix QPixII robot (Genetix, Hampshire, UK) into 384-well plates containing LB-glycerol with 12.5 µg/mL chloramphenicol and 10 µg/mL tetracycline, to be grown up overnight and stored at -80° C. With the exception of colony-picking settings adjusted to preferentially pick white colonies, colony picking was conducted as usual for environmental genomic fosmid libraries.

Ligation, Transformation, and Plating of Libraries in pBAD33

Libraries constructed in plasmid pBAD33 were produced as described above for plasmid pMCL200 with minor modifications. Ligations were transformed into electrocompetent DH10B Electromax cells (Invitrogen, Carlsbad, CA) and plated onto media supplemented with chloramphenicol only. To spot check insert length, clones were digested with SacI and XbaI restriction endonucleases in buffer 4 (New England Biolabs, Beverly, MA) at 37° C for four hours.

Arraying and Phenotypic Screening

Libraries were screened phenotypically for colony color and fluorescence by printing cells from 384-well library glycerol stocks onto rectangular LB plates supplemented with 12.5 µg/mL chloramphenicol, 10 µg/mL tetracycline, with and without 0.5 mM IPTG, and overlaid with Genetix Performa II positively-charged nylon (Genetix, Hampshire, UK) membranes, using a Genetix QPixII robot (Genetix, Hampshire, UK). Twenty-four 384-well plates were arrayed per membrane, and 4 replicates were printed of each membrane (i.e. two for growth with IPTG inducer and two without). One set was incubated for 16 hours at 37° C and one set for 16 hours at 30° C,

after which membranes were moved to room temperature; all incubations were conducted in the dark. Plates were scored for colonies exhibiting a non-wild type color phenotype immediately after the initial 16 hour incubations and then periodically during approximately two weeks at room temperature, since some color phenotypes were visible only after extended incubations. In order to detect colonies expressing genes that confer a fluorescent phenotype on host cells, membranes were scanned and imaged after a week at room temperature using a Fujifilm FLA-5100 imaging system (Fujifilm, Stamford, CT). Scanning was conducted for all membranes at 473 nm, 532 nm, and 635 nm excitation wavelengths with appropriate filters.

To confirm color phenotype, clones of interest were re-streaked from library glycerol stocks onto LB plates supplemented with 12.5 µg/mL chloramphenicol, 10 µg/mL tetracycline, with and without 0.5 mM IPTG, and overlaid with hydrophilic polyethersulfone Supor-200 0.2 µm filters (Chisholm Corp., Lincoln, RI). Fosmid library clones were streaked out as above onto plates with 12.5 µg/mL chloramphenicol, with and without 0.02% L-arabinose to raise fosmid copy number. Provided the expected color phenotype was observed upon re-streaking, single colonies were picked and grown up in liquid culture with appropriate selection. Library vector DNA was prepared and retransformed into the library host *E. coli* strain, and single retransformed colonies were picked and grown up in liquid culture with selection to make 25% glycerol freezer stocks. If the color phenotype of interest was again observed upon streaking out cells as described above from these retransformed cell stocks, the phenotype was considered confirmed.

Arraying with Retinal

In order to screen mid-size insert plasmid library HMS186_25m for proteorhodopsin carrying clones, the library was arrayed as described above for color and fluorescence screening except that exogenous retinal was applied directly to the membrane surface. Genetix Performa II (Genetix, Hampshire, UK) membranes

overlaying solid media supplemented with 12.5 µg/mL chloramphenicol and 10 µg/mL tetracycline, with and without 1 mM IPTG, were sprayed in situ with a fine mist of 5 mM all-*trans* retinal in 95% ethanol using a plastic spray bottle (Fisher Scientific, Springfield, NJ) and allowed to dry before printing. After the cells were arrayed, plates were incubated overnight at 30° C, then placed at room temperature for several days before being scored for color as described above for color and fluorescence screening.

DNA Sequencing and Sequence Analysis

Selected clones in both mid-size insert plasmid library HMS186_25m and fosmid library HF186_25m with confirmed non-wild type colony color phenotypes were end-sequenced. Sequencing of fosmid ends was performed using forward primer T7 and reverse primer EpiFos5R to fosmid vector pCC1FOS (Epicentre, Madison, WI). For plasmid library clones, primers pMCL200-seq-FW and pMCL200-seq-REV (Joint Genome Institute, Walnut Creek, CA), which read into the multi-cloning site of plasmid vector pMCL200, were used to obtain forward and reverse end-sequences (Table S1). In several cases, additional or complete insert sequences for clones in library HMS186_25m were obtained by subsequent rounds of primer walking.

To identify genes responsible for colony color phenotypes identified in fosmid libraries, transposon mutagenesis of fosmid DNA was performed using the EZ-Tn5 <KAN-2> Insertion kit (Epicentre, Madison, WI). After transformation with mutagenized fosmid DNA, cells were plated on solid selective media with 0.02% L-arabinose to raise fosmid copy number and an opaque membrane overlay so that colony color could be easily visualized. Colonies in which the color phenotype of interest had been knocked out (i.e. wild type colored colonies) were picked, and fosmid DNA was prepared for sequencing with primers, Kan-2 FP-1 and Kan-2 RP-1 (Table S1), off the ends of the EZ-Tn5 <KAN-2> transposon (Epicentre, Madison, WI). In cases where additional or full-length insert sequence was desired, additional transposon insertion mutants were picked for sequencing. All DNA sequencing was carried out using the

BigDye v3.1 sequencing system with capillary electrophoresis run on an ABI Prism 3700 DNA analyzer (Applied Biosystems, Forest City, CA). Fosmid sequences were assembled using Sequencher v.4.5 DNA assembly software (Gene Codes, Ann Arbor, MI), and preliminary automated annotations were done with fgenesb software (Softberry, Mount Kisco, NY). Unless otherwise noted, end-sequences, transposon insertion reads, and other short sequences were analyzed by BLASTX (Altschul et al. 1990) against NCBI non-redundant protein sequences.

HPLC Analysis

To prepare pigments from carotenoid-producing fosmid clones for high pressure liquid chromatography (HPLC), liquid cultures of Epi300 *E. coli* cells harboring fosmids of interest were grown up overnight at 30° C with shaking in M9 minimal media supplemented with 0.4% glycerol, 0.1% casamino acids, 0.03% thiamine, and 12.5 µg/mL chloramphenicol. In the morning, cultures were diluted back 1:100 into fresh media and grown at 30° C with shaking for 3-4 hours before induction to high copy with 0.02% L-arabinose. Cultures were allowed to grow for an additional 16 hours before cells were pelleted and stored at -80° C.

To extract pigments, cells from 2 mL worth of culture were resuspended in 300 µL of 7:2 vol/vol acetone:methanol then vortexed briefly, followed by 20, 1 second pulses of sonication. Tubes were then spun for 1 minute at 13,000 rpm on a table-top microcentrifuge, and the supernatants were transferred to clean tubes. Samples were brought to 1 mL total volume with additional 7:2 vol/vol acetone:methanol and filtered through 0.45 µm filters. Filters were washed with another mL of 7:2 vol/vol acetone:methanol, and samples were dried down to volume under nitrogen. During this process, samples were kept on ice and shielded from light whenever practical. HPLC was performed with a Waters (Milford, MA) separations module controlled by MassLynx 4.0 software, with a Waters 996 photodiode array detector. The column was a 5 µm Zorbax-ODS C18 column (150 X 4.6 mm), (Agilent Technologies, PaloAlto, CA). The

solvent system has been used previously (Maresca et al. 2007), although the gradients were modified somewhat for this work (Fig. S1). Solvent A consisted of 62.5% water, 21% methanol, and 16.5% acetonitrile with 10 mM ammonium acetate, and Solvent B consisted of 50% methanol, 30% ethyl acetate, and 20% acetonitrile. A flow rate of 0.8 ml/min was used for all samples. Purified retinal, astaxanthin, lycopene, and β -carotene standards were run during each experiment.

Results and Discussion

Library Construction and Protocol Optimization in Plasmid pBAD33

This work was undertaken in order to construct randomly sheared mid-size insert environmental genomic libraries in plasmid vectors with inducible promoters for the purpose of functional and phenotypic screening. The target mean insert size of 8-10 kilobases was selected to allow entire operons to be captured in single clones. Although numerous commercially prepared library construction kits are available and protocols exist for the construction of randomly sheared plasmid genomic libraries, no commercial products were available that met the specific requirements of this project. Available protocols for mid-size insert plasmid or fosmid genomic libraries and small insert shotgun environmental genomic libraries (Hallam et al. 2006a) required significant optimization in order to yield high-quality libraries of the desired type from our marine microbial DNA samples. Protocols for library construction used for this work were adapted from JGI protocols for genomic library construction (see <http://www.jgi.doe.gov/sequencing/protocols/index.html>) and Epicentre protocols for CopyControl Fosmid library construction. Plasmid vector pBAD33 was selected initially for library construction, although mid-size insert libraries for screening were eventually constructed in the plasmid vector pMCL200 (Fig. S2). The multi-cloning site on pBAD33 is under the control of the L-arabinose inducible P_{BAD} promoter, and it carries the P_{BAD} positive and negative regulator, *araC*, as well as a chloramphenicol resistance

marker. It has the same mid-low copy p15A origin of replication as pMCL200 but lacks blue/white screening for the presence of insert. However, it offers tight repression in the presence of glucose and provides lower expression levels than those obtained with the P_{LAC} induction system (Guzman et al. 1995).

Random shearing of marine microbial community DNA was performed as described using a Hydroshear (Genomic Solutions, Ann Arbor, MI), which fragments DNA to a given size range by forcing DNA samples repeatedly through a shearing orifice at a designated speed using an adjustable syringe pump. Since shearing orifices vary slightly, each one must be calibrated prior to use. To conserve environmental DNA samples, the Hydroshear shearing orifices were calibrated using expendable samples or stock lambda DNA. Fragment size was found to be relatively sample-independent, so calibration with either was adequate. The bulk of sheared sample DNA generally fell within a 3 kilobase size range, although a significant quantity of larger and smaller fragments were also produced. Since smaller genomic fragments generally clone preferentially, for the protocol optimization experiments in vector pBAD33 insert DNA was generally sheared using a pump speed that produced fragments centered around 10 kilobases (usually speed code 16) on the premise that the mean library insert size would be somewhat smaller than the mean fragment size. In later experiments, and for the construction of surface water library HMS186_25m discussed in more detail below, the pump speed was raised to speed code 15 in order to produce smaller fragments such that the bulk of the insert DNA fell in the 7-10 kilobase size range, which was observed to clone with higher efficiency. 5-10 µg of crude environmental DNA was selected as starting material. However, the designated size fraction of no more than 2.1 µg worth of this material was incorporated into the ligation reaction for any given library.

To prepare sheared insert DNA, fragments were treated with End-it DNA End-Repair (Epicentre, Madison, WI), which combines T4 DNA polymerase with T4 polynucleotide kinase to yield blunt, 5'-phosphorylated ends for cloning into dephosphorylated vector. This was followed by size-fractionation on a LMP agarose gel to remove fragments that were smaller than the desired insert size. Size-fractionation of

insert DNA was deemed necessary, although it results in significant loss of material, in order to minimize the effect of the preferential cloning of smaller fragments on mean library insert size. Since a favorable insert to vector molar ratio was found to be essential in the construction of these libraries, even more so than with smaller (1-3 kilobase) size inserts, maximizing the yield on preparation of limiting environmental insert DNA was a key area for protocol optimization. Concentration of the End-Repair product by standard ethanol precipitation improved the yield at this step. Due to the tendency of the Hydroshear to clog, environmental sample DNA was diluted to less than 100 ng/ μ L meaning sample volume after End-Repair was generally too large to fit in a single moderately sized gel well. Concentrating the sample to a volume of approximately 30 μ L allowed the sample to be conveniently loaded in one gel lane for size-fractionation. Several methods were evaluated for the purification of insert DNA from the size-fractionation gel. However, regardless of the method used, yields improved when the entire sample was concentrated into a single lane, run out only to the extent necessary, and the mass of agarose excised kept to a minimum. Limiting gel slices to between 100 and 200 mg allowed them to be processed in single reactions and gave good results.

Initially, End-Repair reactions were heat-inactivated then loaded directly onto a LMP agarose gel, without extraction and concentration. After the gel was run out, gel bands containing size fractions of interest were excised, and DNA was isolated using the Gelase gel-extraction kit (Epicentre, Madison, WI). By this method, polysaccharides in the molten agarose sample were enzymatically digested and nucleic acids were precipitated directly from solution. This method is used to extract 40 kilobase fragments of environmental DNA from LMP agarose gels for the preparation of CopyControl Fosmid libraries (Epicentre, Madison, WI) and was initially selected because it allows the recovery of DNA fragments larger than 10 kilobases in length, which is the upper limit on the commonly used Qiagen QIAQuick gel extraction kits (Qiagen, Valencia, CA). However, for five to 12 kilobase fragments, the efficiency of sample recovery was lower than with other available methods, and there was some question as to whether the

purified insert DNA would be sufficiently clean to perform well in the downstream blunt-end ligation.

Gel extractions in subsequent work were performed using the QIAEX II kit (Qiagen, Valencia, CA) or the QIAQuick kit (Qiagen, Valencia, CA). For sheared DNA with mean fragment sizes of approximately 10 kilobases, the Qiagen QIAEX II gel purification kit (Qiagen, Valencia, CA) was used. Like the QIAQuick column purification system, the QIAEX II kit utilizes the reversible binding of DNA to silica. However, the silica beads used in the QIAEX II method allow the recovery of a broader size range of DNA fragments, including those 10 kilobases or larger which do not elute well from the DNA binding columns of the QIAQuick kit (Qiagen, Valencia, CA). During the purification process, beads were resuspended by gently flicking the tubes, rather than vortexing, to avoid shearing of end-repaired insert DNA. Sample recovery was comparable between the two kits, and, for convenience, the QIAQuick kit (Qiagen, Valencia, CA) was used when mean fragment size was less than 10 kb, such as in the construction of the HMS186_25m Hawaii surface water library. Both of these kits gave improved insert DNA extraction efficiencies relative to the GELase system (Epicentre, Madison, WI) and yielded purified material that was adequate for library production.

Blunt-end ligations are considerably less efficient than those involving fragments with cohesive ends. Frequently, this obstacle is overcome by increasing the concentration of DNA or simply scaling up the ligation reaction (Ausubel et al. 1992), (section 3.16). In this case, however, the aim was to minimize the amount of DNA required for library production. Limited quantities of marine microbial DNA can be collected during each oceanographic research cruise, and by reducing the amount utilized for any given application, it is possible to characterize samples by a variety of culture-independent methods and therefore maximizes our ability to draw meaningful conclusions from the resulting data. Shearing followed by end-repair of sample DNA has the advantage over partial digestion methods (Olsen et al. 1986) in that it yields a more random distribution of fragments. However, this method results in blunt rather than cohesive ends, making high efficiency directional cloning impossible without the use of

linkers. To reduce religated vector background, it was necessary to treat the prepared vector with alkaline phosphatase enzyme to remove 5'-phosphates. This step, though generally necessary in blunt-end cloning, further reduces overall ligation efficiency (Ausubel et al. 1992), (section 3.16). An attempt was made to construct a library without first alkaline phosphatase treating the plasmid by ligating an excess of 8-10 kilobase insert DNA into blunt-end digested pBAD33, but the result, as expected, was primarily religated vector.

A number of protocol modifications were made to improve the efficiency of the ligation of sample DNA into the pBAD33 vector. Since the final concentration of ligation components may be too low for accurate spectrophotometric quantitation, an agarose gel with insert and vector side by side was run then stained with SYBR Gold (Molecular Probes, Eugene OR) to allow a more accurate determination of the concentration ratio, as well as serving as a useful quality check for both vector and insert preparation. Both insert to vector molar ratio and blunt-end ligation conditions, including the amount of enzyme present in the ligation reaction, were found to be critical to library yield, defined for the purposes of protocol optimization experiments by the rough metric of colony-forming units (cfu) per μg of crude DNA starting material. Although rapid ligation protocols are extremely convenient and work well for many cloning applications, Fast-Link DNA Ligase (Epicentre, Madison, WI) ligase was never successfully employed in this protocol. Even blunt-end digested pBAD33 religated to itself with relatively poor efficiency when this enzyme was used for the blunt-end ligation step. For all work discussed here, standard T4 DNA ligase (New England Biolabs, Beverly, MA) was used, and all ligations were performed overnight at 16°C . The use of T4 DNA ligase and long incubation times, rather than the Fast-Link DNA Ligase (Epicentre, Madison, WI) used for fosmid library construction, yielded dramatic improvements in cloning efficiency.

Insert to vector molar ratios between 2:1 and 4:1 were used for all library construction. Early attempts with higher 6:1 insert to vector molar ratios did not improve library quality. Lower ratios between 1:1 and 1.5:1, although often recommended for blunt-end ligation (Sambrook and Russell 2001), (CSH Protocols; 2006;

doi:10.1101/pdb.prot3921), yielded libraries of poor quality with a high proportion of religated vector background. Since the aim of this work was to produce high quality plasmid libraries with a large mean insert size while using as little valuable environmental DNA as possible, a 3:1 insert:vector molar ratio was determined to be optimal. The quantity of ligase included in the ligation reaction was also tweaked to improve efficiency, and 200 units of enzyme per 10 μ L reaction was selected for library construction. Higher quantities of ligase are sometimes recommended to increase the efficiency of blunt-end ligation (Ausubel et al. 1992), (section 3.16). However, increasing the amount of ligase used five-fold to 1000 units enzyme per 10 μ L ligation reaction decreased the yield on self-ligation of blunt-end digested pBAD33, estimated here in cfu per μ g vector DNA, by a factor of at least 20. Significantly lower quantities of ligase also decreased yield. As good results were achieved with the ligation protocol described in “Methods,” enzyme removal prior to electroporation, either by phenol extraction or spin-column purification, was not attempted, nor was the addition of polyethylene glycol to the ligation reaction (Pheiffer and Zimmerman 1983).

The presence of non-insert bearing clones in genomic or environmental genomic libraries can be a significant challenge in the construction of high-quality libraries. Although in principal dephosphorylation of the vector DNA should prevent this problem, in many cases this step is not sufficient to eliminate the presence of non-insert bearing clones or “background” in the finished library. A major advantage of fosmid libraries is that blunt-end ligation products that do not contain vector will not be packaged properly and therefore do not pose a problem in library construction. In addition, selection systems have been developed such that cells that contain empty vector do not survive (Bernard et al. 1994). The pBAD33 plasmid used in the early protocol optimization work has neither a selection against religated vector nor a visual screen to distinguish clones that carry empty vector, and library background was a significant problem.

To evaluate the success of size-fractionation and estimate the fraction of non-insert bearing clones present, a sample of clones selected from partial plating of newly constructed libraries was always checked for quality. Two possible methods for

verifying library quality are colony PCR and restriction enzyme digest. For both pBAD33 and pMCL200, primers are available to amplify the region of the plasmid containing the insertion site (Table S1), so colony PCR could be used to determine mean insert size. However, inserts of length greater than around 5 kilobases can be difficult to amplify using standard colony PCR techniques. In addition, amplification of insert DNA will not diagnose vector rearrangements or deletions. During protocol development, all libraries were checked for quality by restriction analysis of prepared plasmid DNA to look at insert size distribution and verify that clones carried full-length vector. By this method, clones containing insert that is equal in length to linear vector DNA, or clones that yield fragments of that length when digested, could be incorrectly scored. Nonetheless, it was a useful way to evaluate library construction.

Although transformation of self-ligated, alkaline phosphatase treated, blunt-end digested pBAD33 vector in the absence of insert yielded few or no colonies, as did non-ligated alkaline phosphatase treated prepared pBAD33 vector DNA, the majority of clones in the first library construction experiments contained what appeared to be religated vector. Manufacturer's notes accompanying the End-Repair enzyme mix (Epicentre, Madison, WI) state that T4 polynucleotide kinase is not completely heat-inactivated and recommend phenol-extracting the End-repair reaction if excessive no-insert background is observed to avoid kinase carry-over into the ligation reaction and rephosphorylation of vector DNA. This seemed unlikely, since the sample had been size-fractionated and purified from an agarose gel prior to ligation, but the advice was followed nonetheless. This modified protocol produced a better result, most likely because the subsequent precipitation reduced sample volume, which, as discussed above, improved yield on the agarose gel-purification step. Thus the step remained in the protocol although concentration of the insert DNA before size-fractionation, without phenol extraction, should be sufficient.

Protocol development proceeded under the assumption that empty vector background observed was due to incomplete dephosphorylation of vector DNA. Since a longer vector dephosphorylation reaction was not helpful, a very clean preparation of

pBAD33 vector DNA was produced by taking approximately 1 µg of SmaI blunt-end digested, alkaline-phosphatase treated pBAD33 plasmid DNA and subjecting it to an overnight self-ligation reaction with T4 DNA ligase at 16 ° C in order that any incompletely dephosphorylated vector would religate. Following ligation, the vector preparation was run out on a 1% agarose gel to separate linear (dephosphorylated) vector DNA from any religated or uncut plasmid DNA present. This step was effective in reducing the amount of empty-vector containing clones in the final library.

An additional problem encountered with libraries constructed in the pBAD33 vector was the presence of non-insert bearing clones that upon plasmid DNA extraction and digestion appeared to harbor a linear vector band shorter than 5.3 kilobases, the length of full-length pBAD33, suggesting some sort of vector truncation or rearrangement. During an experiment designed to test whether increasing the insert to vector molar ratio to 6:1 would lead to a greater fraction of insert-bearing clones, it was observed that a control reaction containing a 1:1 8-10 kilobase insert to vector molar ratio resulted in a library that appeared to be entirely religated vector. However, the ligation with 6:1 ratio produced roughly 1/3 clones that appeared to be religated full-length vector (~5.3 kb) and 2/3 that appeared to harbor non-insert containing vector of approximately 4 kilobases in length, as evaluated by digesting plasmid DNA with SacI and XbaI restriction enzymes intended to excise any cloned insert. Single digests of these DNA preparations suggested that SacI linearized the short vectors, while the XbaI site may be missing. Apparently truncated vector that were shorter than 4 kilobases or which appeared to contain some insert were also observed, but, in general, clones with inserts of the desired size tended to also exhibit full-length vector bands when digested with SacI and XbaI. This result seemed to suggest that the pBAD33 cloning vector may not stably maintain 8-10 kb inserts in the *E. coli* strain DH10B used for these experiments, but these clones were never sequenced or further investigated.

Library construction in pBAD33 with samples of stock lambda DNA, incorporating the protocol optimizations discussed above, did eventually yield greatly improved results. Beginning with 5 µg of lambda stock DNA sheared into 8-12 kilobase

fragments and ligated into prepared pBAD33 vector, approximately 40,000 clones could be produced per μg of crude insert DNA (Table 1). Eight randomly selected clones were subjected to restriction analysis. Seven of these contained insert with a mean insert size of 6.4 kilobases, while one contained truncated vector and no insert. An environmental genomic library of this quality would be adequate for many functional screening applications, provided that the larger insert clones could be stably propagated. However, when the protocol was attempted with an environmental sample, quality declined significantly. It was not clear whether this was due to sample quality or a poor yield on sample preparation that resulted in an insert to vector molar ligation ratio well below the target ratio of 3:1. However, switching to an alternate cloning vector eventually circumvented many of the difficulties encountered with libraries constructed in plasmid pBAD33.

Library Construction and Protocol Optimization in Plasmid pMCL200

Subsequent library construction, including the surface water library HMS186_25m arrayed for phenotypic screening, generally made use of the plasmid cloning vector pMCL200. This cloning vector is used by the JGI for the construction of mid-size insert genomic libraries, so it was believed to be suitable for the cloning and maintenance of 8-10 kilobase inserts. The vector was selected because, like pBAD33, it is a mid-low copy plasmid, and the multi-cloning site is under the control of an inducible promoter. pMCL200 has a pACYC derived p15A origin of replication so is maintained at approximately 20-30 copies per cell (Lutz and Bujard 1997), and it carries a chloramphenicol resistance gene. The multiple-cloning site is under the control of the IPTG inducible wild type *lac* operon promoter, and the vector has the *lacZ α* gene to facilitate selection of insert containing clones (Nakano et al. 1995). Moderately low copy number combined with the relatively strong inducible *lac* promoter was intended to provide sufficient expression for functional screening of the library, while minimizing the

additional cloning biases associated with high plasmid copy number and elevated expression levels.

The *E. coli* host cell strain Top10-F' (Invitrogen, Carlsbad, CA) was selected for use with vector pMCL200. This strain is very similar to the commonly used *lacI* cloning strain DH10B, which was used for all pBAD33 libraries, but harbors an F' extra-chromosomal element carrying a tetracycline resistance marker and a *lacI^q* allele. The presence of the *LacI^q* allele allows this host strain to make sufficient *lac* repressor such that expression of genes carried on insert DNA from the pMCL200 P_{LAC} promoter are largely repressed in the absence of the inducer molecule, IPTG. Top10-F' cells, and the closely related MC1061-derived strain DH10B used for pBAD33 test libraries, poses a number of features that make them suitable as host strains for genomic libraries. Both are DNA repair deficient *recA1* strains, which reduces the likelihood of recombination events and enhances library stability, and both lack a suite of genes involved in methylation-dependent restriction of foreign DNA. In addition, both carry chromosomal *lacZ* genes with partial deletions to permit blue/white screening by alpha-complementation, *endA1* alleles for cleaner DNA preparations, and *deoR* to facilitate the uptake of large plasmids (Woodcock et al. 1989), (Grant et al. 1990).

To confirm that the *lacI^q* allele in Top10-F' cells provides sufficient *lac* repressor to block expression from the P_{LAC} promoter, pMCL200 plasmid DNA was transformed into Top10-F' *E. coli* cells. As a control, transformed DH10B cells were similarly transformed and plated onto LB plates supplemented with 12.5 µg/mL chloramphenicol and 40 µg/mL X-gal. After overnight incubation at 37° C, all colonies on these control plates appeared blue, as expected given that DH10B cells do not express *lacI* repressor. Transformed Top10-F' cells were plated onto LB plates supplemented with 10 µg/mL tetracycline, 12.5 µg/mL chloramphenicol and 40 µg/mL X-gal, with and without 0.5 mM IPTG inducer to induce gene expression from the P_{LAC} promoter. In the absence of inducer, all colonies were wild type in color, indicating that expression of β-galactosidase protein from the P_{LAC} promoter was largely repressed. Cells plated in the presence of 0.5 mM IPTG appeared blue confirming that the addition of this concentration of inducer

was sufficient to allow significant expression of β -galactosidase from the P_{LAC} promoter. Commercially prepared electrocompetent cells were selected to maximize the efficiency of library transformation.

The preparation of cloning vector pMCL200 for library construction differed slightly from that of pBAD33 since it was prepared strictly according to the JGI protocols provided with the plasmid. In the isolation of plasmid DNA, pMCL200 DNA was not subjected to CsCl density gradient centrifugation purification, while pBAD33 did receive this additional purification step. Although the Hispeed Plasmid MidiKit (Qiagen, Valencia, CA) used here for isolation of pMCL200 plasmid DNA gave a reasonably clean product, some higher molecular weight DNA, believed to be *E. coli* chromosomal DNA, was visible when the product was run out and stained on an agarose gel. CsCl equilibrium density gradient centrifugation could be used to reduce *E. coli* chromosomal DNA contamination in the plasmid DNA preparation. For many functional screening applications, depending on the specific screen and host cells employed, low-level contamination of environmental libraries with *E. coli* chromosomal fragments may not be problematic. However, since there is no drawback to doing so, aside from some additional time in the vector preparation step, it would be worthwhile to add the CsCl density gradient purification step to the vector preparation protocol.

The vector dephosphorylation step also differed somewhat, both in the type of phosphatase enzyme and the reaction conditions. For the preparation of pBAD33, Apex Heat Labile Alkaline Phosphatase was selected and used according to the manufacturer's directions, which call for 5 units of enzyme per μg of plasmid DNA and a 10 minute incubation at 37° C. Blunt-end digested pMCL200 was dephosphorylated with shrimp alkaline phosphatase (SAP), according to the JGI protocol, using almost three times the amount of enzyme called for in the product manual (Roche Applied Science). Dephosphorylated pMCL200 was not treated with ligase, to ligate incompletely dephosphorylated vector, prior to gel-purification. The fraction of pMCL200 library clones exhibiting a blue color when plated in the presence of IPTG inducer and X-gal, indicating that they express a functional β -galactosidase enzyme and thus do not contain

insert, was generally in an acceptable 10-20% range for this protocol. However, the vector self-ligation step was helpful in reducing religated vector background in the pBAD33 libraries, so adding this step to the pMCL200 vector preparation protocol could potentially further reduce the number of empty vector clones in the finished library.

Preliminary tests of the pMCL200 cloning vector with lambda DNA insert, prepared according to the protocol optimized here for environmental samples, showed promising results, so a mid-size insert environmental genomic plasmid library was constructed in this vector using a mixed-depth sample of water column microbial DNA collected on a sampling expedition at the Bermuda Atlantic Time-series Study (BATS) field site during October of 2006. The purpose of this experiment was to verify that mid-size insert libraries of acceptable quality could be produced by the current protocol before using any of the limited quantity of HOT surface water DNA available, as well as to compare the quality of libraries constructed in pMCL200 to those in pBAD33. A sample containing roughly 9 μg of crude marine microbial DNA was sheared on the Hydroshear (Genomic Solutions, Ann Arbor, MI) at speed code 15 for 25 cycles producing fragments centered around 8 kilobases in length. Since this sample was split between several test libraries, the results for each test library represent the use of less than 1.5 μg of starting material. End-repaired fragments were run out in the usual way on a 1% LMP agarose gel, and bands containing the 8-10 kilobase fraction, which constituted the majority of the sample, as well as a secondary 4-8 kilobase slice, were excised for gel-purification. All ligation components were quantitated visually on a gel to allow the calculation of approximate quantities and molar ratios (Fig. 1A). The two BATS DNA size fractions were ligated to both pBAD33 and pMCL200 vectors, along with a control sample of lambda insert DNA that had been successfully cloned in pBAD33 previously.

Although variations in component DNA length and concentration make it impossible to draw rigorous conclusions about cloning efficiency, the purpose was simply to gain a practical sense of protocol outcomes with different size fractions of environmental DNA in these cloning vectors. All ligation reactions were transformed into electrocompetent DH10B cells, as this experiment was performed prior to the

decision to use Top10-F' cells for library construction. The most dramatic result was variation in library yield and quality between the two vectors. Sample libraries constructed in pMCL200 produced dramatically more clones than those constructed in pBAD33 with identical inserts, and the pBAD33 libraries attempted with environmental DNA samples essentially failed. Plates for pBAD33 test libraries had only slightly more colonies than the no-insert pBAD33 vector control. Six clones from each size fraction, as well as for the control lambda library, were checked by restriction analysis. No pBAD33 clones checked for either size fraction BATS libraries contained full-length vector with insert. Plasmid DNA preparations appeared to consist of either religated vector, empty truncated vector, or, in one case, truncated vector with insert. Only 2 of 6 lambda clones in pBAD33 contained insert.

Results for the pMCL200 BATS sample libraries were much better. The library constructed in pMCL200 with the smaller size fraction of insert produced three times as many cfu per μg crude starting material than that with the larger size fraction and had a lower percentage of blue colonies when plated on X-gal containing plates. 22% of colonies were blue compared to 32% of colonies for the larger size fraction, including very faintly blue colonies that were only visible after more than one week. This was consistent with previous observations of reduced cloning efficiency with larger inserts. As with the pBAD33 test libraries, six clones from each library were selected and grown up to spot check library quality. However, only white colonies were selected from the pMCL200 sample libraries; at the time colonies were picked, between 80% and 90% of colonies appeared white. As observed previously, mean insert sizes were several kilobases smaller than the apparent mean fragment size of the sheared DNA (i.e. 6 kilobases for the "8 kilobase" library and 2.4 kilobases for the "6 kilobase" library). All clones checked in the pMCL200 control lambda library contained insert. For the two environmental genomic test libraries, roughly 75% of clones contained insert (Table 1).

Based on these results, the decision was made to prepare libraries of this type from a Hawaii surface water sample collected in October of 2006 during a research cruise to HOT field station ALOHA (22° 45' N, 158° W). This sample is part of a set of

microbial DNA and RNA samples collected on this cruise from specified depths in the water column that are currently being analyzed by a variety of environmental genomic methods. Hawaii Ocean Time-series station ALOHA is an oligotrophic oceanographic field site in the North Pacific Subtropical Gyre that has been studied extensively since 1988 by biological, chemical, and physical oceanographers (Karl and Lukas 1996). The microbial and phytoplanktonic surface water populations have been well-characterized during this time using traditional microbiological and oceanographic methods (Karl et al. 2001), and recent work has taken a culture-independent approach based on sequence analysis of environmental genomic fosmid libraries prepared from microbial DNA isolated at key points in the water column (DeLong et al. 2006). Since mid-size insert plasmid libraries require a significant quantity of DNA to produce a good result, libraries of this type were only constructed from the sample with the highest DNA yield, isolated from the 0.2-2 μm microbial size fraction collected in the photic zone at a depth of 25 meters.

An excess of crude surface water DNA was available, so 9 μg of this was used as starting material for library preparation. A speed code of 15 was selected on the Hydroshear (Genomic Solutions, Ann Arbor, MI) for the random shearing of insert DNA, although previous experience had shown that this pump speed produced fragments with a length distribution clustered around 8 kilobases, the target mean insert length for this library, so the mean insert length in the final library would therefore be somewhat shorter than desired. Fragments in the optimal 8-10 kilobase range had been successfully cloned in both vectors, but no environmental genomic library of this type had been prepared with a mean insert size in this range. The protocol was repeated exactly as in the BATS environmental sample library. Libraries were prepared from two size fractions. A gel slice containing the bulk of the sheared DNA, the length of which fell between 7 and 10 kilobases, was excised for the primary library, and the smaller fragments, which fell in the 4-7 kilobase range, were also excised from the gel for the construction of a second backup library. 25% of the size-fractionated and gel-purified insert DNA for each size fraction was incorporated into the ligation reactions, and 12.5% of each ligation product

was transformed into electrocompetent DH10B cells for an initial assessment of library quality and yield. Aliquots of both libraries consisting of 1% of each transformation product were plated onto LB plates supplemented with 40 mg/mL X-gal and the appropriate antibiotics. For both size fractions less than 10% of colonies appeared blue, indicating that they likely contained religated pMCL200 plasmid, although after several days of storage at 4° C additional lighter blue colonies were visible. A rough calculation, including both blue and white colonies to allow a direct comparison to pBAD33 results, indicates a yield of 198,000 clones per µg of crude DNA starting material, which is more than 5 times the best yield obtained with pBAD33.

A portion of the larger fragment size ligation product was then transformed into *lacI^q* Top10-F' *E. coli* cells for colony picking and library characterization. The transformation efficiency and fraction of blue colored colonies in the presence of 0.5 mM IPTG inducer were indistinguishable from the results with DH10B cells. This larger insert Hawaii surface water library is referred to as HMS186_25m (for Hawaii mid-size library, cruise 186, 25m sample). Before plating and picking the bulk of the library, a sample of wild type colored clones were grown up and plasmid DNA prepared for restriction-enzyme digest analysis to confirm insert size and library quality. 20 colonies were checked in this way, 14 of the larger insert library, HMS186_25m, and 6 of the smaller insert library. All of the smaller insert clones contained insert, with a mean insert length of 4.6 kilobases. For HMS186_25m, 11 out of 14 clones checked contained insert, and the mean insert size was 5.2 kilobases, including analysis of clones in both cell strains (Table 1, Fig. 1B). Two of those that did not contain insert appeared to be truncated vector, potentially explaining the lack of β-galactosidase expression and subsequent white color, and one to be full-length empty vector.

Library HMS186_25m in Top10-F' cells was plated onto large LB plates supplemented with chloramphenicol, tetracycline, X-gal, and IPTG and stored for several days at 4° C to allow blue color to develop. Robot settings were adjusted to favor the selection of wild type colored colonies over those with a blue color, but it was not possible to entirely avoid the picking of light blue colored colonies. In addition, since

quality checking of HMS186_25m revealed that roughly 20% of wild type colored colonies appeared to harbor religated or truncated vector, the blue/white screen for the presence of insert may be of limited utility. Sixty-one 384-well plates were picked for a total of 23,424 clones or approximately 50% of the colonies obtained from the transformation of 1 μ L of the ligation product. This quantity of clones was considered adequate to begin to evaluate the utility of the library for phenotypic and functional screening. To facilitate future screening or transfer to alternate host strains, library HMS186_25m was pooled and 1.5 L was grown up in liquid media for plasmid DNA preparation and purification by CsCl equilibrium density gradient centrifugation.

Colony Phenotype Screening of Libraries HMS186_25m and HF186_25m

When environmental genomic libraries constructed in copy-control fosmid vectors are plated on solid media under conditions of induction to high copy, expression of genes carried on cloned environmental DNA frequently imparts non-wild type coloring to colonies of host *E. coli* cells. By stamping libraries from glycerol cell stocks onto solid media overlaid with opaque membranes, entire libraries can be screened for color under various growth conditions. This basic phenotypic screen was selected as a means of assessing the potential of HMS186_25m and similar libraries for use in gene expression-based screens for a number of reasons. Fosmid libraries constructed from similar environmental samples had yielded positive results using this phenotypic screen previously. Like functional screening, colony phenotype screening relies on expression of cloned genes by host cells, so a library that yielded many positive hits in a colony phenotype screen would likely be useful for functional screening. Also, equipment was available to automate the arraying process making the screen relatively straightforward to carry out. Finally, a fosmid library, called HF186_25m (for Hawaii, fosmid library, cruise 186, 25m sample), constructed from the same sample of photic zone DNA was also available making a direct comparison between the two possible.

The first set of HMS186_25m colony phenotype arrays were grown up on solid media containing X-gal substrate so that the success of blue/white selection in automated colony picking could be evaluated and induction of gene expression with IPTG could be confirmed. Forty-eight of the 61 plates were arrayed, 24 per membrane, in six replicates. One set was grown up then denatured, crosslinked to the membrane, and stored for possible nucleic acid probing experiments. The other four sets were grown up with and without IPTG inducer at 30° C and 37° C, as described. The presence of blue clones on the plates with inducer and, after several days at room temperature, light blue clones on the plates without inducer verified that automated blue/white colony selection was imperfect and also interfered significantly with scoring the results of the screen. Intensely colored blue colonies made it difficult to visually pick out subtle non-wild type colors, and light blue clones due to low levels of β -galactosidase activity were indistinguishable from colonies that may have had a bluish or greenish tint due to expression of cloned environmental genes. However, one distinctly purple colony was identified, as well as several other colonies that appeared to have a subtle yellow or gray tinge, so the library was re-arrayed on plates that did not contain X-gal to allow better scoring.

Many non-wild type colony color phenotypes were not visible after the overnight 30° C or 37° C incubation and appeared only after a week or more at room temperature, so plates were scored after the initial incubation and then moved to dark, room temperature storage and scored periodically. The morning after HMS186_25m arrays were printed, only the 37° C incubated plates had grown up enough for colony color to be evaluated. The purple colony observed previously was visible on the plate with IPTG, as was one pink colony from which the color appeared to diffuse and a few faintly gray-brown colonies. After 24 hours at room temperature, four colonies with non-wild type coloring were visible, but results varied somewhat between plates incubated at 30° C and 37° C, as well as with and without IPTG. For the purpose of this discussion, clones of interest will generally be referred to by their plate number and well coordinates, e.g. “11-H17.” The two most distinct hits observed were the previously mentioned purple colony

(11-H17) and a blue colony (28-P8). 11-H17 was visible on all four plates, though color varied somewhat being pinkish on the plate without IPTG incubated initially at 30° C, grayish on the plate without IPTG incubated at 37° C, and purple on the two plates with IPTG inducer. 28-P8 was distinctly blue on all replicates but brighter in color on both plates initially put at 30° C. Also visible, but only on the 30° C incubated plate with IPTG, was a colony with a pinkish-brown tinge (29-L8). The pinkish clone (10-M20) noticed after the initial overnight incubation had faded in color and was difficult to detect (Table 2).

Three days later when plates were scored again, the phenotypes had changed somewhat. Colors were generally more pronounced on plates with initial 30° C incubations, so those plates are discussed here. 10-M20 colonies no longer had a distinct phenotype. An additional grayish-brown colony was visible (11-D16) on the plate without IPTG; this clone did not grow up at all in the presence of IPTG. 29-L8 could now be picked out with and without inducer. There were several other colonies that had a slight gray or brown tinge but were not included in this analysis, as the phenotype was extremely subtle. 11-H17 and 28-P8 were both visible with and without inducer and, in fact, were subtly brighter without, though this observation is not consistent with later characterization of these clones and may be due to reduced growth on plates supplemented with IPTG. Growth was generally normal but less consistent and slightly reduced on these plates relative to those lacking IPTG, potentially due to the burden of expression off the strong P_{LAC} promoter. This trend could also be simply due to plate-to-plate variability in array printing. From visual inspection of single colonies, it is not possible to reliably distinguish the effects of gene expression level per cell and extent of growth on colony color phenotype. These preliminary results, therefore, are relevant to the usefulness of induction with IPTG for this type of phenotypic screen, but they indicate little about expression of environmental DNA from the P_{LAC} promoter in this library.

The HMS186_25m mid-size insert plasmid library arrays were also scanned for fluorescence after one week of incubation at room temperature to detect the presence of

expressed genes conferring fluorescent phenotypes on host *E. coli* cells (Table 2). The set of plates, with and without IPTG, that had their initial overnight incubation at 30° C were selected for scanning, as this set generally had brighter colony color phenotypes. Scans employing all three available excitation wavelengths, 473 nm, 532 nm, and 635 nm, were performed, as fluorescent colonies may fluoresce more or less at different wavelengths. The 473 nm scans, however, showed very little signal. Overall, the 635 nm scans were the most useful, although at least one colony exhibited more fluorescence when excited at 532 nm. The set of colonies that showed up on the fluorescence scans differ somewhat from the set of colonies that were picked out visually. Of the five colonies that were identified visually based on non-wild type colony color, the three with the most distinct phenotypes were also identified by fluorescence. The two colonies that had faint pink or pinkish-brown coloring were not picked up on any of the fluorescence scans. Five additional colonies, which visually had no distinct colony phenotype, were also identified on these scans; of these, two were strongly fluorescent. Three of the five exhibited greater signal on the membrane grown up in the presence of IPTG, but the remainder were not obviously affected by the presence of inducer. This was consistent with the presence of non-wild type colored colonies on the array plates grown up without IPTG and seems to suggest either incomplete repression of the P_{LAC} promoter on vector pMCL200 or some amount of gene expression from promoters present in the insert sequences.

For comparison, 48 plates of fosmid library HF186_25m, which was constructed from the same DNA sample as mid-size insert plasmid library HMS186_25m, were also arrayed for colony phenotype screening. All HF186_25m membranes were grown up in the presence of L-arabinose for induction to high copy-number, as very little color has been observed when environmental genomic fosmid libraries are plated at low copy-number (Martinez, unpublished data). Membranes arrayed with the HF186_25m fosmid library clones exhibited several brightly colored sets of colonies after overnight incubation. After 24 hours at room temperature, these plates were fully scored for colony color phenotype (Table 3). Between the two membranes, which were printed with cells

from 48 plates containing 18,432 total fosmid clones, 11 clones had distinct colony color phenotypes. Six of those were quite bright, including two orange, one blue, one yellow, and two blue-green. An additional 20 clones gave rise to colonies with subtle tinges of color in the grey-yellow range. By way of a general observation, colony color phenotypes were more distinct than those for clones identified in the HMS186_25m arrays. In total, approximately 0.17% of the HF186_25m fosmid clones exhibited non-wild type colony color, which works out to roughly 5 times the percentage of HMS186_25m clones that exhibited non-wild type colony color, taking into account the set of HMS186_25m plates, numbers 49-61, which were arrayed only in the presence of retinal as discussed below. Given that the average insert size on the fosmid clones is more than 7 times larger than the average insert for the mid-size insert plasmid library, that figure may not be surprising. However, with minimal data concerning either gene expression off promoters contained in the cloned environmental DNA sequences or the relative amounts of transcription into the multi-cloning sites of these vectors under the growth conditions employed, this rough comparison is of some practical interest but is included mainly to give some context to the HMS186_25m phenotypic screening results.

Screening with Exogenous Retinal

A feature of fosmid environmental libraries is that clones carrying proteorhodopsin genes, along with the set of 5 genes required for retinal biosynthesis in *E. coli*, can often be identified when arrayed on solid media by an orange colony color (Martinez et al. 2007). Adding exogenous retinal to proteorhodopsin expressing clones will cause them also to appear orange (Beja et al. 2000). Although proteorhodopsin genes are often located in close proximity to retinal biosynthetic operons in the genomes of marine photic zone microbes, the genome fragment required for an environmental clone to both express proteorhodopsin and allow retinal biosynthesis in *E. coli* host cells is generally around 8 kilobases in length (McCarren and DeLong 2007). Thus library HMS186_25m does not contain sufficiently large inserts to make visually screening arrays for proteorhodopsin-

expressing clones worthwhile without the addition of exogenous retinal. The addition of exogenous retinal to liquid cultures of *E. coli* cells, followed by growth and pelleting of cells, is a straightforward way to visualize the characteristic orange color of proteorhodopsin-expressing clones but is not well suited to high-throughput screening. As entire libraries can be conveniently arrayed onto solid media overlaid with opaque membranes for phenotypic screening as described above, this is clearly a better pipeline for phenotypic screening of proteorhodopsins in small or mid-size insert environmental genomic libraries. Adding exogenous retinal to solid media, however, has been complicated by the low solubility of retinal in aqueous media and its failure to diffuse through the membrane overlays used to enhance visualization of colony color (Martinez et al. 2007).

In order to screen library HMS186_25m for proteorhodopsin genes, a method was developed to apply all-*trans* retinal directly to the membranes. Two clones that carry proteorhodopsin genes under the control of the arabinose inducible P_{BAD} promoter on pBAD-TopoTA plasmids (Invitrogen, Carlsbad, CA) were used as positive controls in developing the screen. Both yielded distinctly colored cell pellets when proteorhodopsin expression was induced with 10 μ M all-*trans* retinal in log-phase liquid cultures; clone pBAD-31A8 produced a pink colored pellet, while clone pBAD-S14PR produced an orange shade. As retinal is quite insoluble in aqueous solution, all retinal stock solutions were made up in 95% ethanol. Dipping the membrane in retinal solution was tested, as was spreading retinal solution both under and over the membrane and spraying retinal solution at various concentrations directly onto the membrane. Spreading retinal under the membrane failed to produce color in cells streaked out in the presence of 0.02% arabinose to induce proteorhodopsin expression. Although some colonies of cells streaked out on plates with retinal solution spread over the membrane had good color, application by this method was uneven, so this approach was not promising for library arrays. Dipping proved to be quite messy, and yellow retinal solution dripped across the membrane as it was removed from solution causing streaky, uneven application. Spraying the membranes with a fine mist of retinal solution was much more effective.

Several retinal concentrations were tried, ranging from 10 μ M to 5 mM, and, in each case, the two positive controls and a negative control consisting of cells harboring empty puc19 plasmid were streaked out with induction and incubated overnight at 30° C. Colonies of positive control cells appeared wild type in color on membranes sprayed with 10 μ M all-*trans* retinal, and color was too faint on membranes sprayed with 50 μ M or 100 μ M retinal to be reliably detected on library arrays. Plates overlaid with membranes sprayed with 500 μ M, 1 mM, and 5 mM retinal all imparted good color to colonies of both proteorhodopsin-expressing clones. However, while the 500 μ M and 1 mM retinal sprayed membranes were light yellow in color, the 5 mM sprayed membrane appeared bright yellow. Surprisingly, the bright yellow background color did not make it difficult to observe the orange or reddish color of proteorhodopsin-expressing colonies. In fact, cells streaked out with this high concentration of retinal on the membrane surface exhibited the most dramatic orange color phenotype and were easily identifiable on the yellow background (Fig. 2). Although it is possible that this background color would make it difficult to detect a more subtle orange colony phenotype, a 5 mM retinal solution spray was selected for the screening of library HMS186_25m for proteorhodopsins.

The complete HMS186_25m library was arrayed with and without 1 mM IPTG onto membranes coated as described with all-*trans* retinal. After overnight incubation at 30° C, colored clones identified previously were visible, but no orange-tinted colonies were observed. In addition, two non-wild type colored clones on the membrane printed from plates 49-61, which had not been arrayed previously, were visible on arrays grown up with and without IPTG. The 50-D15 duplicate colonies were brown in color, and the 54-F8 colonies were dark blue-purple. Plates were moved to room temperature and allowed to remain for one week before being scored again. At this time one light orange clone, 54-B7, could be picked out, also on both the induced and non-induced replicates. The faint orange phenotype was confirmed by retransforming the clone and streaking it out onto retinal-coated membranes. However, efforts to confirm the retinal-dependence of this phenotype have been inconclusive because subtle orange shades are actually more

noticeable on a yellow than a white background, and end-sequences did not reveal a proteorhodopsin gene.

Sequence Analysis and Characterization of Selected Clones

Several clones exhibiting colony color phenotypes of interest, from both the HMS186_25m plasmid library and the HF186_25m fosmid library, were selected for further investigation. Prior to sequence analysis, clones were streaked out, retransformed, and their colony phenotype confirmed, as described in “Methods.” All colony phenotypes observed in the library arrays that were selected for analysis stayed true when clones were retransformed and tested. The sequencing strategy adopted varied somewhat between the two library types, but the goal in all cases was to gain some insight into the identity of genes responsible for the observed phenotype and the phylogenetic identity of the microbe from which the clone derived. Fosmid clones from library HF186_25m were end-sequenced to provide some clues as to phylogeny, and transposon mutagenesis was performed. Resulting colonies for which the distinctive color phenotype had been knocked out or modified were picked and sequenced in both directions from the transposon insertion with the intention of identifying the gene or genes that conferred colony color in the original clone. Although color may be knocked out by transposon insertions outside the specific genes responsible, for example by interfering with induction of fosmid pCC1FOS to high copy, examination of insertion site sequences from 4-6 transposon insertion mutants was sufficient to identify genes responsible for color in the fosmid clones discussed here. HMS186_25m plasmid clones of interest were also end-sequenced, but rather than performing transposon mutagenesis, primer walking was used to obtain additional sequence information. This decision was made based on the premise that for mid-size insert plasmid libraries with roughly 6 kilobase inserts, simply end-sequencing would be likely to provide clues as to the genes responsible for a phenotype of interest. If additional sequence was desired, clones could then be fully sequenced by two or three sequential rounds of primer walking. This

approach was adequate for some preliminary investigation of a small number of clones of interest but would not be efficient for any type of large-scale screening project. Also, end-sequences were not always informative, so a more effective pipeline might include transposon mutagenesis or sub-cloning approaches to obtain full insert sequences of clones identified by functional or phenotypic screening.

Clone 11-H17 in library HMS186_25m was first identified by a purple colony phenotype when the library was arrayed onto membranes. The phenotype was stable when plasmid DNA was prepared and retransformed into both DH10B and Top10-F' *E. coli* cells. Colony color appeared somewhat darker on the array grown up in the presence of induction with IPTG, so prior to sequence analysis, cell stocks of 11-H17 in Top10-F' cells were streaked out onto opaque membranes overlaying LB plates, supplemented with appropriate antibiotics, with and without 0.5 mM IPTG. Plates were incubated overnight at 37° C and then moved to room temperature and allowed to grow at room temperature for two days. Even after the overnight incubation, cells grown with induction had taken on a much darker purple shade, while cells grown without induction appeared pink (Fig. 3A). Color intensified as cells grew at room temperature, with single colonies and edges of dense streaks exhibiting more intense coloration and some appearance of color diffusing from colonies.

Plasmid DNA prepared from clone 11-H17 was end-sequenced, and 3 rounds of primer walking were performed. The sequence was not closed, but approximately 5 kilobases of sequence was obtained, 2.5 kilobases from each end reading into the insert. Both reads indicated that the sequence derived from a marine Gammaproteobacterium. The forward sequence appeared to contain two genes, one with homology to uroporphyrin III C-methyltransferase, which is involved in biosynthesis of cobalamin, and one with homology to glycosyltransferases. Both of the coding regions present on this forward read have greater than 75% amino acid identity to genes in *Alteromonas macleodii* (deep ecotype) that lack meaningful annotation. The reverse read contained a putative diguanylate cyclase gene and a nucleotidase-like sequence (Fig. 3B). The insert in clone 11-H17 is oriented such that the uroporphyrin III C-methyltransferase-like gene

is directly downstream of the P_{LAC} promoter, and it is this gene that is believed to be responsible for the purple colony phenotype. Uroporphyrin III C-methyltransferase genes, notably *cobA*, have been cloned previously and are known cause host cells to exhibit red fluorescence (Sattler et al. 1995). Over expression of this gene in many cell types leads to accumulation of sirohydrochlorin and trimethylpyrrocorphin, which fluoresce red when subjected to UV excitation, and *cobA* from the bacterium *Propionibacterium freudenrochii* has, in fact, been employed as a transcriptional reporter (Wildt and Deuschle 1999). In addition to exhibiting an IPTG-inducible purple colony phenotype, 11-H17 colonies were detected on fluorescence scans of HMS186_25m arrays, with more pronounced signal upon excitation at 532 nm than at 635 nm.

A clone exhibiting a blue colony phenotype from library HMS186_25m, clone 28-P8, was also investigated further. When cells were streaked out from glycerol stocks, no difference was observed between cells grown with and without 0.5 mM IPTG. Both with and without induction, cells were light blue in color, and the blue color diffused from cells causing the white membrane surrounding colonies to also appear light blue (Fig. 4A). Full sequencing of this clone revealed three coding regions spanning virtually the full insert. However, these coding regions are positioned on the opposite strand relative to the pMCL200 P_{LAC} promoter, likely explaining the IPTG-independence of the colony color phenotype. All three encode genes potentially involved in iron binding and transport, with amino acid identity between 50% and 75% to genes in *Gammaproteobacterium Alteromonas macleodii* (deep ecotype) including the permease component of an ABC-type heme transport system, the ATPase component of an ABC-type heme transport system, and an uncharacterized metal-binding protein with homology to the putative heme iron utilization protein HugZ (Mourino et al. 2004), (Fig. 4B).

Since iron concentrations are extremely low in seawater and may limit primary production under some circumstances (Boyd et al. 2000), (Street and Paytan 2005), marine microbes have evolved a variety of iron uptake systems. Many marine bacteria are known to excrete low molecular weight iron chelating molecules known collectively

as siderophores (Guerinot 1994), (Martinez 2000), and siderophore receptors in Gammaproteobacteria, including *Alteromonas macleodii*, have been characterized by binding assays with purified membrane proteins (Armstrong et al. 2004). Although iron chelators in bacteria are generally small molecules rather than peptides and synthesis may involve many genes (Guerinot 1994), the three open reading frames had good homology to iron uptake-related genes and the blue color appeared to diffuse out from colonies, so an attempt was made to characterize clone 28-P8 by streaking cells out onto siderophore indicator CAS plate with glycerol added as a carbon source. CAS plates contain the blue dye chrome azurol S/iron(III)/ hexadecyltrimethylammonium bromide. When cells excrete a strong iron-chelating siderophore, iron (III) is sequestered from the dye and a yellow-orange halo is visible (Schwyn and Neilands 1987).

In preparation, 28-P8 cells were initially streaked out onto a LB plate with chloramphenicol, tetracycline, and 1 mM IPTG, and negative control Top10-F' cells were streaked onto a LB plate, also with tetracycline and IPTG. 28-P8 and negative control cells were then streaked out side by side from the starter plate onto the CAS plate pre-spread with 1 mM IPTG and incubated overnight at 30° C. Wild type *E. coli* produce two types of siderophores, aerobactin and enterobactin, and results were inconclusive due to high levels of siderophore activity in the host *E. coli* cells (Guerinot 1994). The assay was not repeated in a strain of *E. coli* cells with reduced siderophore activity (Luke and Gibson 1971), as the sequences annotated on clone 28-P8 did not have homology to genes known to be involved in siderophore biosynthesis.

Arrays of fosmid library HF186_25m under copy-up growth conditions also revealed the presence of a clone, 22-F8, exhibiting a blue colony phenotype similar to that of plasmid clone 28-P8. The phenotype was not visible when cells were streaked out on membrane overlays in the absence of L-arabinose, and blue color took several days to develop, after overnight incubation at 30° C followed by incubation at room temperature, when cells were streaked out with 0.02% L-arabinose (Fig. 5). Transposon mutagenesis, as well as end-sequencing, of this clone was performed to see if the genetic basis of the phenotype was similar. Both end sequences indicated that clone 22-F8, like the two

HMS186_25m plasmid clones described above, derived from a Gammaproteobacterium. Sequence analysis of transposon insertion mutants of 22-F8 with wild type colony phenotypes revealed that blue colony color was, in fact, likely due to the presence of a ring-hydroxylating dioxygenase gene, which was interrupted in three of the four transposon insertion mutants analyzed. When expressed in *E. coli*, these genes convert indole to indigo, causing colonies to take on a blue color (Ensley et al. 1983), (Drewlo et al. 2001). In the environment, these enzymes are involved in the first step of aerobic catabolism of aromatic hydrocarbons (Eaton and Chapman 1995), and they have been observed previously in environmental genomic libraries prepared from soil samples (MacNeil et al. 2001).

Color screens of fosmid library HF186_25m identified two intensely orange colored clones, 13-D19 and 30-B18 (Fig. 5). The bright orange shade of these colonies was uncharacteristically vivid for proteorhodopsin clones. Fosmid 30-B18 even exhibited a faint orange colony color even when streaked out without arabinose on white membrane overlays. EZ-Tn5 <KAN-2> (Epicentre, Madison, WI) transposon mutagenesis was performed on both fosmids. When products were transformed into Epi300 host *E. coli* cells and plated for color onto selective plates with 0.02% arabinose, approximately 10% of colonies had lost their bright orange color and were either wild type or yellow in color. Partial sequence analysis of both yellow and wild type insertion mutants revealed the presence of genes involved in carotenoid biosynthesis. Carotenoids are a diverse class of lipophilic pigments that are produced by plants, fungi, algae and many groups of bacteria, including cyanobacteria, anoxygenic photosynthetic, and nonphotosynthetic bacteria. They are involved in light-harvesting in photosynthetic organisms; in nonphotosynthetic bacteria, the primary function of carotenoids is believed to be to protect against photooxidative damage (Armstrong 1997), (Pasamontes et al. 1997). Retinal biosynthetic genes are known to be widely distributed among diverse marine microbial groups (Sabeji et al. 2005), (McCarren and DeLong 2007), and the genes required for astaxanthin biosynthesis in the marine bacterium *Paracoccus haeuundaenis* have recently been characterized (Lee and Kim 2006).

To identify the carotenoids produced by these clones, pigments extracted from both carotenoid-expressing fosmid clones and two transposon insertion mutants of fosmid 30-B18 were analyzed by reverse phase HPLC (Fig. 6). Of the two transposon mutants of fosmid 30-B18, one appeared wild type in color when streaked out under high copy growth conditions, and the other had a distinct yellow color. Analysis of sequences off the ends of the <KAN-2> transposon insertion on the yellow mutant revealed that the gene interrupted had homology to β -carotene ketolase (*crtW*). Sequences from the wild type color transposon mutant indicated an insertion site within a gene with homology to phytoene dehydrogenase (*crtI*). Samples were run with purified standards of retinal, astaxanthin, lycopene, and β -carotene. The elution time and absorption spectrum for the sample extracted from the yellow colored transposon mutant matched that of the β -carotene standard. Both eluted approximately two minutes after the lycopene standard, with the 0.3 minute difference in the β -carotene standard peak at 57.2 minutes and the transposon insertion mutant peak at 56.9 minutes understood to be due to the HPLC machine having operated for several hours after the pigment standards before this sample was run. The spectra exhibit two broad overlapping peaks with maximum absorption at 454 nm and secondary peaks around 480 nm. This is consistent with sequence analysis suggesting that the transposon insertion interrupts the enzyme β -carotene ketolase and thus allows the accumulation of β -carotene in these cells. The sample extracted from the wild type colored transposon mutant appeared to be similar to the extracted Epi300 *E. coli* negative control. *CrtI*, the gene apparently knocked out in this mutant, is responsible for taking phytoene to lycopene in the biosynthetic pathway leading to β -carotene (Armstrong 1997), so it is likely that this mutant accumulates phytoene and therefore appears wild type in color.

The elution times and spectra for the samples corresponding to the original orange fosmid clones, 13-D19 and 30-B18, were essentially identical. These spectra appeared very similar to that of the astaxanthin standard, both exhibiting single broad peaks with maxima close to the astaxanthin maximum at 480 nm. However, the elution time for these samples was more than five minutes longer than that for the astaxanthin pigment

standard. This absorption spectrum and longer elution time is characteristic of the closely related carotenoid castaxanthin (Choi et al. 2007). It is not clear why both of these samples elute in two closely spaced peaks, the first smaller than the second. However, spectra of the two peaks are indistinguishable (data not shown). HPLC data indicating accumulation of β -carotene in the yellow transposon insertion mutant with a *crtW*-like gene interrupted also supports the conclusion that the carotenoid produced by these fosmid clones may be canthaxanthin. Both CrtO and CrtW type β -carotene ketolases catalyze the conversion of β -carotene to castaxanthin via the intermediate echinenone, and CrtW type ketolases can also act on the related carotenoid zeaxanthin to produce astaxanthin (Ye et al. 2006), (Choi et al. 2007), which was not detected in these samples.

To permit more extensive analysis of the carotenoid biosynthetic genes carried on these clones and to gain insight into the phylogenetic identity of the organisms from which they originate, additional sequencing was performed for both 13-D19 and 30-B18. The sequences obtained for these two Pacific photic zone fosmids are syntenic over a large overlapping region of approximately 31 kilobases. Based on the sequence of a putative transcription elongation factor encoded on both fosmids, these clones likely have a Betaproteobacterial origin. Other sequences encoded on these fosmids including a DNA repair photolyase, acyl-CoA-synthetase, and FAD/FMN-containing dehydrogenase support a Proteobacterial origin. Sequence analysis of the cloned carotenoid biosynthetic operon indicates that it consists of 6 genes, including a glycosyltransferase, β -carotene ketolase (*crtW*), lycopene cyclase (*crtY*), phytoene/squalene synthetase (*crtB*), phytoene dehydrogenase (*crtI*), and geranylgeranyl pyrophosphate synthase (*crtE*) (Fig. 7). This operon includes all the genes required for biosynthesis of β -carotene by *E. coli* (Martinez et al. 2007), as well as the ketolase required to produce castaxanthin from β -carotene.

Conclusions

In this work, a protocol was developed for the construction of randomly sheared, mid-size insert environmental genomic libraries in inducible expression plasmid vectors

from crude extracts of environmental microbial DNA. Libraries of this type offer an alternative platform to both large-insert, low-copy BAC and fosmid libraries and small-insert (1-3 kb) plasmid sequencing libraries for functional or phenotypic screening of environmental genomic libraries and, in addition, are appropriate for shotgun sequencing projects. Although a larger number of clones can be produced for shotgun sequencing by construction of small-insert libraries, libraries with larger inserts greatly increase the likelihood of capturing complete operons on individual clones, which is advantageous for functional screening. Placement of cloned environmental DNA under the control of an inducible promoter is intended to allow libraries to be produced and prepared for end-sequencing without expression of cloned DNA and induced only for functional or phenotypic screening. This feature should reduce bias associated with cloning into expression vectors and increase library stability. Individual clones can be easily characterized by both functional and sequence-based methods without the need for subcloning. The plasmid eventually selected for use in these libraries, pMCL200, carries a mid-low copy p15A origin of replication and the relatively strong IPTG inducible wild type P_{LAC} promoter. This choice was made to achieve high cloning efficiency of mid-size inserts and to balance the need for inducible expression at sufficiently high levels for successful screening with the reduced growth burden and additional stability that comes with low-copy cloning vectors and inhibition of gene expression during routine manipulations.

The final version of the protocol applied to marine microbial DNA samples produced a yield of approximately 100,000 clones per μg of crude DNA, of which 75% of clones examined contain insert, and mean insert lengths fell in the 5 to 6 kilobase range. Each time this protocol was successfully executed, library construction began with a large quantity, between 5 and 10 μg , of crude environmental DNA. However, no more than 2.1 μg worth of crude DNA went into the construction of any given library, and there is no reason a high-quality library could not be constructed from a starting sample of this size, provided the insert DNA was sufficiently concentrated going into the ligation step. The level of non-insert bearing background in the libraries was not ideal,

but it may be difficult to improve this without using a very large quantity of insert DNA, switching from blunt-end ligation to directional cloning with linkers, or incorporating some selection against empty-vector clones. Insert size was also smaller than the 8-10 kilobase length desired due to the use of insert DNA sheared to a mean insert size of approximately 8 kilobases. Based on observations during protocol development, it seems likely that the mean insert length could be increased from 5 or 6 kilobases to approximately 7 or 8 kilobases by shearing DNA into larger chunks of approximately 10 kilobases in length, as long as a favorable insert to vector molar ratio was maintained. However, few clones with inserts greater than 10 kilobases were observed during any plasmid library construction experiment, so it is not known whether high-quality libraries with larger inserts could be produced without protocol modifications.

Colony phenotype screening of surface water mid-size insert plasmid library HMS186_25m and comparison to fosmid library HF186_25m constructed from the same marine microbial DNA sample yielded some insights into the relative utility of these types of environmental genomic libraries for gene expression-based screens. Not surprisingly, expression of cloned genes appears to be extremely sensitive to growth conditions. With some exceptions, clones with distinct colony color phenotypes spotted onto array plates initially incubated at 30° C had more noticeable coloration than those placed at 37° C for the first sixteen hours of incubation, presumably because the slower growth rate leaves more cellular resources available for expression of non-growth related genes. Color phenotypes became apparent after periods of time ranging from 16 hours to more than one week after arraying, and at least one clone of interest initially had some visible coloration but colonies appeared wild type by the fifth day. These observations illustrate the importance of employing multiple growth conditions and evaluating results at various growth stages. In the case of the mid-size insert plasmid library, induction of gene expression with IPTG did not dramatically alter the result, although some colony color and fluorescence phenotypes were more pronounced in the presence of IPTG.

In comparing the two types of libraries, the fraction of clones that exhibit a non-wild type colony color phenotype in library arrays grown up with induction is roughly

proportional to insert size. The pCC1FOS fosmid vector is not designed as an expression vector, and expression of cloned genes under high copy growth conditions may be due to read-through from vector promoters into the multiple-cloning site or expression from promoters carried on environmental insert DNA. This has not been examined systematically, but both may occur. Expression of genes situated toward the middle of 40 kilobase fosmid inserts, such as those involved in carotenoid biosynthesis in fosmid clones 13-D19 and 30-B18, indicates that transcription from internal promoters may be important, while some fraction of fosmid clones investigated previously (Martinez, unpublished results) appear to be expressing genes downstream of vector promoters. Good gene expression from internal insert promoters is one explanation for the expression of genes responsible for distinct colony coloration in pMCL200 plasmid libraries arrayed without induction. Since promoter-like sequences are quite well conserved among diverse groups of microbes (Lonetto et al. 1992), (Huerta et al. 2006), this seems likely. Another factor, supported by the appearance of light blue colonies after multi-day incubations of pMCL200 libraries arrayed with X-gal but without IPTG, may be leaky expression from the P_{LAC} promoter or read through from other vector promoters. Examination of full insert sequences from a larger sample of phenotypically identified pMCL200 clones for gene orientation and internal promoter sequences would likely clarify this issue.

Growth of mid-size insert plasmid libraries in vector pMCL200 was not obviously affected by the addition of IPTG to induce gene expression, although subtly reduced growth was observed on some occasions and no experiments were conducted to explicitly measure growth rate. The addition of 0.02% arabinose to solid media to raise fosmid copy number, on the other hand, dramatically inhibited growth, and insert instability has also been observed under these conditions (Martinez, unpublished data). By avoiding the physiological complications associated with induction of fosmid libraries to high copy number for screening, mid-size insert relatively low-copy plasmid libraries may offer some advantages, particularly in screens that depend on the detection of subtle growth advantages or change in respiration rates (Henne et al. 1999). However, phenotypes that

require the cloning of full biosynthetic operons in excess of 5 kilobases or so in length will rarely be observed in libraries of this type, and even functional screens that require the expression of only a single gene may yield more hits by screening fosmid libraries simply due to the greater quantity of environmental DNA contained in the larger fosmid clones. One drawback associated with large insert libraries is the potential difficulty of identifying the gene or genes responsible for a phenotype of interest. Transposon mutagenesis followed by partial sequence analysis of mutants lacking the original phenotype is often effective but can give ambiguous results due to downstream effects or novel gene sequences. Mid-size insert clones, which contain fewer open reading frames and can be fully sequenced more conveniently, may simplify this process, but at the same time provide less genomic context and fewer phylogenetic clues.

Since all culture-independent methods of characterizing natural microbial communities have biases that at this time are not well understood, employing multiple approaches may be the best strategy to determine the phylogenetic composition and genetic potential of these populations. Preliminary characterization of library quality and results of colony phenotype arrays indicate that libraries of this type may be useful in functional screening of environmental genomic libraries for activities of interest. Although arrays of a fosmid library derived from the same surface water sample as plasmid library HMS186_25m exhibited a higher percentage of phenotypically distinct colonies, analysis of phenotypically identified clones in library HMS186_25m revealed sequences of likely ecological relevance in marine environments including genes potentially involved in iron transport and vitamin_{B12} biosynthesis. This result suggests that some fraction of genes carried on marine microbial DNA cloned into plasmid pMCL200 are expressed at sufficiently high levels to allow gene expression-based screening and that libraries of this type offer an alternative platform to large insert BAC or fosmid clones for such studies.

Acknowledgements

I thank Julia Maresca for assistance with the HPLC analysis, Jay McCarren for providing fosmid library HF186_25m, Gene Tyson and Kostas Konstantinidis for help with gene annotations, members of the DeLong lab, particularly Chon Martinez, for helpful consultation, and William Braff, Mak Saito, and Ed DeLong for comments on the manuscript. This work was supported by a grant to EFD from the Gordon and Betty Moore Foundation.

Bibliography

- Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. *J Mol Biol* 215(3): 403-410.
- Armstrong E, Granger J, Mann E, Price N (2004) Outer-membrane siderophore receptors of heterotrophic oceanic bacteria. *Limnol Oceanogr* 49(2): 579-587.
- Armstrong G (1997) Genetics of eubacterial carotenoid biosynthesis: a colorful tale. *Annu Rev Microbiol* 51: 629-659.
- Ausubel F, Brent R, Kingston R, Moore D, Seidman J et al. (1992) *Current Protocols in Molecular Biology*: Green Publishing Associates and Wiley Interscience.
- Beja O, Aravind L, Koonin E, Suzuki M, Hadd A et al. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* 289: 1902-1906.
- Bernard P, Gabant P, Bahassi E, Couturier M (1994) Positive-selection vectors using the F plasmid *ccdB* killer gene. *Gene* 148: 71-74.
- Boyd P, Watson A, Law C, Abraham E, Trull T et al. (2000) A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature* 407: 695-702.
- Choi S, Harada H, Matsuda S, Misawa N (2007) Characterization of two B-carotene ketolases, *CrtO* and *CrtW*, by complementation analysis in *Escherichia coli*. *Appl Microbiol Biotechnol* 75: 1335-1341.
- Courtois S, Cappellano C, Ball M, Francou F, Normand P et al. (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Applied and Environmental Microbiology* 69(1): 49-55.
- DeLong E, Preston C, Mincer T, Rich V, Hallam S et al. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311(5760): 496-503.
- Drewlo S, Bramer C, Madkour M, Mayer F, Steinbuchel A (2001) Cloning and expression of a *Ralstonia eutropha* HF39 gene mediating indigo formation in *Escherichia coli*. *Applied and Environmental Microbiology* 67(4).
- Eaton R, Chapman P (1995) Formation of indigo and related compounds from indolecarboxylic acids by aromatic acid-degrading bacteria: chromogenic

- reactions for cloning genes encoding dioxygenases that act on aromatic acids. *Journal of Bacteriology* 177(23): 6983-6988.
- Ensley B, Ratzkin B, Osslund T, Simon M, Wackett L et al. (1983) Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. *Science* 222(4620): 167-169.
- Frigaard N, Martinez A, Mincer T, Delong E (2006) Proteorhodopsin lateral gene transfer between marine planktonic bacteria and archaea. *Nature* 439(7078): 847-850.
- Giovannoni S, Britschgi T, Moyer C, Field K (1990) Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345: 60-63.
- Grant S, Jessee J, Bloom F, Hanahan D (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl Acad Sci USA* 87: 4645-4649.
- Guerinot M (1994) Microbial iron transport. *Annu Rev Microbiol* 48: 743-772.
- Guzman L, Belin D, Carson M, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J Bacteriol* 177(14): 4121-4130.
- Hallam S, Mincer T, Schleper C, Preston C, Roberts K et al. (2006a) Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biology* 4(4).
- Hallam S, Konstantinidis K, Putnam N, Schleper C, Watanabe Y et al. (2006b) Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc Natl Acad Sci USA* 103(48): 18296-18301.
- Henne A, Daniel R, Schmitz R, Gottschalk G (1999) Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Applied and Environmental Microbiology* 65(9): 3901-3907.
- Henne A, Schmitz R, Bomeke M, Gottschalk G, Daniel R (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Applied and Environmental Microbiology* 66(7): 3113-3116.
- Huerta A, Francino M, Morett E, Collado-Vides J (2006) Selection for unequal densities of sigma-70 promoter-like signals in different regions of large bacterial genomes. *PLOS Genetics* 2(11): 1740-1750.
- Karl D, Lukas R (1996) The Hawaii Ocean Time-series (HOT) program: Background, rationale, and field implementation. *Deep-Sea Research II* 43: 129-156.
- Karl D, Bridigare R, Letelier R (2001) Long-term changes in plankton community structure and productivity in the North Pacific Subtropical Gyre: The domain shift hypothesis. *Deep-Sea Research II* 48: 1449-1470.
- Kim U, Shizuya H, Jong P, Birren B, Simon M (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Research* 20(5): 1083-1085.
- Lane D, Pace B, Olsen G, Stahl D, Sogin M et al. (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci USA* 82: 6955-6959.

- Lee J, Kim Y (2006) Cloning and characterization of the astaxanthin biosynthesis gene cluster from the marine bacterium *Paracoccus haeundaensis*. *Gene* 370: 86-95.
- Li Y, Wexler M, Johnson A (2005) Screening a wide host-range, wast-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned trp genes. *Environ Microbiol*.
- Lonetto M, Gribscov M, Gross C (1992) The sigma-70 family: sequence conservation and evolutionay relationships. *Journal of Bacteriology* 174(12): 3843-3849.
- Luke R, Gibson F (1971) Location of three genes concerned with the conversion of 2,3-dihydroxybenzoate into enterochelin in *Escherichia coli* K-12. *Journal of Bacteriology* 107(2): 557-562.
- Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research* 25(6): 1203-1210.
- MacNeil I, Tiong C, Minor C, August P, Grossman T et al. (2001) Expression and Isolation of Antimicrobial Small Molecules from Soil DNA Libraries. *J Mol Microbiol Biotechnol* 3(2): 301-308.
- Majernik A, Gottschalk G, Daniel R (2001) Screening of environmental DNA libraries for the presence of genes conferring Na⁺(Li⁺)/H⁺ antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. *Journal of Bacteriology* 183(22): 6645-6653.
- Maresca J, Graham J, Wu M, Eisen J, Bryant D (2007) Identification of a fourth family of lycopene cyclases in photosynthetic bacteria. *PNAS* 104(28): 11784-11789.
- Martinez A, Bradley A, Waldbauer J, Summons R, Delong E (2007) Proteorhodopsin photosystem gene expression enables photophosphorylation in heterologous host. *Proc Natl Acad Sci U S A*.
- Martinez A, Kolvek S, Yip C, Hopke J, Brown K et al. (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. *Applied and Environmental Microbiology*: 2452-2463.
- Martinez J, Zhang G, Holt P, Jung H, Carrano C, Haygood M, Butler A (2000) Self-assembling amphiphilic siderophores from marine bacteria. *Science* 287: 1245-1247.
- McCarren J, DeLong E (2007) Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environ Microbiol* 9(4): 846-858.
- Mourino S, Osorio C, Lemos M (2004) Characterization of heme uptake cluster genes in the fish pathogen *Vibrio anguillarum*. *Journal of Bacteriology* 186(18): 6159-6167.
- Nakagawa S, Takai K, Inagaki F, Hirayama H, Nunoura T et al. (2005) Distribution, phylogenetic diversity and physiological characteristics of epsion-Proteobacteria in a deep-sea hydrothermal field. *Environmental Microbiology* 7(10): 1619-1632.

- Nakano Y, Yoshida Y, Yamashita Y, Koga T (1995) Construction of a series of pACYC-derived plasmid vectors. *Gene* 162: 157-158.
- Olsen G, Lane D, Giovannoni S, Pace N, Stahl D (1986) Microbial ecology and evolution: a ribosomal RNA approach. *Annu Rev Microbiol* 40: 337-365.
- Pace N, Stahl D, Lane D, Olsen G (1985) Analyzing natural microbial populations by rRNA sequences. *ASM American Society for Microbiology News* 51(1): 4-12.
- Pasamontes L, Hug D, Tessier M, Hohmann H, Schierle J et al. (1997) Isolation and characterization of the carotenoid biosynthesis genes of *Flavobacterium* sp. strain R1534. *Gene* 185: 33-41.
- Pedros-Alio C (2006) Marine microbial diversity: can it be determined? *Trends in Microbiology* 14: 6.
- Pheiffer B, Zimmerman S (1983) Polymer-stimulated ligation: enhanced blunt- or cohesive-end ligation of DNA or deoxyribonucleotides by T4 DNA ligase in polymer solutions. *Nucleic Acids Research* 11(22).
- Rusch D, Halpern A, Sutton G, Heidelberg K, Williamson S et al. (2007) The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLOS Biology* 5(3): 398-432.
- Sabehi G, Loy A, Jung K, Partha R, Spudich J et al. (2005) New Insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biology* 3(8).
- Sambrook J, Russell D (2001) *Molecular Cloning*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Sattler I, Roessner C, Stolowich N, Hardin S, Harris-Haller L et al. (1995) Cloning, sequencing, and expression of the uroporphyrinogen III methyltransferase *cobA* gene of *Propionibacterium freudenreichii* (shermanii). *Journal of Bacteriology* 177(6): 1564-1569.
- Schmidt T, DeLong E, Pace N (1991) Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of Bacteriology* 173(14): 4371-4378.
- Schwyn B, Neilands J (1987) Universal chemical assay for the detection and determination of siderophores. *Anal Biochem* 160(1): 47-56.
- Shizuya H, Birren B, Kim U, Mancino V, Slepak T et al. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci USA* 89: 8794-8797.
- Sogin M, Morrison H, Huber J, Welch D, Huse S et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci U S A*.
- Stahl D, Lane D, Olsen G, Pace N (1985) Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. *Applied and Environmental Microbiology* 49(6): 1379-1384.
- Staley J, Konopka A (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* 39: 321-346.
- Stein J, Marsh T, Wu K, Shizuya H, DeLong E (1996) Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology* 178(3): 591-599.

- Street J, Paytan A (2005) Iron, phytoplankton growth, and the carbon cycle. *Met Ions Biol Syst* 43: 153-193.
- Suzuki M, Preston C, Beja O, de la Torre J, Steward G et al. (2004) Phylogenetic screening of ribosomal RNA gene-containing clones in Bacterial Artificial Chromosome (BAC) libraries from different depths in Monterey Bay. *Microb Ecol* 48: 473-488.
- Tringe S, Mering C, Kovayashi A, Salamov A, Chen K et al. (2005) Comparative metagenomics of microbial communities. *Science* 308: 554-557.
- Tyson G, Chapman J, Hugenholtz P, Allen E, Ram R et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428: 25-26.
- Uchiyama T, Abe T, Ikemura T, Watanabe K (2004) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature Biotechnology*.
- Venter J, Remington K, Heidelberg J, Halpern A, Rusch D et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304(5667): 66-74.
- Ward D, Weller R, Bateson M (1990) 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345: 63-65.
- Wexler M, Bond P, Richardson D, Johnson A (2005) A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environ Microbiol*.
- Wild J, Hradecna Z, Szybalski W (2002) Conditionally amplifiable BACs: switching from single-copy to high-cop vectors and genomic clones. *Genome Research*.
- Wildt S, Deuschle U (1999) *cobA*, a red fluorescent transcriptional reporter for *Escherichia coli*, yeast, and mammalian cells. *Nature Biotechnology* 17: 1175-1178.
- Williamson L, Borlee B, Schloss P, Guan C, Allen H et al. (2005) Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Applied and Environmental Microbiology* 71(10): 6335-6344.
- Woodcock D, Crowther P, Doherty J, Jefferson S, DeCruz E et al. (1989) Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Research* 17(9): 3469-3478.
- Ye R, Stead K, Yao H, He H (2006) Mutational and functional analysis of the B-carotene ketolase involved in the production of canthaxanthin and astaxanthin. *Applied and Environmental Microbiology* 72(9): 5829-5837.

| Library | Vector | Insert | Fragment Size (kb) | µg crude DNA per ligation | insert:vector molar ratio in ligation | cfu/µg crude DNA | fraction clones digested containing insert | Mean Insert Size (kb) |
|-----------------------------|----------------|-----------------------------------|--------------------|---------------------------|---------------------------------------|------------------|--|-----------------------|
| lambda test library | pBAD33 | lambda DNA | 8-12 | 0.5 | 3.3:1 | 4.00E+04 | 7/8 | 6.4 |
| BATS libraries | | | | | | | | |
| BATS- large fraction | pBAD33 | BATS 10/06 mixed depth DNA | 8-10 | 1.4 | 4.0:1 | 3.02E+03 | 0/6 | n/a |
| BATS- small fraction | pBAD33 | BATS 10/06 mixed depth DNA | 4-8 | 1.4 | 5.0:1 | 4.76E+03 | 0/6 | n/a |
| lambda control | pBAD33 | lambda DNA | 4-8 | n/a | 8.0:1 | n/a | 2/6 | 2.7 |
| BATS- large fraction | pMCL200 | BATS 10/06 mixed depth DNA | 8-10 | 1.4 | 2.8:1 | 5.20E+04 | 4/6 | 6.0 |
| BATS- small fraction | pMCL200 | BATS 10/06 mixed depth DNA | 4-8 | 1.4 | 3.8:1 | 1.60E+05 | 5/6 | 2.4 |
| lambda control | pMCL200 | lambda DNA | 4-8 | n/a | 6.3:1 | n/a | 6/6 | 3.9 |
| HOT libraries | | | | | | | | |
| HMS186_25m | pMCL200 | HOT 186 photic zone DNA | 7-10 | 2.1 | 2.0:1 | 1.98E+05 | 11/14 | 5.2 |
| HOT- small fraction | pMCL200 | HOT 186 photic zone DNA | 4-7 | 2.1 | 2.6:1 | 1.98E+05 | 6/6 | 4.6 |

Table 1. Summary of results for selected libraries.

| Plate | Well Number | Color | Fluorescence | Phenotype +IPTG |
|--|-------------|------------------|---------------|---------------------|
| 1 | M15 | wild type | bright | more fluorescent |
| 10 | M20 | faint pink | no | no change |
| 11 | D16 | gray-brown | yes | no change |
| 11 | H17 | purple | yes | darker color |
| 12 | D7 | wild type | yes | no change |
| 16 | I1 | wild type | bright | more fluorescent |
| 28 | P8 | blue | bright | no change |
| 29 | L8 | faint pink-brown | no | no change |
| 46 | E6 | wild type | yes | more fluorescent |
| 43 | I6 | wild type | yes | no change |
| arrayed with <i>all-trans</i> retinal only | | | | |
| 50 | D15 | brown | not scanned | no change |
| 54 | B7 | faint orange | not scanned | no change |
| 54 | F8 | blue-purple | not scanned | no change |

Table 2. Phenotypically identified clones in library HMS186_25m.

| Plate | Well Number | Color |
|-----------|-------------|----------------------|
| 1 | N22 | yellow-green tinge |
| 1 | N24 | yellow-green tinge |
| 2 | A11 | yellow-green tinge |
| 2 | A16 | faint purple |
| 2 | A9 | yellow-green tinge |
| 3 | K13 | yellow-green tinge |
| 5 | P14 | faint brown |
| 8 | P2 | yellow |
| 10 | L20 | yellow-green tinge |
| 10 | P18 | yellow-green tinge |
| 13 | D19 | bright orange |
| 16 | M13 | yellow-green tinge |
| 18 | E24 | yellow-green tinge |
| 19 | E19 | yellow-green tinge |
| 21 | C11 | yellow-green tinge |
| 21 | G3 | yellow-green tinge |
| 21 | P1 | yellow-green tinge |
| 22 | F8 | blue |
| 24 | N8 | yellow-green tinge |
| 29 | B14 | yellow-green tinge |
| 29 | G22 | yellow-green tinge |
| 30 | B18 | bright orange |
| 31 | B3 | yellow-green tinge |
| 32 | B17 | yellow-green tinge |
| 39 | A19 | yellow-green tinge |
| 39 | D7 | yellow-green tinge |
| 39 | E24 | faint brown-orange |
| 39 | M16 | faint yellow-orange |
| 43 | D5 | blue-green |
| 43 | I20 | blue-green |
| 47 | L7 | faint brown |

Table 3. Phenotypically identified clones in library HF186_25m.

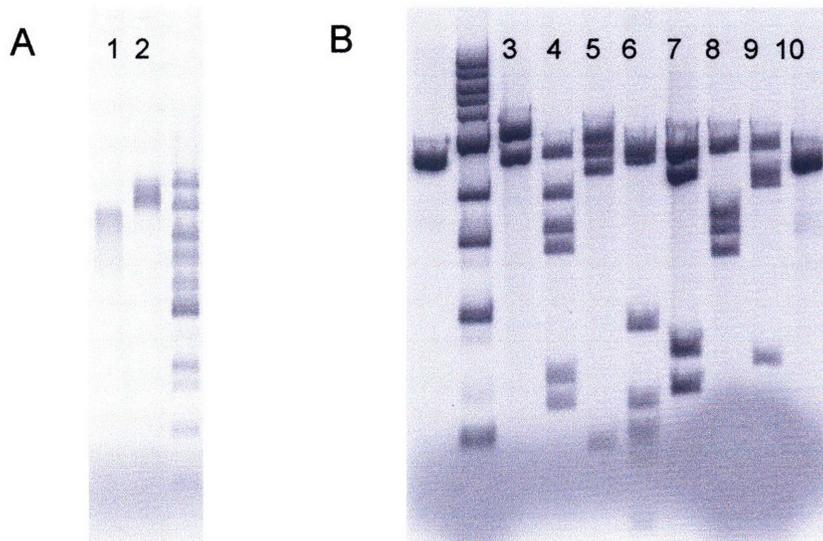


Figure 1. (A) Both size fractions of prepared insert DNA (2 μ L/lane, lanes 1 and 2) for BATS test libraries run alongside 250 ng of 1 kb ladder (NEB). (B) BamHI/HindIII digests to spot check library HMS186_25m in Top10-F' cells (lanes 3-10). Left-hand lane is digested pMCL200 plasmid DNA; marker is 1 kb ladder.

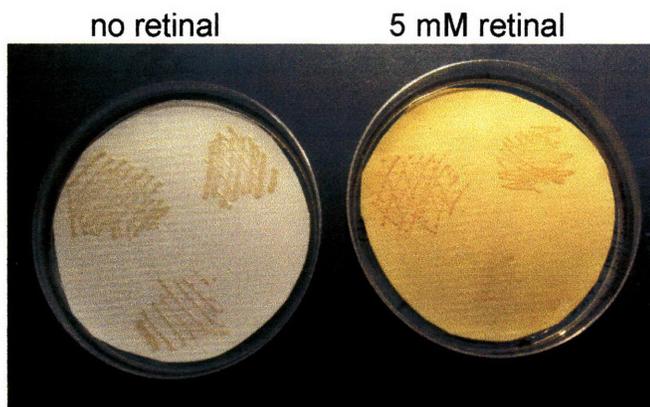


Figure 2. Proteorhodopsin (PR) expressing clones streaked out on agar plates overlaid with opaque membranes, with and without 5 mM all-*trans* retinal. Clockwise from bottom: puc19 (no PR), pBAD-31A8, and pBAD-S14PR.

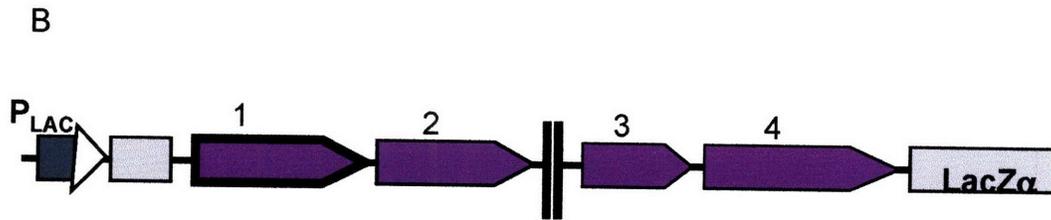
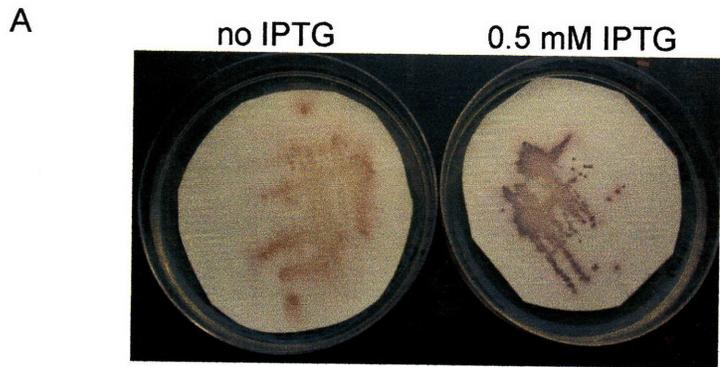


Figure 3. Clone 11-H17 in library HMS186_25m
 (A) Streaked out with and without 0.5 mM IPTG
 (B) Schematic of sequence annotation: 1. uroporphyrin-III C-methyltransferase
 2. Glycosyltransferase 3. nucleotidase-like 4. diguanylate cyclase

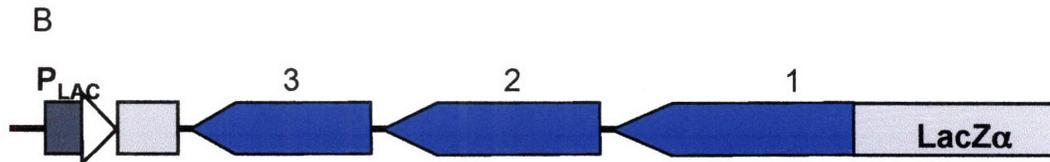
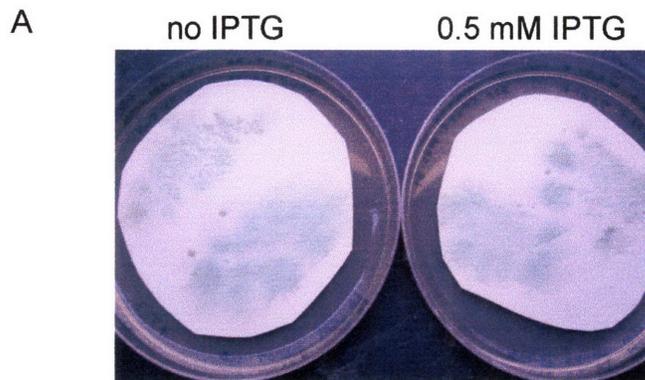


Figure 4. Clone 28-P8 in library HMS186_25m
 (A) Streaked out with and without 0.5 mM IPTG
 (B) Schematic of sequence annotation:
 1. ABC-type hemin transport system (permease component)
 2. ABC-type hemin transport system (ATPase component)
 3. uncharacterized metal-binding protein



Figure 5. Clones from library HF186_25m streaked out on membranes overlaying agar plates supplemented with 0.02% L-arabinose.
 Left to right: 22-F8, 13-D19, 30-B18.

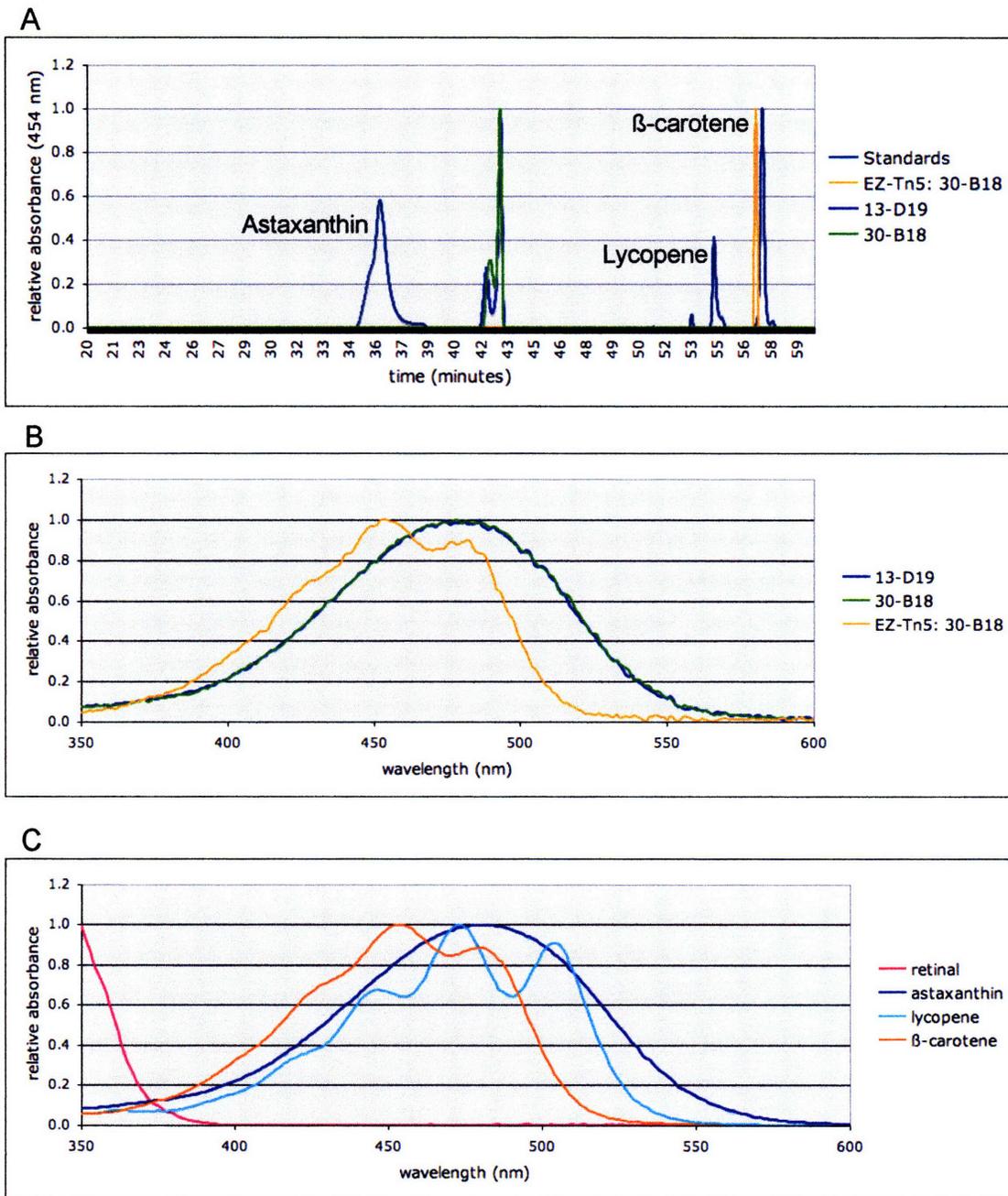


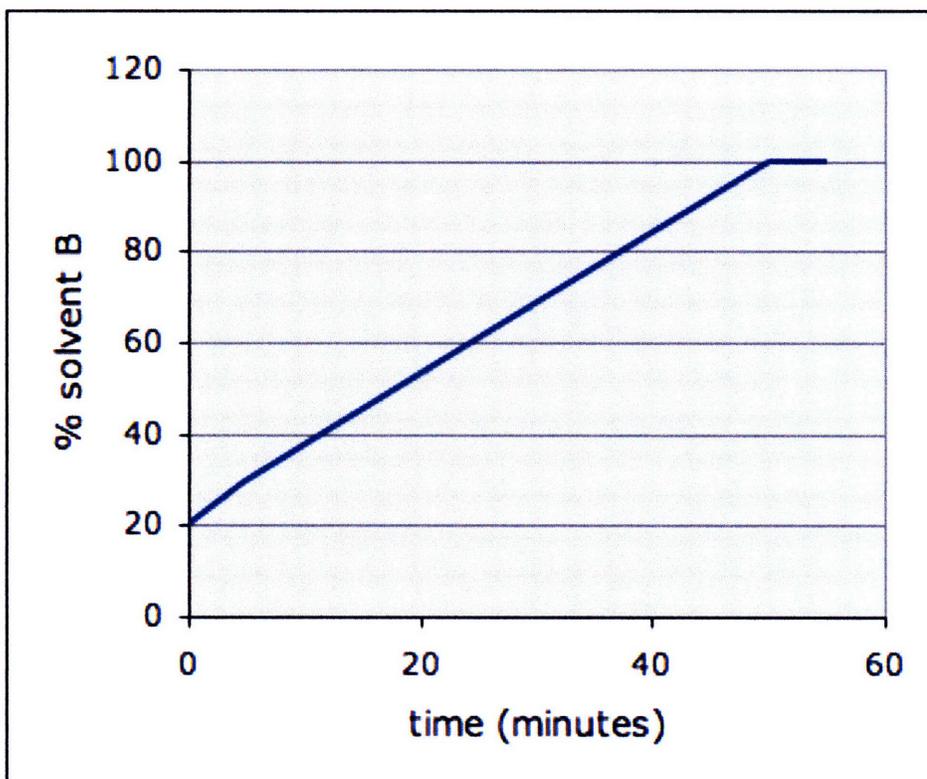
Figure 6. HPLC analysis of pigments produced by HF186_25m fosmid clones 13-D19 and 30-B18 and transposon insertion mutant of 30-B18.
 (A) HPLC profiles of pigment standards and experimental samples.
 (B) Absorption spectra of sample peaks.
 (C) Absorption spectra of pigment standard peaks.



Figure 7. Schematic of the carotenoid biosynthetic pathway carried on fosmids 13-D19 and 30-B18 in library HF186_25m. The last gene in the operon encodes a putative glycosyltransferase.

| Vector | Primer Name | Description | Sequence | Source |
|----------------|-----------------|------------------------|--------------------------------|---|
| pCC1FOS | T7 | forward end-sequence | 5' TAATACGACTCACTATAGGG 3' | Epicentre, Madison, WI |
| pCC1FOS | pEpiFOS RP-2 | reverse end-sequence | 5' TACGCCAAGCTATTTAGGTGAGA 3' | Epicentre, Madison, WI |
| pMCL200 | pMCL200-seq-FW | forward end-sequence | 5' GGTCGACGGTATCGATAAGC 3' | Joint Genome Institute |
| pMCL200 | pMCL200-seq-REV | reverse end-sequence | 5' CGCTCTAGAAGTGTGGATC 3' | Joint Genome Institute |
| pMCL200 | pUC-pcr-F | PCR forward | 5' CTTTACACTTTATGCTTCCGG 3' | Joint Genome Institute |
| pMCL200 | pUC-pcr_R | PCR reverse | 5' GCAAGGCGATTAAGTTGG 3' | Joint Genome Institute |
| EZ-Tn5 <KAN-2> | KAN-2 FP-1 | forward off transposon | 5' ACCTACAACAAGCTCTCATCAACC 3' | Epicentre, Madison, WI |
| EZ-Tn5 <KAN-2> | KAN-2 RP-1 | reverse off transposon | 5' GCAATGTAACATCAGAGATTTGAG 3' | Epicentre, Madison, WI |
| pBAD33 | pBAD33 F | forward | 5'-CTGTTTCTCCATACCCGTT-3' | American Type Culture Collection (ATCC) |
| pBAD33 | pBAD33_R | reverse | 5'-CTCATCCGCCAAAACAG-3' | American Type Culture Collection (ATCC) |

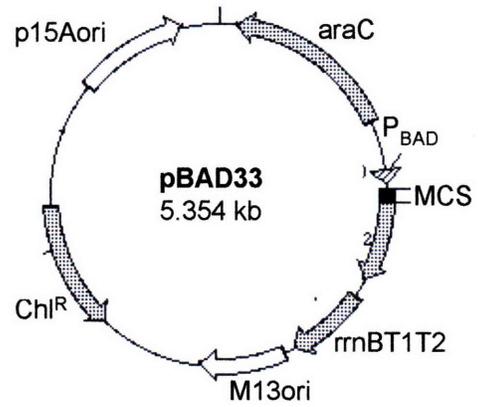
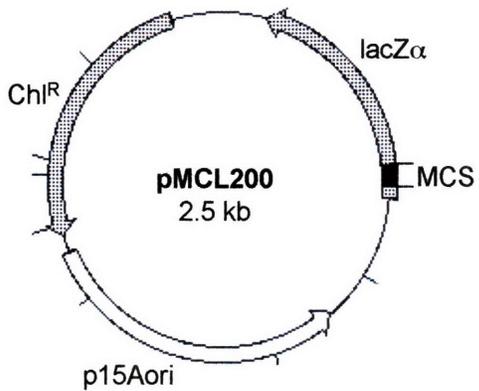
Supplementary Table 1. Primer sequences.



Supplementary Figure 1. Solvent gradient for HPLC analysis.

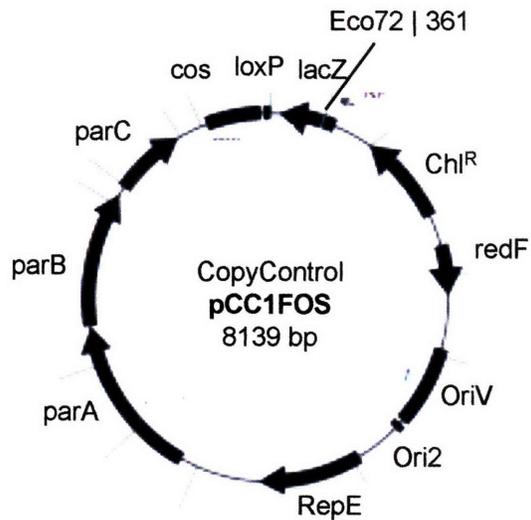
Plasmid Vectors

(<http://gillnet.lab.nig.ac.jp/~cvector/map/>)



Fosmid Vector

(Epicentre CopyControl Fosmid Kit)



Supplementary Figure 2. Vectors used for library construction.

Appendix A

Protocol for Construction of Mid-Size Insert Environmental Genomic Expression Libraries in Plasmid pMCL200

Adapted from Joint Genome Institute protocols for library construction (see <http://www.jgi.doe.gov/sequencing/protocols/index.html>) and Epicentre protocols for CopyControl Fosmid library construction (Epicentre, Madison, WI).

[1] Shearing of Environmental DNA

If using crude extracts of environmental DNA, filter DNA through a 0.2 μm filter before shearing. Always wash Hydroshear as directed before and after use. For precise fragment size, each shearing orifice should be calibrated before use.

- 1.1 Dilute DNA to ~ 100 ng/ μL in filter sterilized TE. Run 5-10 μg DNA (if available) in a volume of 50–150 μL .
- 1.2 Incubate sample 30 minutes at 37° C; vortex every 10 minutes.
- 1.3 Spin 20 minutes at high speed on a table-top microcentrifuge. Any pellet indicates incomplete resuspension.
- 1.4 Transfer sample to a clean tube.
- 1.5 Shear DNA on a Gene Machines Hydroshear (Genomic Solutions) at speed code 15 (gives approximately 7-10 kilobase fragments) or speed code 16 (gives approximately 8-12 kilobase fragments) for 25 cycles.
- 1.6 Chill sample on ice immediately after shearing.

[2] End-repair of Insert DNA

- 2.1 Combine the following to give a total volume of 80 μL . Scale up as needed.

- 8 μL 10X End-Repair Buffer
- 8 μL 2.5 mM each dNTP mix
- 8 μL 10 mM ATP
- 4 μL End-Repair Enzyme Mix (Epicentre, End-It DNA End-Repair kit)
- 52 μL sheared DNA

- 2.2 Incubate 1 hour at room temperature.

- 2.3 Add 7 μ L 125 mM EDTA (to 10 mM), and incubate at 70° C for 10 minutes to heat inactivate,
- 2.4 Chill sample on ice.

T4 polynucleotide kinase contained in the End-Repair Enzyme Mix may not be completely heat inactivated, although the enzyme should be removed by gel purification. If kinase carry-over and phosphorylation of vector DNA during ligation is a concern, the reaction can be phenol-chloroform extracted prior to concentration.

[3] EtOH Precipitation of DNA to Concentrate

- 3.1 Add 2 μ L Pellet Paint (Novagen). Be sure to fully resuspend Pellet Paint before use.
- 3.2 Add 1/10 volume 3 M NaAcetate (pH 5.2) to sample and mix.
- 3.3 Add 2 volumes of 95% EtOH. Invert to mix.
- 3.4 Incubate 2 minutes at room temperature.
- 3.5 Spin 14K XG for 5 minutes. Remove supernatant with a pipette.
- 3.6 Wash pellet with 70% ethanol. Spin and remove supernatant as above.
- 3.7 Wash with 95% ethanol. Spin and remove supernatant as above.
- 3.8 Dry pellet 10 minutes, and resuspend sample in 30 μ L TE.

[4] Gel Purification of Insert DNA

- 4.1 Set a 1% LMP SeaPlaque (FMC BioProducts) agarose-TAE gel. Chill gel at 4° C before use.
- 4.2 Load the entire sample in one lane, and load 500 ng 1 kb DNA ladder (New England Biolabs) in a non-adjacent lane.
- 4.3 Run gel at ~100 V (for large gel) for 3-4 hours. Change running buffer after 2 hours.
- 4.4 Stain gel with SYBRGold dye (Molecular Probes) for 15 minutes with gentle shaking.
- 4.5 View gel on a non-UV light table, and cut smaller (~6-8 kb) and larger (~8-10 kb) size fractions of sheared DNA.
- 4.6 Cut away any excess agarose.
- 4.7 For fragments 10 kilobases in length or larger, gel purify insert DNA using the QIAEX II kit (Qiagen) according to manufacturer's directions. For best results, do not process more than 150 mg of agarose per tube.

or

If fragments are shorter than 10 kilobases, insert can be purified using the QIAQuick kit (Qiagen). Include the optional isopropanol and buffer QG washes. Let buffer PE sit 5 minutes before spinning, and repeat the buffer PE wash step. To elute DNA, add 50 μ L warm (50° C) buffer EB to the column, and let column sit for 10 minutes before elution.

4.8 Run a 1% agarose gel with 1-2 μL each of prepared insert and vector DNA and stain to check vector and insert concentrations.

[5] Blunt-end Ligation

Linear, blunt-end pMCL200 vector DNA should be prepared by digesting with EcoRV, followed by dephosphorylation and agarose gel purification. An insert:vector molar ratio between 2:1 and 3:1, as estimated by quantitation of DNA in step 4.8, works well.

5.1 Combine the following and mix gently:

1 μL 10X T4 DNA ligase buffer
200 units T4 DNA ligase (New England Biolabs)
15 ng prepared pMCL200 DNA
insert DNA (as prepared in steps 1-4)
to 10 μL with sterile H_2O

5.2 Incubate at 16° C overnight.

5.3 Heat inactivate at 65° C for 10 minutes.

5.4 Dialyze against water for 30 minutes on a 0.25 μm filter (Millipore).

[6] Transformation and Test Plating

The *E. coli* host strain Top10-F' (Invitrogen) used below carries a lacI^q allele and Tet^R marker on an F' plasmid to permit IPTG inducible gene expression from the lac promoter on plasmid pMCL200.

6.1 Mix 1 μL ligation product with 50 μL electrocompetent Top10-F' cells on ice and transfer to a chilled 1 mm cuvette.

6.2 Electroporate samples using the following settings: 25 μF , 2.0 kV, 200 Ohms.

6.3 Mix samples with 1 mL room temperature SOC media and incubate at 37° C with shaking for 1 hour.

6.4 To check library, plate 10 μL of transformation product diluted 1:20 into SOC (200 μL total volume) onto a small LB plate supplemented with 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol, 10 $\mu\text{g}/\text{mL}$ tetracycline, and 40 $\mu\text{g}/\text{mL}$ X-gal and the same onto similarly supplemented LB plate with the addition of 1 mM IPTG.

6.5 Store the remaining transformation product as a 25% glycerol stock at -80° C to plate once library yield and quality are confirmed.

6.6 Incubate test plates overnight at 37° C.

6.7 Count colonies and score fraction of blue (likely non-insert bearing) colonies on the test plate with IPTG.

[7] Library Quality Check

- 7.1 Pick 6-10 white colonies from the test plate with IPTG added and grow up 5 mL liquid cultures overnight with appropriate selection.
- 7.2 Prepare plasmid DNA from 4 mL of liquid culture.
- 7.3 Digest plasmid DNA from each clone by combining the following and mixing gently:

- 2 μ L NEB buffer 2
- 5 units BamHI (New England Biolabs)
- 5 units HindIII (New England Biolabs)
- 200 ng plasmid DNA
- sterile H₂O to 20 μ L

- 7.3 Incubate digests at 37° C for 4 hours.
- 7.4 Heat inactivate reactions at 65° C for 20 minutes.
- 7.5 Run out digests on a 1% agarose gel and stain to visualize bands and check insert presence/length.

[8] Library Plating for Automated Picking

- 8.1 Pour large rectangular LB plates supplemented with 12.5 μ g/mL chloramphenicol, 10 μ g/mL tetracycline, 40 μ g/mL X-gal, and 1 mM IPTG. Avoid bubbles and take care that plates are level and plate volume is appropriate for machine settings.
- 8.2 Plate glycerol stock of library transformation product at a density of <3000 colonies/plate (bring volume of glycerol stock to 600 μ L with SOC media for plating).
- 8.3 Incubate plates at 37° C overnight, then store plates for 2 days at 4° C to let blue color develop on *lacZ* expressing clones.
- 8.4 Pick clones as for fosmid libraries with a QPixII robot (Genetix), except set QPixII to pick only white colonies.

Appendix B

DNA Sequence for Selected Clones in Libraries HMS186_25m and HF186_25m

HMS186_25m 11-H17

>forward contig_11_H17

AGCAGCGCCACATTCANNAAGCACATGGTCAGTACCTAACGAATGGCTTG
CCAGTCTTCTCAACACATCTATTTCTACAGACATACAGCGCAATTTAATA
CGTGGCCAAGTAGACGACGCATTTTTAAATGGTGATGTAGTTTGTAGCTG
CTTTGAAGTGCGCGAAAAAACCATTACNCATGCCATCGAAGAAGGCTGTA
GTTTCGGTGGCAGAGCTTGGTGAAGCATTAAAGTGTGGTACGAATTGTGGC
TCGTGCAAGCCCGCCTTGTCTCAGCTGATANACAAGCATTAGTCATTGA
AATTCAATCTGTTTAGGAGCGTTAAATCATGTCAATTGCATCTAGTTGGT
CGTTACCGTTTTTTCGCGACCTTATTTAAACCGCAGAAACCTTTGTAAAC
CTCTTTTATCGCGATGCTATTTCTACCTCGGTCACGACCGTGACGACGCG
CCAGCGCTCCGATAATCCAAAAGCTCCAGCTAAGCATTTCATACTTTGGCA
AGTGGTTGAAGTTAGCAAACAGGCAAACCTGAAGGTCATAAACAAGGGCAG
GTGTTTCATCGTCGGTGCTGGCCCTGGTGATGCCGAGTTACTTACGTAA
AGCGTTTCGTCTTCTGCAACAAGCCGATGTCGTGCTGTTAGATGCTTTAG
TCAGCGAAGATATTCTCGCGCTAATCCCTCTAAAGTGGTGAAAGAATAT
GTAGGGAAACGCTGCAAAAAGCATAGCTTTACCCAAGACGCTATATGCAA
GCGAGTGGTTGAATTGGCAAGCAACGGTCACACTGTCGTACGGTTAAAAG
GCGGTGACCCAGCCCTATTTGCAAGAACCTGCGAAGAAACCGATGCACTC
ACAAAAGCCAACATTCCTTTTGCTATTGTTCCCTGGTATTACGGCAGCATC
GGGTGTGTCAGCCTATACGGGGATCCCTTTAACCGACAGGCGTTGTGCTC
AGTCGGTGAGTTTTATGACCGCACATTTTAAGGACGTGGATCAATGGCCA
GAGATGGCACCCATGGCGCAAAATGTATTGAAGCAAACCATGGTGGTTTA
TATGGGGCTAAGCCGACTTGAAGGGCTATACAAAGGCTTGATTCATCACA
ACGTGCCAACACATGGCCAGTAGCGGCAATTGAAAATGCCACTTCGCCT

GAACAGCGCGTGATTACTGGCGTGCTTGCCGATATTCATAGCAAGGTAGA
AGCAGCAAATTTAACGGGGCCTACGTTACTTATATTTGGAAAGGTGGTAG
AGTCTCGTCAACAGGTAAATACTCTTCTATTACATTCATCCAAGCATGCA
GCAACCATTTAGTGAATACATAAAAATTATTGGAAAAGGTCAAAAAGGGG
GGCGTAGCCTTACCCGTGATGAGGCTTATCATGCCATGAAAATGGTATTG
GCCAATGAAGTGACAGGCGATCAACGCGGCGCATTCTTATGTTGCTGCG
TACCCGTGAAGAAACACCAGAAGAAGTTATTGGCTTTGTCGATGCATGCA
GAGAGCTGATAGAGCCAGAAGTTAAAAACCTCACCCCCACGTTAGATATT
GGCTGCTACGCGGGAAAGCGTCGCCAGTTACCGTGGTACTTATTGGCCGT
AGCGGTTCTAGCCAACAATGGTTACTCGGTTATGCTCCACGGTGCGCATG
AACCGGGTTCGGGACGCCTTTATGCAAGTATAGCGTTACCCCAATTGGGC
CTGTCTGAAGTTAGCAGTATTGAAGATGCCAAACAGCAGTTAACGGATAA
AGGGNTGACTTACCTTGANTTGTCTTATTTGCTGCCTGCGCTCGACACAA
TTATTAAGCTTCGTGAGGNATTTGGCCTGCGTTCATGCGCTAACACCCTC
GCAAGGTTGTTGAATCCCACCGCTGCACAATTTTTCGTACAGGGCGTTTT
CCATATGCATCTTGACCACAAACACTGCCGCGTTAATGAAAATTATCCAT
CTATAGATGCGCTTTGTTTTCGTGGTGATGGCGGAGATCCTGAAGTTAAT
GGTGAACGGGAAACTGAGCTATTTATTACCAGAAATGGTGAGACCGAANA
AGTAATAATGCCGGTGGTTGCTGACAAATGGGCGCTTAAAGATAAAGGCA
TGAACGTTAACGACATGCTTGATGTGTGGGCGGGTAACGCGAGTAGCCTT
TACGGTGAACAAGCCGTTTTGAGTACACTTGCCAGCTATTTAGTGCTGCT
ACAAAAGTGTGATTCCGCAGATGCCCTGTCCCAGGCAAGGCAGCTTTGGG
CACAGCGCGATAAAAGCGCAACGCCATTTAAGGTCACCTGCGTTTTCCG
AGCGGGCAATTATCTATATTGCTTGATATAAAGCCCTTCTTTTTAAGGAA
AAATANCCNATCGTCNCATTGACTTAATCGTTGCTATACTTGTTAATAGN
AATAACNNNTTTTCCTTAAGTGGTGTATGCGAATCACGCANTGGNTCGN
ATTTNANCTNNTGTCGCTCATATTTGAGCAGGTNNTGCTGTATCTG
AAATCNANAACAGGTCACAACCTCTAGCGNAACTTTCGTACNANCCGCAN
ATATGCCTAANATTANCGACGCACGCNCANGGNGTTNCGCTNATCTCNN
NNNNNNNAAAACNNNNNNGNNNANTTANTGNAACCGNCTTCTTCATTTCC
GNGGTGGCTCGATTGNGNCGANNGCGCTNTCCNNNTTTGACCGTGGCTCC
NCACATTATNGATATNCTCAATTTCCCTTGAGCCAGATGNGANGNGGTCN
NAAAGAGAGAATTNNGTTTTCTNTGAGGNNGAANNATCTCTCNTNCNTAN

>reverse_contig_11_H17

ACATCCCGCGCTCCTAATGAACTTCGTTTCACGCATTATTTATCAGCCGA
CTTAGCTCATAACGAAGCCACCCTGACAAAATAGGCACTGGCGTATCACA
CTGAAATACGGTTAGTTGTCTGAAACGATAAGCTAGTATTTATTGGTTTT
GATTAAAGATAGCAGAGGAAGTGAAGTTTATTGATGTTTCACACAAATAA
AACTAAATTTTTTTGTTCAATAAAGATTAAAGATAGCCCAGTTTGGGCAT
TGACATAAACGCGCTACTTTTTGCTATACAAAAATACTCATGGCAGCGAT
GAAAACACGCAGCGATTCCGGCCGCCCGTTTTGGCGTCGTACAAAGCTCT
GTCAGCGGTATTAAGGATCGAAGTTAACGTCCAGTTGTGGTGATTAACGA

AATTGACCGTTGCAACGCCAAACTTGCGGTAATTACGGGTGAAACTGTG
CTCTTGCTGTGCTCTATCCCTAATGATTCTATCGCATCTCGAATCTCTTC
AGCTTTCACCTTCGCGCCTTTCTCATCGATATTAGGCAGAATCGCTACGA
ACTCCTCACCACCGTATCGAGCACAAAAATCAGTATTACGGGTAACTACT
TCATTAAGCCCAATAGCGACGGCTTTCANGCATTTCATCCCCTTTAGGGTG
ACCATAATTGTCATTGTAAAGTTTGAAACTATCGACATCCAAGATAATCA
CCGTCAACGGCCATTTCATACCGTTTTGACATTGCTATTACCTTTTCCATT
TCCATTTCAAATGCCCGTCGGTTTGGGATGTTAGTCAGCGGATCGCTAAG
TGACAAACGAATCAATTCGGTTTCTTGGCTCTTCACTCTCTCAAATACA
GTGCGAAGGTTTTAGCCAGGCGAGATATCTCGTCACTGCCTTCGAAGCTT
GCTACTTCTTCACTATTTTTTTCAACCAAGTTGGTTAAAGCAATTAACG
CAGAATAATACGTTTGTAGATATAGAAGATGATGGCGCAACAAATAATCA
TCACAATTAATGAAAGTGCCAATGACAGGGTTATTTGTTTGGACACAGCC
TGTTGGCAGACAATGCGTAAAATGAGGTGTCTTTTGTGATTGAAGATGC
CCTAAACTCAATATTTGACGCGACTTCTTCAGCCAAGTGCTCAACAAAAC
TGCTCGCCCTCTGAACGTCCACGAATACGCATGACATTAGCTTGTITA
TCAATCATGCCCATTTACCTAATACCCCTCCTCTAGCGCACTTACCAG
TTCGCTGCTGCGCTTGTCTAGCGATGTGCGCTCACGCAAGGAAGCGATGC
TTTTTTCAATACTGTCCTCTAACTTGCGTAACCTCATGCAGTTGGAAGTGC
TGATTTATTTGGGCAATTTGTAATATCAGAGGAACCAGCGTTTGGAGAGGA
TTGAATACCCACATACTGCGATTTATTGTCTTCAAAAATATTGTTCAAGT
AAGTAAATAGTACTTCCCTTTCCACCCGCACTTCTTTTTTCGACATCAATC
CGCTGACGCACTAATGCGTTTAGTTCTTCTATTTCTAATGTAAGTACATT
CAGCTGCGTCGCTAAGTACTTGTCTCCTTGTCTCAAGCTTCAAATTGGAGG
AATCCAAGCGCTCTATCGTTTCATCCAAGTGGGTTTTGACGATACGCCGA
GACGATTGGCTCTCAGACGATGTCAATCGCGCAGTAAGTGAAATTAAGTG
CTGAACCTCACGCGTAATTTCAAGTAGCTTGGCTAAGAGTGGGGATAGAAG
ACGAGGTGATATCGTCAAGCACTTTTCGCACGTCATTTAAACGCAGCAGC
GACAAAACCAGTATCAAAGAAAAAATCCCATCACACTGCATTGTGCAAT
GATAAACAGATACAAGGTCTTTTTCTTATGCTGCATCTGATCAACCGCCT
TTTGCCATATTTATTATGCGGTTATCAGTTTTAGGCTTTATATCGTATTG
ATTGACGATGGCCTGCATCACAAGGTCTGATATTTGAGGGGGACGTGTAC
TCAAGGCCGGACGCTTTTCGGTATTCTTTAACGCATAGAAACCGTCTCCG
CCATTGGCCAGAAAGTTACTTGTAGCCAGTGAATAGGTTTTTCGACTTATC
TATTCGCTTTCCATTAATCGTGACATTTTTTACTCGTTGAAAAACTTTTCG
CTGAAGGGTTCGAAACTATAAGACATACCCGATACATGAGGAAAACCACTT
TTTAAATTTTGATATTGCGAAAGCCCCTCTTCCAGAGCTTGAATCAACTG
AGCGCCAGTTATATCTAAAGCTACTACATGCGCGCGAAACGGCATTTCGG
TGACAATATCTTGCCGAGTAATCGGTGTACCTGCCGTATATTGTTTGTG
CCTCGGATACTGCCTCCGTTTATCAAGGCAATTTTCAGTATTGGCATAGCG
TTTAATGGTATCCGCGACAAAGTTTACAAAAGGATTCTCTTCACTTCTCA
CTTCTTTGCGCGCGGTTCGAAACGGCAATCTTCCATTCGCCAATTTGGATA
TTTAAACAACCGCTTTANGCGCGCTTCATACTCAATGAGTTGCTCGTTTAA
CTCAGCATCAGCCTCTATATCTTCTAACGGCTGTTTCGGTATACGACGTTG

CTTGCCACCCAGCACTTTTCACGAACTAGCAACGACAGCGGTGCCCTGC
TGCGTAAGC

HMS186_25m 28-P8

>full_28_P8_seq

ATATGTCTTACTATAAATGATGACACGCAATACGAGAGCCTCATAGCCCG
CCAAAAATTTAACTTTACAGTGTATAAAATAAACCTTAATAATATGAAA
ACACGGAAGAAGATATTGAATATAGTGATCTATGAACAAGCTACCCTTTA
GTTCTCTCAGTTATTGGTTTGCCTTAATCAGTTGTTTTTTGTTTATCTTC
GGCGGGTTCGCGTTGATTTTCGCCATATTATCTCTGGCGTTGACGATAAA
AGCACAACGGCAATATGATGAAGCGCCTAATGGTTTTACAAACATTAGCA
AAGTTAAAAAAGCTAGGGTTATCGCCATCATAGGTATAGTGTTGAACTTG
GTCATTCTAGGCATCACGATTTGGACATTAATGACCGTAGGATGGGACGC
CTGGTCAGAAGAGTTTATTCGTCGATGGAATGAAGGACTGGAGAGCAGTG
GCAGATATTGAACCACTGAGGAATCTCCTTCCTTCACCGGATTAATCGTT
TTAATTTTCTATCGCTGGGCTTAAAGCTTCACGTGCCTGTTTAGTTTGA
ATCACTAGCTGTTTCGCGAACGTGTTTGGCCGACAGGCACACGTCCTTCGAA
AGGCACAAACCACACGTTGTCGCCATCGGTGTAATGAGCGCCTTCAGGAA
ACACACCACACATTTTAGGGGCAAGAACGCCAACATTCAGTGTATGCTTT
AATATGAGTGACATTGCATCCACATGATCTTCGTTCCATATGGGTGATCAT
GCCGTCGACATTACTGTGCCATTCCTGTTTTTCTTGGTGCCAGTAGGCTT
TATCTACCCAATGGATTTTACCAAACCCGCCTATGTAACGTACCCGTTCT
GTTTTAATCACATAAAAAGTTAAAGTCATGGGTTTGCTTATACCCTTTGGC
TTGCGGAAAAAGACGAATGTACTGGGCTTCAATGTGTGACTCATTGACAC
ACATTGCATTCCCCATAATTGTGACTCGCCCATTCGCTTGGGAGTCGTCT
TCTTCGTTGTCGAAAATCGTTAAGCTCACTTTGTCGTTGGCTTTAATATT
GCGCGTATGCAGCGCAATGTCACTGATATAAATAACCGCATCACCTTCAG
GTGTCAGGTAGTATGGCACCACAGAACCAAAGGGATAACCAGGCATCGAC
GTAGAGTGCGTGCCAAGTACGCCACTGTGTTGTGCTCTGGCTAATTGTTT
TGCAGCAAGTGCAATTCGGGCATGCCCATGGTTTCTTCCTCTTCTATTT
CGCTTAATGAGTAACAAAAACACCATGTTTCAGGATGTTTTATAAGGTGTA
AATCAGTTTGGTAACATTGCTTTAAAATCCGCTGATCAAGCACATCCGCG
GGCGCTCCATCAGCAATGACACCTTGATTAGCTAGCACCAGCATTCTTGT
ACAATAGGGCGTCACAAGATTAATATCATGGCTAATACATATAATGCTAA
GGCCGTGTTGGCTTAATTTTGCCAATAACGCCATCACCAAGTTTCTGGTGG
CGTAAGTCTAGCGCCGATGTGGGTTTCGTCCAGTAAAAGCAACTTGCCAGT
TAAATCGTCCTCTAGTGACGAGGGGACAAGTTGTAGCAAGGCTTTGGCCA
AATACACCCGTTGTTGCTCTCCTCCTGAGAGCGTTTGAATATTCCGGTGC
AACAGGTGGCTTAAATCAAGTGCATCAGACACAGTCGCAATACACCCACG
ACGAATGTGGCTTGGTTCTATCATTGCGTGTGCGCCAAGGTAAAACCTT
CATCCACTAAAAAACAAAAGGCGAGGTTTGATGCTGACACATCACCGCT

CTCACTGTTGCTAACTGTTGAGCGCTGTAGTGCTCTAGTGGCGTGTCATG
CACATATATCTTTCCGCTCACCAGTGCGCGAAATGCCCCCTCTCCACCGC
ATATAATGTTAAGCAGTGTGGATTTACCACACCCATTTTCTCCGGCTATA
GCAATAAATTCTCCGGCCTTTACGGTGAAATTAACCTTGTTGAGCAGTGA
ACGCTCCCCCGTTTTACCTGCACCGCTATAACCTGTAACACCGTATTTT
CTGGTGTTTGAGTAATGCTATTGTCCACCGCTGAACACTCCTTTTTGCTG
CCACAAAAGATACAAAAAGAACGGGGCGCCAATGGCAGAAGTCACAATGC
CAATCGGTATTTCCATTGGTGAAAACACGCTGCGTGACAATGTATCAGCC
GCACTTAATAAGAGTGCGCCAACAAGCGCTGACGCTGGAAGCAATATGCG
ATGGTTTACCCCAAATAGCGTTCGACAAAATATGGGGAACCACTAAACCGA
CGAAACCAATTATCCGCAAAGCGCTGTGGCCAGCGCCACACAAAATGTC
ACCAACCATAACATGCGCTTTTTATAACGTTTTACATCTAATCCCATGTA
GGCAGCCTCGCCCTCACCCAGGGCCAATAAATTAAGTGCATTTTGCATTC
TGAATAGCCCAATAAAAAGCGGCCACACAGGTAATACTGACGATGATGGCT
TTGTTCCAATTGGCACCTGCTAAGCTCCCCATTGTCCAATACGAAATCTG
ACGTAAGGCGTCATCACTGGCGACGTAGCTGACGGCTCCTATAATCGTGC
CGGTAAACGCATTAATCGCTACGCCGGCTAAGATTAACGCGGCCACAGAT
ACGCCTTGCGGGCTTTTTGAAAATTTGAGAACCAATAACACTGCAGCTAA
GGCACCAAAAAACGCCCAAACACCTACCCAGAGAGATGCCATAGCAGGGG
GGATAAAAGAAAGGGCGCTGGCCGGCATCACAGCCATGCTTACCGCAGCT
GCCGCTGCCGCTCCCCCTGAAATACCAATTAAGCCGGGGTTCGGCTAATGG
ATTCGCACTAAGCCCTGCATGGTACAACCTGCCGTTGCGAGCAACGCAC
CAATTAATGACGTCATAATGACCCGAGGCAGTCTCAGCTGGGTAAATACA
TACCAGTGAGNGGNGTTA

HF186_25m 22-F8

Individual Transposon Insertion Sequence Reads

>4a_kanR

GACTACGCACTAGCCAACAAGAGCTTCAGGGTTGAGATGTGTATAAGAGA
CAAGGTCTCGATTCGCCTGACGAGCAAAACCCCTATGAAAGGGGCCAAGA
GCGGCCGAGCCTAGCTAATTCGCTGGAGGTATTTAAATCGACCGTAGAGA
ATGAAGATTATGTTATGGGCGAAATGCAGCAAAGGGCGGCGGAAAGTGGG
CTTCTAAAAGAGATAATCTTTGGTCGTAATGAGCCTGCGCTTCATCACTT
CCACAACAACCTCCGCGAGGCATTGGGAGAACCTCGCTTAGAACAGTTAT
AGCCTAGAAGCTGCCCTATGTTGGCTACAAGACGTGTGACCACCATAAAA
ACGCTCCTGGAGCGAAGCTCCAGGGGCGTTGAATGCAAGGCGAGCACGCA
AGTCGGATGCTAGTTGAGCTTGGCGCTGGTGTTTTGGGAATTCGCGGCTA
AATCTAGTGCTCCCGCCTCCATCGCGCTCGCAATGGCACGTCTCGCAGCC
TTGCGGCAGAAGGTTACGGTTTTAGCTTCGCCAACGTACATTCGGTTCGCC
CAATCGGTTCTTGAAACAACCTCATGGTCAATTTCTTTTCGATTGAGATAG
TAGACTCTCCGGAAGGCAGCTGGGCATTAGGTTTCGATAGCCATTGTGGAT

GATGCAAAGCTAAGGGCGCCAATGGACAAAAGAGCAGACAGTGTTTTTCAT
GAACTTTCCTAAATCTGTTACTGGAATCGGGTTACAAAAAGTAACATACG
ATTCTGCTTAATCATTACCATGAGTAACACTTTAGTTGCGACAAATTGTC
GACATAGTAGGGGTCTAGCCGTTTTCCCTCCGCATAGTCGGCTTC

>4b_kanR

ACTACGCACTAGCCNACAAGAGCTTCAGGGTTGAGATGTGTATAAGAGAC
AGGTGCTGGCCACTCGCGATAAGGAAGGACGGTTCGCGCTTTTCTAAAC
GCATGCAGACATCGTTCGGTTAAGGTGGCGCGGGAGGAGAGGGGCACTAA
GCACGTTTTTACCTGTCCATTTACCATTGGAGCTATGCAAACCTCGGGTG
AGTTGATGACCATTCTAATAATGATCACTTTGGTGAAATCGATAAGTCC
TGCCACGGGTTGGTGGAGCTGCCTGCTGTCGAGAATAACGGGCTGCTATG
GGTTCATCCAAATCCGGAGGGACAGATCGACATCAGCGCGCTCCTCGGTC
CACTGGCGCAAGAGTTCTCCTCCTTTGGTATGCAAACCTCAGATCGCTGTT
GGACGAACAACATATAGAAAAAGCACTTAATTGGAAGCTCGCCAATGATAC
GTTCCGGCGAGACCTATCACTTTGGAAAGCTACACAAAGACACTCTGGGTC
GCTTGTACTACGGTAATAATCTGCATTTTGAGGAGTTCGGCCGACACCAT
CGTTTTGTGACAGCAAATCACGGAATGCAGGAAATGCGTAACTTACCCGA
GAGCGAGTGGGATATCGCCCGTTGTACCTTCGTGCTTTACCACTTGTTTC
CAAACGTACAATT:GATTACCAGTCATTTTTCAACCACGCTTATTCGCAT
ATATCCTGTCAAGGACCAACCCAATCGATCAGTCACGCAAATCAGTTTTT
AT

>4d_kanR

CTCCAATGGTGAAATGGACAGGTA AAAACGTGCTTAGTGCCCCTCTCCTC
CCGCGCCACCTTAACCGAACGATGTCTGCATGCGTTTAGAAAAGCGCGGA
ACCGTCCTTCTTATCGCGAGTGGCCAGCACTGGCACACCGAAATCCTCA
AGGGTTAAGAATGAACCCTTAACCGGTAAATCCCAGAGAGACCAATTA
CTGGGGGTGATTTAAAAAAACTCCTGCTGTTTCGCGCTCGGCTATCTCTG
GACAAGTATAGGAACTTGTTCGGCATCCGATACTGCACGCCC GCATCTACG
TTGCGCCCTTCATCAAGTTGGGCCATGAGTTCTTTGATAATCGTTACCTG
AAGATCGTGCTGCATAGGACCGCCCCTTAAAGCGTTTATTATTGAAAATT
CTANAAGCTGAATGCTTCTGGCCCATGATTTAACCTTTACTCTTTTTCA
TAAGTCGATTGAGCCTGATTAAGATAGGTGAACTCATCTTGCACATCTA
CCATACTAAAGGATAAGCGATGATTTTTGATCGCTTCTCAAGTGTTCAA
AGCTAACACAACAATCCATCGCCAGCTTGAAGAATTAAGATATGCTCACT
TGCGTTTTTTCTGTANCAGATGTGCTCGCTTCGTCAGCTCATT

HF186_25m 13-D19

>13_D19_forward

CTCAAGCTCGGCCCGACCCTGAAAGGCGGTCAACGCCACATGGTGCCGA
GAGGTGGCAGCGTGAGCGTCCCGCTGCCTCGCCAACGTTCACTCACAAC

CATGGCCATGACACGACGCGCTGACGCACGAAAGGCCAGCTTGAAGGAG
CCGCGGAAGGGCACGACAGCGTGAGACGAAGGCCTCGCCCGTGGGGCCAC
CCCATCGGGTCACAGCGAGCGAGCGAGCATGGAGAGGGCGTGGGGCA
CCGCTTGGGTACGCTGGCCCTCAGCGTCGTAGCGCCGATTCCAATCCCA
GAAACGATAGGTCACCCCGCGGCGCCCGGGGTACGAACTGCGGCTCAA
TCAGCGTGACCCCCCGCGATCGCGTGCGGGGTCCCGGCGGCGATCACG
AAGGCGTCCCCGGGGAGACCATGACAAAGTTCATCAGGCGATCGAGTGG
GCCGCTGCGCTCAAGGCAACGGACGACATCCTCTCGAGCGACGCCATCTT
GGAACCAAGATAGAGACCAGCGCCCGCGCTCGGAGCACGATCCAA
GCCTCGGGCTTGCCCGATTCTGTCGAGCGCTAGCTCAGGGTCGCCATCTCG
CGGATGAACCTGAACCGAGAGGTTGTCTGCGGCGTCGAGGAGCTTGATGA
GCAGGGGCGTTTGTCCGCGTAGCGCGCGGCGACCTCCGGACCAAGCCAG
CCGACGGGGTCGGCGGCTATCGCGTCGCTCAGCAGCGTCTCGGTCCCCGC
GAGGAGGCTCGGGAAGGAGGGGCTCTACGGAGACCTCCACGACTCCCCGA
CCTTACCTGTCCGGTGGAGGTCGAGATCGGCCTTGTAGTGACCTAAAATC
TCCTCGCCGCCCATGGCGTTCGCGTGGGCGGGGTGAGGTTATCAGAGCG
CAGGAGGCGAGGAATCATGCGTTCGGAGTATCTCGGCGATGTGGCCCGCC
CGAAGGGCAGAGCGCGGGGAGAAAAAAGGGCCCCGAGGTGGCAACCTC
GGGGCCCTGGGCGAACCGCCGTCCGGGGGGTGACGGCGGGAGGGGGGTG
CTCTTAATCACACCCGTTGGCGTGAGCAGGGTTGTGGTCTTGATGGAGGG
TGAGCACCTGCGCGATCTCGCCGGAGGCGAGGTCACGAATATCATCGGCA
CGAACCCCGAGGTCGCTGATGGTGTAGACAATGTCATCCATGAAGACGCT
TCGACGGATCTGCGCCTCCAGCCGTAGGCGTCACGATCGCCGCACGCAC
CATAACCTTGGTAGGCGTCAGCAAGGCTGTCATGGGTAATCGCGCCGGTG
GGGGTGACCCCGATTGAACGTCTACCTCAAAGAGCTGGAGCTTGGAGAC
GTAGCGACCGTAGTAGCCCCAGTCCTGCTCCTGACCCTCATCGTACTCCC
ACCCGCTGACCGGGAGCGCAAGAGGCCTCGCGCCCCGAAGTAGTTGAAG
GCGTGGTGGTCGTATTGGGCTTCAGACCAGGTGTTCCAACCATCTCCGAG
GACCGTCTTCTGCGCCAGGGTGGGCTGGGTAGGATCGCTGACGTCAAAGA
TCTGAAACTGAAGCCCCTGCACGACACCCTCGTCGGTTCGCGTCACGGCCG
ATGGTGAGCAGGTGGTCCCTCGCCCAAGGGGTGCATGTAGGAGGAGAAGCC
GGGGATCTTGAGCTACCAAGTCACCCGTGGCGCGCTGGGCTTCGAGAGAT
CGAGCACGTAGAGAGGGTCCACTTGACGGAAGGTCACGACATAGCCGCGC
TCCCCATAAAGCGTACCGAGTAGATGCGCTCACCCCTTGCCGAGCCCCC
TACATGGCCGAGTTCTACGAGAGCGCCGGACGCTTCGCCGAGCACGCTCA
CGCGGCTCTCGCTCTTTCATCATACCCACCAGCCCCAGGCGCTGGGC
TCCGTCGTCGCGACCCGGAGCGCGCCCTGGTACTCGTCCATGGAGAACTG
ATTGAGGACATAGCCTGGCACCCGAGCCGCTGGCCACGTAAGCGGCTCGAC
CGCTTGTGGGCGAGATGTGCAACTTGTGGATGTGGCTCATCACGCGGGGA
GCCTCGTCATCGGTCATCCAGTACCAAGCGTAGCCCCAGTTGGTGGACGC
GATGTACATCGCGTCGGCCGACGCGTAGACCTCGCCCCAGGTGCCGGGGA
CAGCCGTGCCGAAGGCTTCGGTGGTCTCGAGGTTCAAGGTGACGACCGTC
GTCAGGTTTCGCGCCCCGAGTGCAGCGCTCGGGAGGTAGACATGCTGGCAGTC
TGCCGAGAAGGCTTGAGACGAGAGATCAACCTCGCCGTCTCTGAGAGGG

TGAAACGGTAGGGGAGGAAGTCATCGAGGGTCAGCGCATCAATACGGCGC
AGGTTCTCTGCACCATCTCGTTCAGGGCGGAGAGCTGGACCTGAAGAGG
GGCGTCCCAGTCAACGTCCGGCCAGTACTGAAGCCCATCAATGTAGGTCC
ACGTGTTTTGGACGAGGTAGACGTCGGAGCCGATACGACGGGTGGCTGAG
AGGTCCCCTTCAAAGGCCTGCGCATCCGTGATGCGTGGAGAAGCGGGGTT
GGCGATGTCGACGGTGGTGACAAGGGTCAACGACTGCCAGCGGTGCCAAT
CTCACCCCCACGCTCCTCGCTCGAGCTCCCTGTCTCCCGGAGCTCGCCG
CGATAAGACGAGGAGAGCACCACGACGTGATCGTCGTGCAGATAGAGCCC
GTGAGGAGACCCTGCCACCGTGATCCGGCTCACCTCTACCATGTCATCTG
CAGGCCAGGCGGCGATGATCACGAGGTCGTTTCCAGAGATCGTGTAGAGG
TACTCAGCGTCGGTCTTACGCGGTACGCTCATCGACACCCGCGACCTG
AACATTGGTGTGAGAGACGTCATTGGGGGAGCTCCCGGAGGCCCGCCTG
GCGCCGAGAGGGCTCGGCGCTCTCCACAAGGGCGACGTCTTCGGCGACG
CCGCGGTCCAACCACATCCCCATTCCCTCACCGTCTATCATCTGTGCGCG
AACCTGCTCGGCCTGCAGAAGCATCTCCGTGCGCGCGTTGTCCTGAAGCT
GCGAGAGGGCATGGTCGCAAGAGCTTGAGGCCCCCAAAGAGGCCGTGCGC
ATCCAGCTCATTCTGTGACGCTCGCGCCCTCACTGAGGAGGCTCTCTGT
GACGCTGGTAGGCGCCTCGTGGCACCCCGAGAGCCCGACGGTGAGCGCAG
CGCTGAGCGCCATCAACGCGGCGTGATAGAGGGGGTGAAGCATGGGGGAG
CGCATCATGATGACCTCCTGGTTCGTCTCGGGTGAACCTCGATACGTCACC
TCTGCAACCCGCGCGCCAAAACCATAAGCAGTTGATATTACGAACATCTC
ATCACTTCACTGCGCGACAAGGTGTGTAACAAAAACACCACATGATGCGAT
TATGCTATGCGACGCGGGGAACATGGCGTGCCTCTGTGGCGTCGTGACCC
GCCAGAGACAGAGAGCCTTTACCTCGTCGCAGCGCTGAGGTGAGATAACC
TTGAGCGGTCTGAAGAGGGTGGGGGTGAAAAGATGGAGCAACGAGAGCTA
GGCAGAGCGGCCTCAAGGTGAGCGTTATCGGGCTCGGAGGGATGCCGAT
GTCGATCGTCGGGCGCCCCACCGAATCCGAGGCCATCGATGTCATCCACG
CCTCGCTGGAAGCCGGGGTACATTCATCGATACGGCGGATGTCTATTGC
TTGGATCATCGCGACATTGGCCATAATGAGCGCCTCATCCAAAAGGCGCT
CCGTCACTGGGGGAGCGCCACGAGGTGCTGGTAGCGACCAAGGGGGGGC
TTGAGCGTCCGCAAGGCGACTGGACGACTAACGGCCGCCCCGCGCACCTG
AGGCGCGCATGTGAGGCGAGCCTCCGCGCGCTCGGCGTGAAGCCATCGG
GCTTTATCAGCTCCATGCGCCCTGTGATGTTTCATCCCTTCGAGGACAGCG
TTGGGGCGTTAGCGCGCCTTCAAGAGGAGGGGAAGATTCAACACATCGGT
CTCTCGAACGTCGACGTCGAGCAGATACCCAGGCTCAAGCACTCGCGCC
CATTGTGAGCGTACAGAACCGCTGCAACCCTCATGATACCCAGGCGTTCA
GCGACGGTGTGCTCTCGCGCTGTGAGGCGGACGGGATGACCTTTTTGCCT
TACTCGCCAGTCGGGGGCTTCGACAAAGAGGACATCGGCCAGGACAGCAC
CCTGGGTGAGGTGCGCGGCGACATGACGCAAGTCCCTACGAGGTGGCGC
TCGCCTGGCTCATCGCCTCGTCCAAGGTTCATGCTCCCCATCCCTGGGGCG
AGCAAGGTGCGGAGCGCCCGCTCGAGCGGAAGGCAGGGTCCCTCAGGCT
GACGGCTTTGGACATGGCGACCCTTGACGGAGCGTACCTATAAGAGCGCG
ATTTGCCGCCCTCATGGGGCCGTGTGAACGTGTCAGTCTCCAGGGGGTCT
TTCCTTTGAGAGCATGCCGAGGCGACGCGCCGTGCTCACAAGAGAGCCT

CAATCGCCTCGTAGCGCGACCAACGCCTCCGAGGCGACGACCAAACGG
TGATCGTAGGTCTCTCCATTGCGACACGCTCGAAGCTGATTCATACCCG
ATCGGCGCGCTCGGGGAGCGCGTCCCTCAAGCGGGCCTCAAAGACCTCGC
TGACGACCGGCGAGAGCCGAAGGAGGACGAAGCTCGCGGCGCGAGCGCCT
GCGTCGTGGGCGGCCTCTAAGATCTGAGGGATCTGCGCATCGGCGAGCCC
CGGGATGACCGGCGCGATGCTGACCCGCACGTTGAGGCCCGCGTCACTCA
GACGCCTGATGGTCTTCAACCGACGCCTGGGGCTTGGAACATAAGGCTCG
ATAGCTCGACTCATCCCCGGATCCATAAAAGGGAGGCTGACGGTACGGT
ACACGCTGCGCGCGCATCCAGGTCCCTGAGGAGATCGATGTCTCGCTCGA
TGAGCGTGGACTTCGTGATGATGGTCACTGGGTTTCTGTAGTCGCGGCAG
ATCTCTAAGCAGGCGCGGGTCAACCCGTAGCTGAGCTCAAGGGGCTGATA
GCAATCGGTGTTGCCTGAGAAGAGGATGGCATCACCGGACCATGAGGGGC
GATCGAAGGCCTGTCTGAAGCAGCTCAGCAGCCTGAGGCTTAATGACGATG
CTCCGCTCAAAGTCGGTCCCAGCGCCAAGGTCCAGATAGTGATGGGAGGG
TCGCGCGTAGCAGTAGGCGCATCCATGATAGCACCCACGATAGGGGTTCA
GGCTGAAGCGGAAGGGGAGATCGGGGCTCGCGTTCTTGAGAGGATCTGG
CGGTGTGATCCTCAAAGACCTTCAGCGCCGTCGGCGGAGGCATGCCATC
AAACTCAAGCTCGGAGGCTTGCCAAGGGTTCGGCGGGTTGGCGACCCGAC
GGATGCTCTTCTCCGTGCTCATGAAGGCATCTTATAGGATCATCCGTTCA
GTACCAAGCCCGTGCCACCGTGGGGCGCGCTATGCTAGGCTGGCGGTCTG
AAGAGGGTGAGCCCATGGATAGCCAAGCCTTTGAGCGCAGCGTCGCCGA
CTGAGGCGGGCGGAGCGGGTCTTCGTTCGTACGCGGCGCGGGGATAAGCGC
CGAGTCGGGCATCCCGACCTTTCGAGGTTCAAGGGGACTCTGGGAGGGGT
TCAGAGCCGAGGAGCTCGCGACCCACAGGCCTTCGCTGCAGACCCCGAG
CGCGTGTGGCGATGGTACCGATGGCGCGCCGAGCAATACGGCGATTGCGC
GCCAAACGCGGGGCACGACACCATCGCGGCGATGGCCTCTCGCTATAGGC
GACTTTTGGTCGCGACCCAAAACGTCGACGGCCTCCATCGCCGCGCAGGC
AGCCCAGCGGTGTCGAACCTTCACGGCTCGATCGCTGACATGCGTTGCAC
CGCTGCGACACCTTGGCGCCCTTCCCGAGGGGCACAGAGGACTCAGACC
CCGTGCCGCGATGCGCGCTCTGCGGCGAGCGCATGCGGCCGCACATTCTT
TGGTTCGGGGAGACCTATTGGCCAGGGATCCTTGAGCGCGCGCAGCAGGC
CGCAGAGGAGGCGGACGTTTGCCTCGTCGTCGGCACCTCTGCGCAGGTCT
GGCCCCCGGTGCGCTTGGCGCTCCATGCGCAGGCCTCTGGGGCCTTCTC
ATCGACATCAACCCCGAGGCGACCGAGCTCTCCGGGCGCGCGGACGCGCA
TTTAGCCGGCCCTTCAGGGCGCCTCTTACCCGAGCTCTGGGAGCGTGTTG
GGCGATGACGCGGTCTTTCGATGGCTTTCACGCCTGCCCGGATCGCCTTG
CCTTCTCCGTGCGGGGCGTTCGATCTCAGCTATCGCGACCTTGATCGGGCGC
TCCGCGTCGGTCGCACAGGCCCTTAAAGCCAAAGGGATACGGCCTGGTGA
GAGGGTCGCTGTTAGTCCGAGACCTCAGCCTCGTTGCTGGGCGTCATCG
TGCTCACCTCCGGCTCGGCGTGATCCATGTGCCTATCAACACCCGCTAT
CAGGCCGATGAGATCGACCACATCCTGGAGGACTCGCAGGCGAGCCTCGT
CTTGGTTCGATGACGGGGCGCCCGCCGGATGTCGCGGAGGCGCGTGGGA
CGCCATGGGTCCCTGTGGACTCCCTCATTGGCGACCCTGCAGGCTTTGAG
GAGAGCTGGGAGAGCGACACCCAGCCATGCTCATCTACACCTCTGGGAC

GACCGGGCGCTCCAAAGGCGTCGCCCTAAGCCTCAACGCGCTCAGCGCGA
ATGTCGGCGCGACGACCACCTCTGGCATTGGTCACAGGACGATCACCTT
GTGCTCGCTCTGCCGCTCTTCCACGTGCATGGACTTGGTCTGGGTGTCT
CGGGACCCTCCTCACTCAGATGAGCGCCACGCTTCACCCTCGCTTTGACG
CGGAGAGGGTCATCGAGGCCGTCGCCGCAGGGGGCAGCATCTTCATGGGC
GTCCCCACCATGTACGCGCGCCTCATCGCGCACCTGAAGGGCGAGGGCCA
AGCCCAGAGGGCCCTGCGCGCCGCGCGCCTCTTTACCTCGGGGTCCGCCG
CGTCCCCGCGTCGGCCCATGAGGCCTTCGAGCGGCTGACTGGCAAGCGC
ATCCTCGAGCGTTACGGGATGAGCGAGACCGGCTTCACCCTCTCCAACCC
CTATGCCGGTGAGCGACGCGCCGGCAGCGTCGGCTTCGCGGTCCCTGGCT
ACGAGGTCCGGGTGGTCACTGACGAGGGCGAGGACTGCGGGCCTGGGGAC
GCTGGCGAGATCTGGGTACGAGGCGATGGACTGATGCTCGGCTACTGGGG
CCAGCCAGAGGCTAGCGCAAGCGCCTTCACAGAGGGCTGGTTTCGGACAG
GTGACGTCGCCCCGCGTCGATGAGGACGGCTACCACCATATCCTTGGACGG
CGCTCTGCGGACATCATCAAGAGCGGGGGCTTTAAGATCTCCGCCTTGGA
GATCGAAGACGTGCTACTAAGAGACCCCGCGGTCGCCGAGGTGGCCGTCG
TCGGGCTACCGGATCCCGAGTGGGGAGAGCGCATCGTCGCGGGCGCTCGTG
CTGACGCCAGGCGAGACGATCGACAGCGAGCGCCTCGACGCGCTCAGCAA
CGCCACCTCGCGCGCTTCAAGTGCCACGCGCGTACTACTCGCTCACCG
AGCTCCCGAGAAATGCGCTCGGGAAGCTCCAAAAACATCACCTCACCGCG
CAGCTACTCGCCATAAAGTAGCGTCACGATGTCGCCGACATTGGTCCCCG
AGGGACCGGTCTGATGAGCGCACCGACGGCATCGAGCGCAGGGTGAGAG
TTGTTATCTGCGAGGGCCTGGTCGATGGCGTCAAGCCCGGTGACGCGCTC
CGCGTCAATGATCGCGCCCGCGGCGTCTGTGGGCCATCTCTGCCATCGG
TCCCAGCGCTCAAGAGGGCCAGGGCCTGCCAGAGGGCTCAGCGCGTCGA
AGCTGAGCAAAGGCCAGGGCCAGCTCTTGGCTTCGGCCCCCTCGACCAG
ACCCGTGATGGTGACCGTGGTCTCGCCACCGGCGATGAGCGCGATTCGCC
CCTGAGGGCGCAGCTTGGTCAGCGATCCATCGCGCCGCGTCCCGCGCCTCG
CCTTCTAGATAGGGACTCAAGAGGATCACCCGATAACCGAGCGCCTCGGC
GCAGGCCCGCGCGCCCTCGATCGCCTGGTCGATCCCAGCGACGATGGTCA
TGTCGAGGCCCGCGATGCGAGGGTCTCGCGCATCCGGCGTGTCAGGGAAG
GCCCCTGCGTCTCCGAGCTCCAAGTGATTCCGCACCGCCTGCGTCACCGC
CCCTAAGAGCCCATGGCGCTCAAGGACGCGGAGGGCGTCCCGATAGGTGG
TGGGATCAGGCGCCGCGATCCCAGAGCCGATCGCCTCGGGAGCATTGCCT
AACACGTCACTGAGCAAGAGCGCGCGCACCCCCTGTCCCTCTTGAAGCGC
CTGACCACCTTTGATCCGGGAGAGGTGGCGACGCACCGTGTTACGCTCCT
CAATATCCGCGCCGCGACGCATCAAGAGCTGGGTGCGTTACGATATCA
TGAAGCTCCAAGCCGAGGGCTGGCGCGCAGACCATGGCGCTCGCGCCGCC
CGAGACCAGCACCAAGAGCTCCTCATGCTCTTGGAAGGCTTGCTCGACCG
CGTCACCACCCGCGATGCTCCCCTCCAGGCTCGGGGTGGTCACCCACC
AGGACGCGGAAGCCAGCCGAGGCCGCGCATTGGCCATGTTTCGTGACGAT
GAGCCCGCGACGCGCGCCCCTATCCCGGGCCGCCCCCGCCATGGCGCTCG
CGGCTTTACCGATCGCGACCACGTAGGGGTCTCCCTCTAAGGGACGCCTC
TCAAGCGCCGCCCGCACCGCCTTCGCGGGCTCGACGGCGTCGACCGCCG

GTCATAGATCGCGCGCGCATCGGGCGGAAGCGCTGCTGTCCCTTCACTCA
TCGGCACGCTCCTACAAAGGCCGCGACGCGCTCTGCGCAGACCGGGCGCGT
ATACCTGCCCTCCCCGCTTGAGGGCCGTCCCCACGATCGCCCCATCCGCG
ATGGCGAGCGCCTCTGCGGCCCGTCAAGGGTTCATGCCGCTCCCGATCAG
CAGAGGGACCTCTGGGCAGGCCGCTTACGGCCTGCATGTCTGTTATCG
ACGTCGGTGCCTGTCTGTGAACCCGAGACGATGAGCGCGTCTGCGCGC
CCACGCGCGGCGGTCTCCTCGGCGAGGAGCCCCAGGGGGCGAGGCTCACC
GAGCGGGCTCGCATGCTTGACGTCAACATCGGCAAAGAGAGCGGCCCTGG
CGCCAAGCCTCGCCCTTGTGCGGAGCGTCGTGGCGGGCGGACCCCTCGATG
ATCCCTTGATCGGTCAACGCCGCGCCGATGTGGACGTTACACGGATGAA
GTCGAGGTCGAACGCCGAGGCCAAGCCCAAGGCGGCTTGAGCATCATTGC
GCAACACATTGCAGCCCACGAGCACATCACCCCGAGGGAGCAGGGCCAGG
GCCCGGGCCATGGCGGCCACCGTCTCGGGCCCAACCTGCTCCGGGTAAAA
AGGGGCGTCACCAAAGTTCTCCACGACATAGCCGTCCATGCCCCCGGCGA
TAAGGGCCTCAGCGTCGTGGATAGCCGCCTTAAGCACGGCGTCCATGTCC
CCGGAGAAAGAGGGGCTCCCCGGCAGAGGCGCGAGGTGGACGACCCCAA
GAGGAGCGGGGACTATCGGGCGACCCGAGGGTCCGGGTCTGAGCTTTA
GCATGGCTGGCGGCTAGGTGAGAGAAGGGAGGGAACGTCTCCATCGTAAG
AAGAGAAGGGCCAAGATGGCCAGCCCAGCCCCAAGGGGAGCGATGGGTG
CCGACCTGCATGACATCCACTACGTCCTTCAACCCAGACGCTCCCGGCCG
CGATGCAGGCATCCTGCGTCTCAAGCGGGCCCCAGCTGTGACCCGCGGGC
AGGCAGGCATCTCGGGTCGTGAGCGCACCCCAAGTGGCGCCAGACGCGAC
GCAGTCAGCGCTCCAGTGAAGCTTGGATCCGAGCACCCCTGAGGCCGAGC
ATCCCTCGCTGAGCTCACACTCAGGGACGCTCGCCGTGCTCGCGTACCG
TCGCCAGAGCTGCTCTGCGAGCCAAGGGTGCACCTTGCCGGAGGCGCTGTC
ACAGACAAGGCCGTCATTACAAGGCGCCTCCGCCGAGCATGCACGCGTGC
AGATGCCGTCAATACAGACGCTGTAGCTGCAATCCGCGTCTCGGTGCAC
ACGCCGCCGGCCTCGACGAGTAGACATTGGTTGATTTGCACATCGCAGCG
GTAGCCCGCTGGGCAGCCCACCGAGCCTCCGCAGCCCCGCGTACAAACC
CGTCGGCGCACATCCCCGAGTCGCACTCGTCGTCGAGCGTGCAGCCTCCT
CCGACGCTCGCGACCTCACACATCTGAGACTCGAGAGCGCAGAAGAGCCC
CGGCGGACAGGGCGCCTCATCGGAGCATGTCCGAGAGCAGATCCCCTCCA
CACAGACGGTGTAGCTGCAGTCGAGGTCACTCTCGCAAGGGCTCCCGGCT
GGGACCAGCTCACAGATCCCCACCGCAGCGTCACAGAGGTAGCCCGCTGG
GCAGGGCGTGTTGTGCGCGCACGCACGGCTGCAAATCCCGTCCAAACAGT
ACCCACCCGGACACGCCAGGTCGTCTTTGCAGCCCATGCCAACTTTGTCG
AGCAAGCACGCGCCGTTGGTGCAATTCCAGCCCTCAGGGCAGTCGAGATC
GGCGTCGCACTCACGGGTGCATAAGTCGCCGAGACAGATATCAGAGGAGC
ACTACTATTGTGAGACAGGTGACGCCGATCTCCGGACGACAGCCGACT
GAGACCGGGTTCGCACTCGAAGCCCGTTGGGCAGGGCATATCCTGAGAGCA
GGTGCGGGAGCACGTCCATCAAGGCACCACCCACCGAGGCACTCCGAAG
GCTCAAGGCAACGGCCACCGACGGGCACGACCTCACAGAGCTCGTCACCG
TTACAGAAGAGACCCGTCGGACAGGGCGGGCGTCAATTGCAGCGCCGCGT
GCAGATGGCGCCACACAGTAACCAGACTCACAGTCACTCCCGCCAGCGC

ACCAGCCGCCGATGGGGTAGGAATAATCAGGGTGCCTCGGCCAGCCTAGC
GCCAAGGCGGCACATCCAAGCCGCCGAGCTCCGCGCGTAAACCACGGT
CTCTTTGATCCAGGCGCCCCAGGTGTTGACCCCCGAGTACACGGGGCTTG
AGCACTCCGCGCCCCCGCGGACGCGACACCAATGATCCGATTGAAGGCA
TCGATGGCGGGCCCCGCGCTGTCGCCCTGACAAACCCCGGTGTCGCCGAT
CCACTCCGACGGCGTCAGAGGATACCAGCCGGGGCACTCGCCCGGGCCAC
AGTTGGTGTAGAGACCGTCTCGTCGACGCCGAGTCCCAGAGCCTTGTCCC
TGTCCATTGGTCGAGCCGTAGCCGATGGCGGAGTAGGGGTCGTTGGCGGT
ATGACCGGTGTCGACGCGCGGTATCATCGGTGAGGCCTCCCACCACTCAG
GGATAGGCTCCTCCAGGATGAGGAGCGCAATATCGTACCCACAGGCCAG
TTGGAGGCCTGCGGGACGACGATCTGCTGGCATCGGATATAGGCGCTGTC
ATCCAGGGCATCGGCGATCGGTTGGTCACGAAGACATACTCAGGGGTCT
GCGCAGGCCCAAAGGTGCTTTGGTTGCAGTCCACGTAGTTGTTTCGTCGAC
ATGCTCGCGATGCAATGCTGAGCCGTACGGATGACGTTCCGGGCTGATGAT
GGTGCCGCTGCAGGTGCCGAAGCCTCCCTGGCTCGCCATCCCGAGGCCGA
CAACGGCCTTGTCCGTCGAGTTGTAGCCGCCCTAATCGCGAGCTGCGCC
GTCTCGAGTTGCACCTTATCCGCTGTGCTCTGCTCGGCCTGGCACCCCAA
CGCAACCCATGAGAACGCCATTAACGTACTGAATACACGCCAATCCACCG
AAACACCTCCGCCTGGCCCGTCACGAGGACGAGAGCATAACAGGGTTCGAT
GCGTTTCACGAGACCGTAATCCCGAACCGACGCGAAATCTCTGCCCTCT
CCCGATGGGTGAAGCGGCCGGAGCCCTCCACAAACCCGTTGCAGCATCAC
GAAGACGCGCCGCGCAGGGGTCCCGTGCACCCTGCCCTCGTATGTGGGCA
GAGAGGGGTGATCACAGGCACCGAGCGCTCAAGGGGCCCGCACCTCGCGC
ATCGAGAGCGTCCAATCGGCCTCCTGAACCGGCTAATCATCAGCTGGGCG
CGCGCCAGGGCCACCCGAAAGAGGGTGTGGCAGACGCGCGAGGGAGAGCT
CGCCGACGATCGGCGCTCAGTCTCCCGACGAAAGGTCGCGCAGCGATGCT
ACTGCTTGGGCTTGATCACAACCTTGGGCTTGGCCTGCTTCGGCTTGATC
ACGACCTTGGGCTTGGCCTGCGGCTTATCTTGCTTCGGCTTGATCACCAC
CTTGGGCTTGGCCTGCTTGGGCTTACCACCACCTTGGGCTTGGCCTGCT
TGGGCTTACCACCACCTTGGGCTTGTACCTGCAGGCTTCACGACCACA
GGGGCTTGGTTGGGCTTGGGATCACCTTGGGCGCCGCGCCAGGCTTGG
CATGACTTGGGGCTTGGCCATCCCTGCTTTTGGAGGGGCCATGTTGGGCT
TCAGCAAGACCTTGCCCGTCGTGGGCTCAGCTGGCGCCTTGGCGGGCGCC
GGCGCCGCGGCTTCGGCGCCTCCTTCTTCGACGACCTTGATGTGAT
CGAGGCCGTGATCGCGGGCTCATAGGGGCGGTGTGAACCGTCCGCGAAGA
GCAAGCGCAGGGTGTGAGCGCCTGGGCTCAGGGTCACGTCCGCGCAAGAG
GAGCCGTCTCCAAGGTGGATGTTCTGAGCGTCCTTCGGGATATCTTTGCT
CAGGTCATCGGGGACCGGAATGTGACGAGGATGTGGTGGTGACCCTTAC
CGGGGTTACCGGGCCAGACGCCTCAAGCTCCAACCCTCCGAGCCGATG
CAGACCGTCACAGGGCTCGTGACCTCGTACCGTTTTTGAGGTTGTCGAA
GAAGACCTTGGCCCCCTTGGGCGCTGGCAGCGAGGTCGGCGCAGCCGGCT
CCTCTTGAGCCGCTTCCACCTTAGCGGCGGGCGCTTGGAGCCGCGGGCTCA
GGGGCCTTATTCTCACAAGCGACGAGGCTGAGAGCGGCAGCGAGGGCGAT
GACACGGAGGGTGGCTTTCATGGGGGACTCCTTACAATGGAGCCCGATGA

TTAATCCGACCCGAGCGCGACGTCAACGCATCCCTGGCCGGGAGGGCCTT
GTCTATGGCCTAGAAGGTCTCGGTCTTGGGTCCCCGTCGACTCCTCGCC
CGCGTGCACGCCACCCCGACCGACGCCTTGTGAGCAGGCCGCGTGCCTTT
GGGGTTGTGATTCCCCGCTGGCGCTGTCATGAAACACCTAAGACCGACCC
AAACGTGAGCCATGACACGCGATCCAGACAACGCCCTCGACGCGCTCCGC
GAGCGCTGCCCCGCGCTGGAGTTATCCACCCAGCCGGACGATCTCACGCA
CTATGGCCTCGATTGGACGCGCTTCTACACGCCCAATCCGCTCGCTATCG
CGTTCCCTCGGAGCGTCGAAGAGGTGCAAGCGCTGGTCGGGTGGGCCAGT
GAGTCCATGATCGGCCTCGTGCCTCCGGGGGGCGCACGGGGCTCAGCGG
CGGCGCTGCGCGCTCGCCGGCGAGGTTGTGGTCTCCTTTGCGCGGATGA
ACAAGATCCTCTCCTTTGACCCCGTCGACCGTCTGGTCTGCTGCCAGCCC
GGGGTCATCACCGAGCAGCTGCAGGCCTACGCGCGCGAAGAAGGCCTCTT
CTACCCCGTTGACTTCGCGGGCGGGGGTCTGCTCAGATCGGCGGGAACA
TCGCGACCAACGCTGGGGGGATCAACGTCATCCGCTATGGCATGACCCGT
CAGTGGGTTCGCGGGGCTCAAGGTCGTCACCGGGCAAGGTGAGCTTCTGGA
CCTGAACCGAGGCCTGCTGAAAAACAACGCGGGCTATGACCTTCGGCACC
TCTTCATCGGCTCAGAGGGCACGCTCGGCTTCATCGTCGAAGCGACGCTC
TCGTTGACGCGCGCGCCAGCTCAGAGCGCTGTCTTGGTGCTCGGGGCCCC
AAGCATGGGCTCCGTCCTCAACATCTTAGACGTCTTTCAAACCCGCGTGG
TCTTAAGCGCCTTTGAATTCTTTTCTCACGAGGCGCTGACGAAGGTCATC
GAGCATGCCAGGTGCCGCGCCCGATGGCTGGAGACCACCCTTATTACGC
GCTCTTGGAGCTCGAGTGCGTACCGAGGAGGACGAAGCGGTGGCCCTCG
AGCTCTTCGAGGCCTGCGTTCGAGAGCGGTGACGTCTCCGATGGCGTCATG
AGCCAAGACAACGCGCAAGCCAAACGGCTCTGGAGGTGTCGCGAGGAGAT
AAGCGAGACCATCTCGCGCTGGACGCCCTACAAAAACGACATCAGCGTGA
AGGTCTCAAAGCTCCAGCCTTTCTCGCCGCCATCGAGGAGATCGTCAA
CGGGAGTACCCAAGCTTCGAGATCATCTGGTTTGGGCACATCGGCGATGG
CAACCTGCACCTCAACATCCTCAAGCCAGAGGAGATGGAGGCGAGCGAGT
TCAAAGCGCGATGCGAGCGCATCAGCGATCAGGTCTACGGGGCCGTCGCG
CGCTTCGAGGGATCCATCTCGGCGGAGCACGGGGTTCGGGCTGCTTAAGCG
CTCGCACCTGACCTTCAGCCGAGGCGCAGAGGAAGTCCGCCTCATGCAGG
GCATCAAAGCGGTGTTTCGATCCATCGGGGATCATGAACCCTGGGAAGATA
TTCGAGGCGGAGACCTGAGGTCGAGGCTTCCCTCAGGCGCCTCATCATCG
GGGTTCGAGGAGATGGGGACATAGACGAGCGCACACAAGCGCTCAGGGCAG
GGAGTAGACCCAAGAGCTGAGGTGTTCTAGCGTCCCTACGGTTCGGTCACG
TGACCAGGAGCTTCGGATGCTGTGGTTAACTATGGAGGGCCCCGTCGACG
CGGTTGTCATCGGCGCGCGTGGCGGCATCGGTGAGGCGTTCACCGAAGCG
CTCGCTCAGGCAGAGTCTGTCCGACGCTCTACGCGACGAGTCGAGATGC
GCGCTGGGCACAGGGGTCTCATCGCACCCCAAAGTGGAGCGCGCACTCC
TGGACATCAGCGATGAGACGAGCCTGAGCGCCTTTGCCGACGGGCTCGCG
CGCGTAGAGCGCGCGCCCCGGTTCGTCCTGAACTGCAGTGGATTACTCCA
CGATGGTGAGCTCCGCCAGAGCGCACCTGGCGAGAGCTCAACGCCGAGC
ACATGGCCCCGCTCCTTCGCCGTCAACGCCATCGGCCCGGCGCTCGCCATC
AAACACCTCGTCAAAGCGGTCCCAGGAGGCAGGACGCGCTCTTTATCAC

CCTCTCGGCCAGGGTCGGGAGCATCACCGACAACCGCCTGGGCGGCTGGT
ACAGCTATCGGGCCTCCAAGGCCGCTCAAAACATGATCGTGAAGACGGCG
AGCATTGAGGCGGCGAACCGATAACCCCAAGCTCATCATCTGCGCGATGCA
CCCTGGCACCGTGGCGACCGAGCTCAGCGAGCCCTTACCAAGCGCCTCG
CGCGCGGTACGTTCGTCTTCTCGCCCGAGGAGAGCTGTTCGACATCTCTGC
GAGGTGATAGGTGGCCTACTCCCTCGGACTCCGGGGCGCTCTTCGCCTG
GGATGGCGCTCGGATCCCGTGGTGGAGCGCATGACCCCGGAAGCGCTCTCC
TGGGCCCTGGTGTCTCGCGCACCTCAGCGCCACCCTCTTCATGACGGGGCT
GATCTGGTTTCGTCCAGGTTCGTCCATTACCCGCTCAAGTCATCGGTTCGGGG
ACGCCACCTTTCGCGACTATCAGGCGCGCCACGTACCCCGGACAGGGTGG
GTGGTCGGTCCCCCGATGTTGATTGAGGCGATGAGCGCCGCGCTGCTCGC
CCTTGCCCCGCCCAATGACGCCGCGGCCCTGCCCGCGATCATCGGGCTTA
TCACCTTGGTTCGTTCGTCTGGGCAGCGACGGCGCTCTACTCTGTCCCGGCT
CACGGGCGCCTCGCCGCCGGCTTCGACCAAGTGACCCACCGGCGTTTGGT
GCGCTCCAACCTGGGTGAGGACATGGGGCTGGAGCCTGCGGAGCGGCGTGG
CCATCTGGATCCTCATGACGATGCAAGGCGCGCTGAACACCTTGACGGGC
GCTTGACTGCGCCTGTTTCATCGCGCACAATCCTGACCTCGACGGACGCAT
CGTCGTGTTCTCATCCTGGAGGTAAGCCCGTGTTCCTTCTCACGCCGCCG
ATGACCATCCCGCCCGCGTCTGAATGCCTTCCCGGCCGCGACCAAGCGGT
GGCGGTGGCACCCAAGAACATCGTCAACCCAAAACCCCATGGTGGGGCCCT
TCGGCGACCAGCTCGAGGTGATCCTGGTTCGGAATGGGCTGCTTTTGGGGC
GCTGAGCGCCTCTTTTACAGCCTCCCTGGCGTGTGGACGACTGCGGTCGG
CTACGCCGACGGCCACACCCCGAACCCACCTATGAAGAGGTGTGCACGG
GGCTCACGGGACACAACGAGGTGGTCCCTCATCGTGTTCGACCCCTCGGTC
TTGAGCAACGAGGCGCTGATCGCGCACTTTTGGGAGAACCACGATCCAAC
CCAAGGCATGCGTCAAGGCATGGACCGAGGGACGCAATATCGCTCCGGGC
TCTATGTCACCACGCCGTCACAGCGTGCAGAGGCCCTCGAGAGTAAGGCG
CGCTACGAAGCGCGCCTGGCGAGCGCAGGTTTCGGTCCAATCACCACCGA
GATCCTAGACGCGCCGACCCTCTACTACGCCGAGGATTACCATCAGCAAT
ACCTGCACAAAAACCCTTCAGGGTACTGTGGCATCGGCCGGCACCGGCATC
TCTTGTCCGATCGGCTTGAACCCCGAAGGTTAAGCCTCCCCGGCAGCGCT
CCTTCGAGGAGCTCCCCCTCTCGGTCCGCTCCATCAACGCACGAAAGCG
CGCCACAGACCGCACCCGATCGGCGCTCAATCGCAAGGTCTGGGCCCCCG
AGCGCCAACCGCGGACGCCGCGGTGCGCGCGCTTGAAGAGCGCATCCATG
ATGGGCGTATTCGACGGACGAGAGACCCATGTGGGAGGAACACGTTTCGCG
ATGAATGACTTTTGAACACCCAAAGTCATTGCAAAACATGTTCAACAGAT
CGTATCCTAGCGCCTGTAAGACGTCAGCAACCGCTGGGTCTGCGGAGCGG
CCACGCACGCTTACGCCAGCGAGACGTCCGAGGAGTTCGCATGGTCCCGA
AACAGTCTGCCTCGCTCCGACTCGTTCGCGGAACCGACGTCGCAGAGCATC
CTCAGGCGCCGTGAAGAGCGCACTCAGGCCACGCCCTCGAGCGTGGCCT
CGAGGCGGTGAGGGACCACGCGCTTCAGACCGCGCACGCGCACCTCAATG
AGCCCTTGCCTCGATGGTTCGTCCAGCACCTTCAAACGGGGCGGCAAGCAG
GTCAGGGCACGCATCGTGTCTCGAGGCGGCCCGCGCGCTCGGAGAGGAGCC
GATGACTCTGGTGCCCATCGCGGCGGCCTGTGAGCTCCTCCATAACGCCA

CGCTGGTTCACGACGACCTGCAGGATGGCGACACCGTGCGACGCGGTGAA
CCTGCCGTATGGGTGCGCCATGGTGAGGCGCAAGCCATCAACGTCGGCGA
CGCCATGTTGATGCTCCCGACCCTCGCGCTCGAGCAGATGCTGCTCATGC
CCTCTGTGCGTTGGTACATCGCCACGGCGTTCACGCGTCGCGCGACGGAG
ACCGCCTGCGGTCAGGCGCTTGAGCTCGAGCTCTTGGGTTCGAGGATGGCT
GGGCCGTGGCCCTTAACCTGAGGGCGGCCCTCGGGAAGTCCGGGCCCTTC
TTCTCTCTGCCGATCGAGACCGCGGCGCTCGCCGCGGGGCGGCAACCGAC
GCGAGCGCGCGCCATTGGCGACGCGCTCTTGGCGCTTGGGGCGCTCTTTC
AGATCCGAGATGACGTGCTCGACCTCTTCGGGGACAAGGGGCGCGGGCAA
CGCGGCAACGACCTGCGTGAAGGCAAGGTGAGCGCGCTCACCGTGA CTCA
CCTGGAGCGGCGCCCTGAGGACACCGAAGAGCTGGTCGCGCTCTTGAAAC
GCGACCGCGACGAGACCCATGACGACGACGTGGCCGAGTGGTCTGAGCGC
CTGGTTCTCAGTGGCGCGCTCGAGGACACGTGCGCGTGGGCAGATGAGCT
CCTGCGGGACCTCGTCGCGTCCGCCGAGCTCGCGATGGAGCCGGGCCTCC
GGGCCCTCGTGAGCGCGACGGCGCAAACCATCGCGGCGCCTCTGACAGGG
TTGAGGAGTTAATAATGATGACAGATATCGAGGCAAATGAAGGCAAAGAG
GCCATCGTGATTGGCGCCGGCTTTGGCGGGCTCGCGATGGCAATGCGGCT
TCAGAGCATGGGCTTTCAGACGACGATCATCGAGGCGCTCGACAAGCCGG
GCGGCCGAGCCTACGTCCGTGAAGCAGACGGATAACATCTTCGACATGGGC
CCGACCGTCATCACCGTCCCCCATTTTATCGAGGAACTCTTCGCGCTTGA
GCGCGGGGTCCACGGCCTCGACCAGCCCGACTTTCCCCGGAGACCCTAG
GACCCGATGCCCGCGTGACCTCTGGGGTGAGCGGTGGTCCGCGGACCAGC
GAGTACTGCGAGCTCCGTCCGGTGAGCCCCTTCTATCGCATCTACTTTGA
TGACGGGACCTACTTCGACTACGACGGCGATCGGGACAACACGCGGGCGAC
AGATCGCGGAGATCTGCGAAGAAGACCTTGAGGGGTACGAGCGCTTCCAT
AAAGCCGCTGGCGCCATCTTTGAGCGCGGGTTCTTGGAGCTCGGCTATAC
CCGGTTCTTAGACATCGGTTTCGATGCTCAAGGTGCTGCCAGACCTGCTCA
AGCTCGACGCAGTCAGGAATCTCTTCCGCTTCTCCAGCCGCTACTTTAAG
AGCGACAAGCTGAGGCAGGTCTTCTCCTTCGAGACCCTGCTCGTGGGCGG
CAACCCCATGAGCGTGCCCGCGATCTACGCCATGATCCACTTCGTTGAGA
AGACCTGGGGGGTACACTACTGCATGGGCGGCACCGGGGCGCTCGTCTCC
GGGTTCGTTTCGAAAGTTTGAGGAGATGGGGGGCACCGTGCTCTTGGACGC
GCCGGTTCGATGAGATCCTCGTACCCACACGGGACGGCGGGGAGCCCGGGC
GTCTGACGAGCAAAGTCNCGCGGGGCGTCAGGNTCGAGGATGGNCAAGAG
TTCTATGCCGACGTGGNGGTCTCCAACGCCGACTACGCGACGACGA

>13_D19_reverse

GTCNTACCTTAACCGACCCCGCTATGGCGCCCGCTGGGCACCACGCGGC
CTACACCTTGGTCCCTGTCCCCACAACGGGAGTGGGCTCGACTGGGAGG
CCCTCGGAGAGGGGTTTCGTTGACCGCGTCTCACGATGCTCGATGAGGAG
GGCTACATCCCAACTTGAAGGAGCGCCTGACGCACCTCAGCTGGGTAC
CCCGGATTATTTCCAGAACACCCTCCGGAGCCACCTTGGCAACGGCTTCG
GGCTCGAACCTGTCTTGACGCAGTCCGCTTGGTTCCGTCCGCACAACCGG
GTCAAGTCGATCGAGGGGCTCTACCTCGTCGGGGCGAACACGCAACCCGG

CGCTGGGACCCCGTCGGTGATGATGAGCGCCAAGATGACCGCGCGGTGCG
TCGCCGAGGACTACCCCGAGGTCCAGGTCCAGCGTCCAGGACGAGAGCCC
GGCGAGCGCCACGAGGCCACCAAGAAGGGCGCGCTCCGAAAGGCGAGCTG
AGATGACGAGCGCGCTGGCAACCCGGACGATGGCTGCAGAGAGCCGTCAG
GTCCTCGCCTCGCACGCACGGACCTTTAACCTCGCGAGCGTCTTCCTGCC
GCGCGACCAACGGGACGACGCGGGCGATCGTCTACGCCTTCTGTGCGACCG
TCGATGACTTGGTCGACGAGGCGCGGCACCCCGAGGAGGGGCGAGCGCGAC
ATCGGGCACCTGGCTGAGATGCTGCGCGGTGAGAGGCCCGCCAGCCCGCT
CGCGGAGGGCTTCCTTGAGGTCGCCGAGCGCCGGGGCTTCCACTTCAGG
CCGCCATGACTTGATGCAGGGGGTCTGAGCGACGTGGGGGAAGTGACC
CTTGTGACGACCGAGATCTCATCCAATACGGTTATGCGGTTCGTGGGAC
CGTGGGGCTCATGATGTGTGGGGTCACTCGGGGTGAGAGAGCCATGGGCGC
TGCGCACGCCATCGACTTGGGGGTGCGGATGCAGATCACCAACATCTCA
CGGGATGTGGTCGAGGACGCGGGCCGAGGGCGCATCTACATCCAACTCG
GCGCCTGGTGGCGGGCGGGGATAGACCCTGAGCAAGTCCGCGTCGGAGATA
TCGACGGGGAGGCCTTCGCGCCCGTCTCAGGGACGTCTTGAGCTCGCC
GAACGCTACTATGCGAGCGGTGATGTGAGGGCATGCGCGCCATCCCGTGG
CGCCCTCGGCTCGCCATCCAGTCGGCGAGCAGGGTCTATCGGGCCATCGG
CCTTGTCTCCGCGACCGAGCCTATGATGTCACCCAGGGGCGCGCTGTTG
TCCCAGGCGCCGCGAAGGTCTCTTGGGTCTCCGCGCGGTGGTTCGAGCG
AGCCTCTGCGGGCTGAGCGCACGCAAAGCGCATGACCCTGCGCTCCACAC
TGCGCTGCGAGGCCCGCCCTGGCACCGACGCGCTCGCACCCCGATGGGGA
CGAGGGTGAAGGCGCTCGCTCGATGACGGAGCGCTACGACTTGCTTATCG
CCGGAGCGGGCTGCGCGGGCTTAAGCCTCGCGTGGCACCTGCTCGACGCC
GGCCTTGAAGGCCAGCGCGTTCGCGGTTCGTGGATCGGAGCTTTGAGGCCCA
AGCGGACCGCCTCTGGTGCTTCTGGGGCCCCGATGACGCTCCCTTCAGCG
CCTTGGCCGATCACGCCTGGCAAAAAGGCGCGCGTTCGCTTCGACACAGAG
GTGCGTGATGAGCCGCTCGAAGACACACGCTATCATTGCGTCCGCAGCGA
ACACTTCAAGAGGGAGGTGCTCAGGCGCATCGAGGAGAGCGATAACCATCA
CCTTGATCGAGGGGGACATCCTCGCCATCGGCGAAGACGACAATGGCGCC
TATGTTGAGACCCAGAGAGAGAGCTTCGAGCCGCATGGGTTCGCGCANAG
CGTCCGCTACGGGGCCGACGACGAAGCCCAACCTCTCAAGTACCCGNCA
AGCAGCACTTCGGAGGCGTAGAGATCGTGACGGANGANCCCGTCTTCGAC
CCANAGCGCTNCACCATGATGGATTTTAGCGTTCCTCCAAACGAGGCGGGGG
TCGCCTTCGNCTACGTGCTCCCCTTCNCGCCGAACAAAGCGCTTATCGAG
CATAAGATGTTNTCAACCGACCCGCTGCCGCTGAGGTCCACAAAGAGGC
CGCGCTCGCCTATATCGCGCGGAATACGAGACTGAGTACGCCATCGAGC
GCGAGGAGTATGGGTTCGCTGCCGATGGACGATCGCTACCCTGCCAACGA
AGCAGCCCACACGTCTTCAATATCGGGATCGTTCGGTGGGCAGATGAAGCC
CAGCACCGGCTACGCATTTGAGCGTATCCAGCGCCACACGCGCGCCCTCG
CCGAGGGCTTTGTGCGAGAGGGGTGACTCTCTCAGGTCCAGAGGCAGCG
ACGCGCTTTGCTTCTATGACCTGCTGCTCCTCAGGATCCTCCACGAGGA
GCCCCGAGCGAGGCGCGAGCATCTTCCAGAGGCTCTTTGGCATAAAAGGCA
TCGCGCCGATCTTGACCTTCTGGATGAGAGGACCCGCCTCATCGATGAG

ATGAGCCTCTTTCTGCGTCTCCCATGGGGACCCTTTATCGGAGGGGTGCC
AGCGGTCCTCGGACGCCTCTCGGAGGGCGAAGGCCCTCAGGCGCCCACTC
GGAGTACCCCTGGGCTCTTGACGGCCGCGGCGGTCTCAGCGCTTGGGCG
CTCGCCATGTTCTGGGGCTCCGCGGTGAACACCCCGGAGCAGGCCATAT
CATCGCCGATGGGCTCTGGATCGCGCTAACGACCTTCTCTCCACCGGCG
TCTTCATCACAGCGCATGACGCCATGCATGGCCTGATCGCTCCGGGCAAC
GCCGCTCTCAACCGCGCGATAGGTTGGTTCGCGACGTGGAGCTACGCGGG
CCTCAACTACGCCGCGCTCGAGCGCGCTCATCACGAGCACCATGCCAAAC
CAGCGAGCGGCGACGATCCCGACTTCCACCGTGGCGACCCAGGGTTCGTC
GCGTGGTACATCAGCTTCATGCGCCAGTACCTCAGCTGGGCTCAGTTTGC
ATGGCTCTGTCTCCTCTTAGCCGTGATGATCGGCGTCCTTGAGATGTCCG
TCGCCC GCGTCGCGTTGTTTTGGGCGACACCGCTCATCCTGAGCACCTT
CAGCTCTTCATCGTTGGGACATGGTTGCCGCACCGCCCAGGAGCCTACTT
GGGGGACGGGCCGCTGCGGGCGCGAAGCGTCGACCTGCCTCCTTGGCTCT
CCTTCTCGCGTGCTATCACTTCGGCTATCACTTCGAGCACCCACGCGCGC
CCCGATCTGCCCTGGTGGCGACTCTGGCGCGTCAGGGGCTATCGCGGCCT
CAACGCCTCATAAGAGCGGCGCCACGCACGACCGATTGGACCCATGATAA
GCCTGAGCCTCTGCGCCATCTCGCTCCTGAGCCTGATGGTCTATCTCGTC
GTCTTCTACCGAGCGCGCGGGCGTGCCCGAGCTCATAGAGACCACAGTCAC
GCCCTGCGCGACTCGACCTGGCCACGGTGTGCTCATTGTCCCTGCATG
CAATGAAGCGTCGACCATCGAAGCCGCGCTCAGAAGCATCCTCGCTTGGC
CGTACCCCGCGCTCGAGGTCTTGGTTCGTCAACGACCGCTCCACCGACGCC
ACCGGGGCGATGATAGACGACGTGGCGAGAGGGCGACGCGCGCGTCGTCGT
CCACCACATCGAGGCGCTGCCTGAGGGGTGGCTCGGGAAGGTTACGCGA
TGCACATCGCGAGCAGGGAGGGCGAAGGGCGACTACGTCATCTACGCCGAC
GCCGATGTGCACCTCGCCCCGGGGGCACTTCAAGCCGCCGTGCGCTGGGC
CGAGCGCGAGCGCATCGACATGATGGCGCTCATCCAAAGATGCACGCTG
AGGGCGTCCTGCTGAATGCGACGCTTGGGGGCTTCGGGGGTTTCTTCTC
GCGGGCACCAACGCGCGGGCGGTCAACCGAGACTCTGAGGAGGCCTACGC
GGGTGTGCGGGCCCTCAGCATGGTCAGGCGCTCTGTCTTTGAGGCGAGCC
CAGGGTGGGAGTGGCTGAGGCTGGAGATCGCCGACGATCTTGGCCTGGCC
CTCATGATGCACCGCGCCGGGGCACGGGGACGGATCGCGCTTGGCGCCGCG
GCTCTTAAGCCTCCGCTGGTATAACGAGCCTCGGGGGTTTACATACGCGGCC
TTGAGAAGAACATGTTTCCGGTTCGTGCGCGCTTAGCCCCCTCAGGGCG
TGCGTCATCACCGTGGGCGCCTTGATGATGCCGCTGAGCCCTCTGCTCCT
CCTCTTCGGGCCCTACGCCTGGCTCGGCGGGCTGACCCTGGGGCTCAATG
GGCTACTCTGCTTTTTCGCGAACCCCTGCGGGCGCCCGACCCTTAGCCTAC
CTCTTGAGCCCCTTGGCGCCCTGCGTGATGGCGCTGATCTTGATGCGCTC
GACGTTCAAGGCCCTCATCCATGGCACGGTCTCCTGGCGCGGCACGCGCT
ACGATCTAGCCTCGCTCAAGACGAATCAGAGGGTACACATCATCTAAAGA
AGGCCACATGGCATACCCGCAGAGGACCGCACGATATGACCCGCACGAG
CGCTCCGAGGATGACGACGGAGATCCACCGCCGAGGGTTCGCACGGCGGGC
CACTGCGTGCTACGCATGAAGCCTCCCCTTGGCGCTCATCCGAGGGTGGTC
AAAGACGGGAAGAGCGCAGGGACCAGCGGGTGTAAAGGAGAGCTGAGGGG

TCGCCCCTCAGGGTTCACAAGGGGCGAGACGGCGATGACGCGACGCGTAGA
CAAGCGATGACACAAGACGCCACCAAAGGGGGTGACGTGATGGTCGCCGT
CGCGGGAGCATCGGGATTCATTGGACAAGCCCTGGGACCCAACCTCTCCG
AGAGCTTCGCGCCATCTCGCTCTCGCGCTCGCTCAGAGAGCCCGGGGA
GGCTACCGGAGCTGCCGTCAGGTCGACCTCTTTTCGCTCTCAGGCGCGAC
AGAGGCGCTCCGGGGCGCGCGCTTCGCCGTCTACCTGGTTCACAGCATGA
TGCCCGCCGCGAGGCTCGTTCAAGGCCACTTCGAGGACCTCGATGTCCTC
TGC GCGGACA ACTTCGCGCGGGCGGCCGCCGAGGTCGGTATCGAGCACAT
CGTCTATGTCGGCGGCCTCCTACCGAGCGGGAACGCGTTGAGCGAACACC
TTCGGAGTCGTCAGGAGATGGAGCGCGCGCTCGGGGCGACCGGCATCCCG
GTCACGACGCTGCGCTGTGGCCTGGTCGTCGGCGCCCGAGGCTCCTCCTA
TCAGCTGCTCGCGCGGCTGGTGAGGCGCCTGCCGCTGATGGTCTGCCAA
GCTGGACCTCAAGCCGCATGCAGCCCGTTGCGCTCGGAGACGTGGTCGCG
GCGATCACACAGGTCACCGACGATGTCCCCGATAAGAGCCGCGTCTTTGA
TCTCGGGGCGAGCGGACGCCGTGAGCTACAGGGAGCTGATGGCGGCGACGG
CGCGTTCCCTTCGGACTGACGCGGCGCTTCATCTCGGTCCCTGTCCTCTCC
CCGGGACTCTCTCGGCTCTGGGTACGCTCACGACCGGGCGCGCCCAAGGC
GCTGGTGGCGCCGCTTTATCCAGAGCCTCCGCCACGACATGCTCGCGCGC
CCAGACGCCAAGCTGCCGACAGAGCGCCCGACCACCTCCCTCGTCGAGAT
GTTGCGCGAGGCCGCGAGACGAGCAAGGTGATGTCACGATCGTCCCACGCG
CCTTCCGTCGCGCCACGCGCCCCAAGGGGCCCTCGACGGTCTGCAGTGTG
CAGCGCATGACGACGCCCCGGGGCGAGACATCCGCTGGGTCCCGACGA
GTA CTTCGCTGGCTCGCGCGTAAAATGCGAGGGCTCATCCACATCACGC
GCTCGGAAGACGGCCAAGAGATCCAGTTCGTCTTTGCGCTCACAGGTCAA
CCCCTCTTGGGGCTACGCCGCTTGGCGCACCGAAGCGGCCCTGACCGCGA
GGTGATGCTCGTGACCGAGGGGCTCCTCGCACGCACGACCGAGCGCGGGC
GTCTCGAATTCCGAAACGTTCTCGGAGACCGAACGCTCATCGCGGCGCTC
CACGATTTTGTACCGCGCTTGCCATGGTGGATTTACCGCGTCACCCAAGG
GACCTTTCACCGCTGGGTGATGGTCCGCTTCGGAGCTCACTTGAGGCGTC
TCAGCCGAGACGAGACGCCAAAACGGAGACATCATGACCTCACCCATCC
CCCAGAGCCGCGTCCAGACCATCCAACGGGCCCCCATACGACCCGAGAGG
CGCTTTGTCCTCTACTGGATGGTCTCGGCGCGCCGACGACCGATAATTT
CGGGCTCCAGCATGCCCTCTATCACGCGGAGACCCTCGGCGTCCCCCTCG
TGGTCTTTGAGCCCTGCGTGTGGGCTATCGATGGGCGTCAGAGCGGTTT
CATGCCCTTCGTCCTCGAGGGTATGCGGGACAACGCGGAGGCCTTCGACGG
GGTGGCTGGTGTACCTATCTCCCCTACGTCGAGCCGGAGGCGGGGGCGG
GGAGCGGGCTCCTCGAGGCCCTCGCCGCGCAGGCCTGCATGGTCTGTCAGC
GACCACTTCCCCTGCTTCTTTGTCCCTCGAATGGTTCGACAAGGCGGCCAA
GCGCCTGGACGTGCAGGTGGATGTGGTCGACTCCAATGGGATCTTGCCGC
TCTTCGCGGCCCCGAGGGAGTTCACCACCGCCACTCCTTCCGACGCCAC
GTCCATAAAAACGTGCTGGAGCACCTCGCGGAGATGCCGGAGGCCGAACC
GCTCGCGTCTGCCGAGCTGCCCCCTGCGCCGGCGCGTGACGCGCTCAATA
TCGCCCCATGGGCAGCCTGCACGGGCGAAGGCGCCCTCAACGTGACGTTA
AGCGCGCTCCCGATCGACCACAGCGTGACGCCCGTCAGCACACGTGGAGG

GAGGCAAGAAGCCTGGCGCCGACTCGACGCCTTCTTGGGGCGACAGCTCA
ACGCCTATGGCGAGCGGAACCAAGTCGAGCGTGACATTGGGACTGGCCTC
TCACCCACCTCCACTTCGGGCACATCGGGCGACACACGATCCTCCGGGA
GCTCATCGAGCGCGAGCGTTGGAGCCCTGACCAGGTGGCTCCCAAGCCGA
CGGGGAGCCGCGCGGGTTGGTGGGGGATGAGCGAGGGCGCTGAGAGCTTC
CTCGATCAAGTGCTGACGTGGCGGGAGCTCGGCTACACCTTCTGTACAGT
CAGGTGGCGCGACTATGATCGCTTCGAGAGCCTCCCCCGTGGGCGCTCA
CGACCCTCGAGGCCACGCCGAGACCGTCGCCCTGTCCTTACGATTTC
GAGGCGCTTGACGCCGGCAAGACGCACGACATGCTCTGGAACGCGGGCGCA
ACGACAGCTCCGCCGTGAGGGGATGATCCAACTACCTGAGGATGCTCT
GGCAAAGAAGATCCTTGAGTGGACCGAGCACCTCGGGTCGCGCTCGAC
ACCCTTATCGAGCTCAACAACCGCTACGCCCTCGATGGCAGAGATCCCAA
CTCGTACTCCGGCATCTTCTGGTGCCTGGGTGCTTCGACCGCGCATGGG
GCCAGAGCGTCCGATCTTCGGCAAGATCCGTTACATGTCGAGCGAAAAC
ACCGCCCGAAAGGTTTCGCGTCGGCGCGTACATCGCGCGCTACAGCTGAAC
AGCCTCAGTAAAGCGGTCCGAGGACACACCTCGGGTCGGGGAAAGCGCTG
AGAGCCACGGCATATCCGCGTCACCGACCGCCCCGGGAGATCGTGGGT
CTCTGGGACGACCAAGCTCAGCTGAAATGAAGGTGCGGGTTCAGCCCC
CGTTGACCTCGCGGACCCACCTCGGCGAGGCGCTCTCCAGCCTCAAAG
CGCTCCCCTTCAGACCAGCGAGAGGGCTCCCTCGGCTCAGGTGCCATG
AAGCGCGAAGACCTCTTGGTCAAGCCACCTGTGGCGCGTGATGATCGTCG
GGCCGTAGTCGTAGGGCAGGGCGTTGTCTCCAGCTTGAAGATGCTCCCC
GCGTAAAAGGCATACACCGCCTCGCCTGCGGGCGCGGCGATGTCGATGCC
GACGTGGATGTCTCGACGGCCTTCGCTGAACTGCTACCGAGGTACATCC
CAGGGCGCCGCTCGTTGTAGCGCCCGATGCCATAGGGGCTCGTGAGCGTT
CGTTCAGGGTCATAACCCTCGCTGAAATCGTAGACCTCATAGGTCTGGGG
CAAGCGGATCACGGGCTTGAAGGTAACCCCTCGACGCGCAGGGGGGCGT
AGCTATCGCGCTTTAGGGTGTGCTTCATCGAGGCAACGTGCCATGCGCGC
CATCCACGCGGCAAGCGTGGTAGGCTGCCGATGGACCCCGCGTGCGCCAG
GAGTTAACCGTGATACGTCGCGCCTTGATGCTCAGCTCCCTCGTCCTCGT
GACCGGGTTCGAGAAACCTGCACCGCCTGGAGAGCAGACCCAGGCAGATG
TGCACGCTGCGGAGGTGGCGGCGTTGGTGGAGCGCGCCTATCCCGGGGGGA
ACAGTGATCTACCCACCTTGGTGATCGGGCGTTCAGCCGACGAAGGCCGC
GCTCGAGGAGCTCGCCACGCAAGGACTGAAGCACGTCATCAACCTGAGGA
CGCCAGCGAGCTCGACTTTGATGAGGAGAAGGTGTCGTCAAAGAGCTGAAG
ATGTCCTACCAAATGATCCAGTGCCAGGCGCCAAAGGCCTCACAGAGTC
GGCGATTGACGACCTCCACGCTGCCCTCAACCCTGAGGTCCCGACCCTCG
TGCACTGCGCAAGCGGCAACCGCGTGGGCGCGCTCTTCGCCCTTCGAGCC
CAGCGCCACCTCGACGCCAACTTCGACGAGGCGCTCGCGACCGGCGTCCG
GCATGGACTCAAGGGGCTCAAGGGCGCCGTCTCAGAGCGCCTCCATGCGT
CACCGAAAGACAACTTCCAGAGCGCGCCTCCGCGCGTGTGACGCGCTTC
AAGGAGGCGCTGAGCCGCGAACTCGGCGCCGCCATCAAGGCCGAGGGGCT
GGCTGGAGCCATCGATGTCTGTGCGCAGAGGGCGCCAGCGCTGAGCCGCG
AGCTCTCAACGGAGACGCTGAGCCTGCGTCGCGTCGGGACACGGGTCCGA

AACAGCGAGACGAACACCCCAAGTGAGGCGATGCGCAAGGTCCTTAAAGG
TCTCAGCAAAGGCGCGTCGGTCTACGTCGGGACCGTCGACGACACCCAAG
CAGCGGTGCATGGGCTCTTCATCGATAAGCCCCTCTGCTTGAGCTGCCAT
GGTCCACAAGACACCCTCGCCAAGCCCGTCCTTGAGGCGCTCAAGGCGCG
TTACCCTGATGACGAGGCCACGGGCTACGCGATGGGCGACCTCCGTGGCG
CCATCGTGTTGGAGCGGCGTCGCGCACCTTGAGTGGGGGCGCCGCTTATT
CGGCCTCTTGCGGCAGATAAGCCACCTCAAGGATCTCGATGGGCGTCACC
TCCCCGGCGTGGTTGAGCGTCACCTCATCCCCCTCGCTCTTCCCGATCAA
GAGGCGCCCGAGAGGGGAGCGCCAGCTGATGCGCTGCTTATCAACCTCTA
TTTCATCGACGCCGACCAGCGCCACCTGCTTCTCGGTGCCGTCCATGTAG
CCCAGGGTCACCGTCGCCCCGAAGAAGATCCGCTCGCCACGGTCAATCGC
TGATCAACGATCTCGGTCAATCGAGGCGCTTATTGAGCCAGCGCAAGC
GACGGTCGATCTCTCGAAGGCGCTTTTTCCCATAGATGTATTCGGCGTTC
TCCGAGCGATCCCCATCGCGGCGGCGCGGCCACCTCATCGACGATATG
AGGCCGCTCGACGAACCAAAGCTGGTCATGCTCGGCGCGCAGCCGATTGA
AGCCGCGTGGGGTGATATAGTTCTTAGGCGTCGCCATCGGAGCGACCCTA
ACCCAAGACGAAGAAAGGCTCACGATGTGCGGTGATTTGGCATCAGCGC
CAACGCCGAGCAGCTCGCCCTGCTCCTCGAGATAACCCCGAGTGAGGTGCG
GTGAGGTGCGCCACGCTTCAACATCGCGCCGAGCACGCTCGCGCCCATC
GTGCGCCATAACAGCGCCGAGCGCAGGGCGCTTCATTGGGTCCGATGGGG
GCTCGTCCCTTCGTGGGCCAAGACACGCAAGATTGGGCATAAGCTCATCA
ACGCCGCTCGGAGACCGTGAATATCAAGCCCGCCTTTCGAGCCGCCTTT
GAACGACGACGTTGCCTCGTCATCGCCGATGGCTTCTACGAATGGAGGCG
AGACGAGGGGGCGAAGACCAAGCAGGCCTATCATATCGGGCTCAGCGATG
AGTCCGCCTTCGCGATGGCGGGTCTCTGGGAGCGCCACACCGACCCTGTC
GCCGGCGACACCCTCGACACCTTCACGGTGTTGACCACAGAGGCCAACGA
TGTACTGGCCCCCTTACCACCGCATGCCGGTCATCCTCCCGCCCCAGG
ACTATGAGACTTGGCTGTGCCGAGAGAGCGACCCGAGAGCGCTCTTGAAC
CTCCTAAGACCCTGCCAAGCGAGATCCTCGTGACCTGGCCCGTCTCCCC
CTTGGTCAATAGCCCAAAGCATCAAGGCGCCGAGTGTCGGAGCGCCATCC
AGGTGTCGACAGACGCACCCGGAGCCAAGGCGTCATGAAGAAAGCCCACC
TCAAGAAAACGCAGCGCATCTTGGTCCGGTTTGTGGTGCTATGCGTCCTG
GGCTTCTTCGCCCTTGCAGCGCTCTGGGGGGCCTCGATCTCCGTTGGCTA
CCTCAACCTCGCCCTCTCCTATGAGGCGCTCAAGCCCCAAGCGGAGCTCG
CCATGCACTACCCCTTGACCCTGGCCAGCGCGCGCGCTCACCCCTTGAG
GGCGAGGCGCTGGTCGCCAGCCCCCTCGGGCGGATCACGATCTTGCAAGC
GAGCGAGGCGGAGCGCGTCGTGATCTTCGACCGAGGCGGAGGTCACCTCG
GCATACAGGGCTACCTCTACGCCCCACAACCGAGGACGAGGAGACCGTC
TATGCGATGGCCTTTGGGGACCTCGAGGCGCGCGAGGTAAACGCGCTCAT
CGGGGGTTGGTACAGCTATGACTCGACCGAGGATTAACCTCGGTCTGCCT
CACCCCTCGAAGGCGCGTCGACAACCTCAAGCGCCTCGTCGATGATCGC
AAAAGCGTGCCGAAGCTCCTCCTCGGTGATGCAGAGGGGGCGGGATGCACA
TGAACTCGCTCCAGCGCACGTAGGTGTAAAGCCCGAGCTCCAAGAGCCGC
GCCTTAAAGGCGTCACCGACGGCTGAGTGGTGCCCATAGGGAGCGAGCCG

TTTGCCTGCGCTGTCGCGCTGAAGGTCGAGGATCCCAAAGAGGCCAATGA
CGCGCCCCTCTTTGACAGAGGGGTGCGTTGCCCTCAAGCGTCCATCTCT
TCACGCATCACAGGCTCAAGCTGCGCTGCATTCTCGATGAGCCCCATC
ACGGAGGACGTGGATACAGGCCTCAGCGGCCGCCAGTCCGCTTGGGTGCG
AGTTGTAGGTGAGCCCGCCGTAGAAGACGTTGTCTTTAAAGTGGTCGGCG
ATCTTCTGCCGCATCCCCATCGCCCCGAGCGACTGGTAGCAGCTCGTGAG
CCCCTTCGCCATCGTCACGATGTCCGGGACGATCCCGGCGTGCTCGAAGG
CGTACATCTTGCCGGTCCGCCCCGAAGCCCGCCATGACCTCGTCGCAGACG
AGCAAGATGCCGTAGCGATCGAGGAGCGCTCGGAGGCCCTTGAGGTACCC
CTCGGGTGGCGGCAGGACGCCATTGGTGCCTGTGACGGTTTCGATGAACA
TCGCCGCAATTTGGTCCGGGCCCTCGTAGTGGATGACCTCCTCGAGGTAC
CGGAGGTTTTGCTCGGTCTTCTCGGCGTCCGGTCTCGCCAAAGCTGTAGGT
GTAAGGCCTCGGGTCCATGACCTTACGAAGCCCTGCGCGCCGGGCTCGT
TGGGGATCCTTCGGGGGTCAACCGTCAGCATGGCGGGCGGCGTGGGTGCGG
CCATGATAGCTCCGGTAGCGACTCAGGATCTTGTGGCGCCCCGTGAACCA
CCGCGCCGCTTGATGGCATTCTCGTTGGCCTCGGCGCCGCTGCAGGTGA
AAAAGAAGACGTCGATATCCCCCGGCGTGATCTCAGCCATCAGCTTGCCC
AAGCGCGCTCGGACCTCGGTGGCCGACCCCGGCATCGCGAAGGGGAGGAC
ATCAAGCTGCCTCTGGACCGCCTCGATGACCCGCGGGTGACCATGACCCA
CGTTGACGCTCATGACATGGCTATTGAAGTCGAGGATCTTCCGATTATCC
GGGCCCCAAAAGTAGATCCCCTCAGCCCCTGTGATCGGGATGGGGTCGAC
GGCGTTGCCGGCAGACCATGAATACATGGTGTGCGCCTTGACACAGACGAA
TCATCTCATTGCTGTCCATACATCCCCCTGTCAGCTCGCCGAGCAAGCC
ATGGTGTGTGCACAAAGGCAACGGCAATCACCCGTGTTTGCGATGCGGCG
TGGAGGAAACACACTCGGGCTGCATCCCATGTCATGGCGACCGCATTTC
GGTGCATCTCAGGTGGAGGAGAGCAGATCGCCTGTCCGGTGACGAGAGT
CCCTGAAGCGTCGCCTGAGCCCGAGCGCCCTGACCCCTGTTCCGGCCAAGC
GTCTCTTGAGCCCTTCGGGTGTGCTCGATAGCCGCCTCAAGGGTCAAGGG
ATCGCAGGGGGTAACAAGGACACGATGACACACGCTCACGAGCGCCTGAG
CCCCCTCAAGCGTGTCCACATCGGCGAGCACCAGCGCACACAATCGGGC
TGCTGATGGCGCAGGTCATCGAGCACCCCTGTGCGTGACCCTGATCGGA
GCCCAAGACGATCAGCGCCGCGTCAAAAGGCTCCTCAGCGATGACGAGCC
GCGCATCCTCGAGCCGCGGTGCATGCGTCAACGCCACGGTGGGCTCGCG
CCAGAGAGGTGGTCACGGCAAGCGCGGGGCAGCTCTGCGCGCTCGTCGAC
GATAAGCAAGCTGATACCCATAAGACCCCGAAGCACGTGATGAAACACC
CCGGGGCGGACATCAGACGAGGCTCTTGTCTCCTAGGGCGCCTATCGAC
TTTGGGGCGAAAGAGGCAAGTCAGCCATTTTCTCGTCTTTGGGGAGGGGT
TGCGCGGCACATCGCACCCCGCTGGGGGCTGTGCTACGCGCAACGCGT
CGCGCAGGGGATCGGGCGCTTGGACCCTCAGTTCAGGGCGTTGGGCGCCG
CCAAGGTGAAGGGCGCGATCACCGCGTCAACCCCTCGCCGCTCTTCGAG
ATCATCTGATAGGTACCCTGCATGGTACCACCGGCGTACTGAGAGGACA
GGCGCTCGTGTACTCGTAAGACTCCGAGGGCCTCAAGATCGGTTGGTCAC
CGACGACGCCAGGTCCCTTGACCTCCTCCACCTTATTCTCGCCATCGGTG
ATGACCCAATGACGGCTGACGAGCTGCACCGTCTCGTCACCCACGTTGGC

GATTCGGACACGGTAGGCGAAGAAGTAGTAGTCCTGCATGGGATCGCTGC
GCTCGGCCACGTAGAAAGACCTCACGCTTACAGCCACACCACGGGTAATC
GCTTCGCTCATGGTCTCTCGACTAGCGCGATGCGGCAGAAGGATCAAGGG
GTGGTGAGGATAGGGGCCCCCAAGGTGATCCCCTCCGGGCTCACCTGCGC
GCTGTAGACGAGGACCTCAACCCCTGCGTCCTTCGCGGCGAGGACGCCCT
CGGCGTAAGCGGGGTCGATCGACCTGGCGGGAGCGAAGCGCTGACAATCG
GCGCGCTGAACCAGGTAGAACATGATAGCGCGATGACCTTGAGCGACGAC
CCGCTCAAGCTCTCTGAGGTGCTTCAAGCCTCGCGCGGTGACCGCATCCG
GGAAGAGCGCCGTCCTGCATCAGCCAAGGTGACGCTCTTGATCTCGACG
TAGCAAGGTCGCGCGTCCTTAGGGTGCCCCTCCGCCAACAGGTCAATACG
GGAACGCTCGGCGCCATATCGGACCTCCCGCCGGAAGGAGGTGTACCCAG
TCAGCGCCTCAATGGCGCCAGCCTCAGCGGCCGCGGCGACGAGATGGTTA
GGGAGGTTGGTGTTAACCCACCAGGGCAAGCCCCCTCATGGGCGATCTG
CCAGGTCCACTGGAGCTTTCGCTTCGGGTCGTCATGAGGGGTCAACCAGA
CAAGGCTGCCCTCGGCGGTGAGCCCCCTTCATGCTACCCGGATTCATGCAG
TGGGCGGTCACCTCCCGCCCGTCGCTCAAGCGGACGTCGGCAAGGAAACG
CTTATAACGACGCAGGAGCGTCGCCTGGACGAGCGGAGGGGAGAAAAGCG
CTCGAATCTGAGTTTGAGAGACATCGACATGGAACGTAGGCATGAAAAC
CGCTATAGTCCGCCCTCAGAAGCTCTCCACGCCTGCAGAGCGCGGACATT
GGACACGCTTCGATGGATTGCAAGGAGAGCAGGATGCTGCGGTGGTCAGT
GTTAATCGGTGTCGGGTGCGCAGCGCTCATTGGGTTGACCGCCGCGGCAT
CGGCGACGATCGTCAAGCCCTTACCCTTCGAGGCCTCGCCACCGAGGCG
CATGAGGTGGTCCGCGGCGAAGTCATCGATGAAGAGGTGATCTACGACGC
CTATTGGGATTCGTGTCTACACCCACAACGTCGTGCACGTCTTAGAGGCG
CTTGGTGGGACCGCGCGCCCCGGTGACCTCATCGTCGTGCGCCAGATGGG
CGGGGTCTTGGATGGGATCGAGAGCGTCCTCATCGGGACCAGCGACCTTC
AGCTGGGTGATGAAGTGGTGCTCTTACCCGCACCGACGGAGCCATGCAC
TACCTTGTCGGGATGGCC

HF186_25m 30-B18

>full_30_B18_seq

CATCACAGAGGGCCCCATAGGCCTTATTTCCGGCCTTGAGGCAGACGCCG
TTGACGCAGGTCATCCCGGAGGGGCACCAGCTTGTGCAGGTCTGAGAGCA
GTAGATCTCATCGCTGAGCACCAGGCACACCCCGCCCTCGCAGGTCTGGT
TGCCATAGCAGAACTCTCCGTTGGGGATGTTGCCACCGGCGTCATGACAG
TAGTCATTTTCAGGGATGCACTGGCTTGGGATGTTCGTCATTGATGAACTG
ACATCGAAAGCCCTGGGCGCAGTCGCTGTCACTGGTGCAAGTTCTCACCCG
AAAAGGCCTCACCAACGCCCCAGGAGAGGCAGTAGCCGTTGGCGCAGTTC
GCGCTCTCATCGCAGGCGTCGCACGCGATGCCCTCGGTGAACGAGCCGAC
CGGGTAGAGGAAGCAGGCGCCACGCTCATCATCCTCCGCGAGGCTCCTCA
GCTCCGCTGAGCCCCCTGAGAAGAACATCGTGGAGTCCGGATAGCGCGGG

TGATCGAGGCCGATAGCGTGTCCGATCTCATGGGTCACGACGCCCTGCAC
GTCCTGGTAGCCGTATCCGGGGTTCATGCTGGTGGTCCATTGAACCTCGG
CCGCGTTGAAGACGACGTCGACACTCTGGATCACGCCTTGCCAGTTGGGG
TCGGGGAGCGCATAGGCCAAGGCATCGCCCACCGAAGGGTCCCAGTTCGT
GGTCTGTACGTAGACATGGATTCCCTGGTTGTTATCCCAGCTCTTGAGCC
CCTTGTAAGTGAAGCCGATGGTGGCGCAGTCGACGTCTTCCCAGGATTCA
AACGCGGCCTCGATCGCGTCCTGAAACTCGTCGATATTCATCCCAGGCAT
CCC GCCCGTGTGAGCGTCCAGTCGACGTTGTCCGTCCCCCAGCGCAGCG
TGTTGCCATTCTGGAGCCGGATCGTGACATAGGCTTGGGCGGAGGCCGAG
AGGAGCACCCCGAGGAGGACCAAGGTGATCAGCGTGAGCTTGC GCCTCAT
GGTGCCCCCTGTACGGGTCGGA ACTGCGCGATGGCGGCCACAGGTTGGT
GAGCAGCGTGTGGTTTGGAGCCTTACGCGCTGCGGTA TGGGGCAGCGGTC
GGATGTGATCGCGTCGAGGTAGCGCGTCACCCGCCCGATGCCAGCGCCG
TCTCGCCACACCCAGTAGGCGCCTTGGGCCATCCC GACAAGGTAGTGCAT
CGCTCCGTCCGGTGC GGGTGAAGAGCACCACTTCATCACCCAGCTGAAGGT
CGCTGGTCCCGATGAGGACGCTCTCGATCCCATCCAAGACCCCGCCCATC
TGCGCACGACGATGAGGTCACCGGGGCGCGGGTCCCACCAAGCGCCTC
CAAGACGTGCACGACGTTGTGGGTGTAGACACGATCCCAATAGGCGTCGT
AGATCACCTCTTCATCGATGACTTCGCCGCGGACCACCTCATGCGCCTCG
GTGGCGAGGCCTCGAAGGGTGAAGGGCTTGACGATCGTCGCCGATGCCCG
GGCGGTCAACCCAATGAGCGCTGCGCACCCGATAACCGATTAACACTGACC
ACCGCAGCATCCTGCTCTCCTTGCAATCTATCGAAGCGTGTCCAATGTCC
GCGCTCTGCAGGCGTGGAGAGCTTCTGAGGGCGGACTATAGCGAGTTTTC
ATGCCTACGTTCCATGTCGATGTCTCTCAA ACTCAGATTTCGAGCGCTTTT
CTCCCTCCGCTCGTCCAGGCGACGCTCCTGCGTCGTTATAAGCGTTTCC
TTGCCGACGTCCGCTTGAGCGACGGGCGGGAGGTGACCGCCC ACTGCATG
AATCCGGGTAGCATGAAGGGGCTCACCGCCGAGGGCAGCCTTGTCTGGTT
GACCCCTCATGACGACCCGAAGCGAAAGCTCCAGTGGACCTGGCAGATCG
CCCATGAGGGGGCTTGCCCGGTGGGGGTTAACACCAACCTCCCTAACCAT
CTCGTCGCCGCGGCCGCTGAGGCTGGCGCCATTGAGGCGCTGACT:GGGT
ACACCTCCTTCCGGCGGGAGGTCCGATAT:GGCGCCGAGCGTTCCCGTAT
TGACCTGTTGGCGGAGGGGCACCCTAAGGACGCGCGACCTTGCTACGTCG
AGATCAAGAGCGTCACCTTGGCTGATGCAGGGACGGCGCTCTTCCCGGAT
GCGGTCACCGCGCGAGGCTTGAAGCACCTCAGAGAGCTTGAGCGGGTTCGT
CGCTCAAGGTCATCGCGCTATCATGTTCTACCTGGTTCAGCGCGCCGATT
GTCAGCGCTTCGCTCCCGCCAGGTTCGATCGACCCCGCTTACGCCGAGGGC
GTCCTCGCCGCGAAGGACGCGAGGGGTTGAGGTCCTCGTCTACAGCGCGCA
GGTGAGCCCGGAGGGGATCACCTTGGGGGCCCTATCCTCACCAACCCTT
GATCCTTCTGCCGCATCGCGCTAGTCGAGAGACCATGAGCGAAGCGATTA
CCCGTGGTGTGGCTGTAAGCGTGAGGTCTTTCTACGTGGCCGAGCGCAGC
GATCCCATGCAGGACTACTACTTCTTCGCCTACCGTGTCCGAATCGCCAA
CGTGGGGTGACGAGAC:GGTGCAGCTCGTCAGCCGTCATTGGGTCATCAC
CGATGGCGAGAATAAGGTGGAGGAGGTCAAGGGACCTGGCGTTCGTCGGTG
ACCAACCGATCTTGAGGCCCTCGGAGTCTTACGAGTACACGAGCGCCTGT

CCTCTCAGTACGCCTGTGGGTACCATGCAGGGTACCTATCAGATGATCTC
GAAGAGCGGCGAGGGGTTTCGACGCGGTGATCGCGCCCTTACCTTGGCGG
CGCCAAACGCCCTGAACTGAGGGTCCAAGCGCCCGATCCCCTGCGCGACG
CGTTGCGCGTAGCGACAGCCCCAGCGGGGTGCGATGTGCGCCGCCAACC
CCTCCCCAAAGACGAGAAAATGGCTGACTTGCCTCTTTCGCCCCACAGTC
GATAGGCGCCCTAGGAGGACAAGAGCCTCGTCTGATGTCCGCCCCGGGGT
GTTTCATCACGTGCTTCGGGGGTCTTATGGGTATCAGCTTGCTTATCGTC
GACGAGCGCGCAGAGCTGCCCCGCGCTTGCCGTGACCACCTCTCTGACGC
GAGCCACCGTGGCGTGTGACGCATGCACCGCGGCTCGAGGATGCGCGGC
TCGTCATCGCTGAGGAGCCTTTTGACGCGGCGCTGATCGTCTTGGGCTCC
GATCAGGGTCACGCGCAGGGGGGTGCTCGATGACCTGCGCCATCAGCAGC
CCGATTGTGTGCGCTT:GGTGTCTCGCCGATGTGGACACGCTTG:AGGGGG
CTCAGGCGCTCGTGAGCGTGTGTATCGTGTCTTGTACCCCTGCGAT
CCCTTGACCCTTGAGGCGGCCATCGAGCACACCCGAAGGCTCCAAGAGAC
GCTTGGCCGAACAGGGGTGAGGGCGCTCGGGCTCAGGCGACGCTTCAGGG
ACTCTCGTACCCGACAGGCGATCTGCTCTCCTCCACCTGAGATCGCTCC
GAAAATGCGGTCGCCATGACATGGGATGCAGCCCGAGTGTGTTTCTCCA
CGCCGCATCGCAAACACAGGTGATTGCCGTTGCCTTTGTGCACACACCAT
GGCTTGCTCGGCGAGCTGACAGGGGGGATGTATGGACAGCAATGAGATGA
TTCGTCTGTGCAAGGCGCACACCATGTATTTCATGGTCTGCCGG:CAACGC
CGTCGACCCCATCCCGATCACGGGAGCCGAGGGG:ATCTACTTTTGGGGC
CCGATAATCGGAAGATCCTCGACTTCAATAGCCATGTTCATG:AGCGTC:
AACGTGGGTTCATGGTACCCGCGGGTTCATCGAGGCGGTCCAGAGGCAGCT
CGATGTCTCTCCCTTCGCGATGCCGGGGTTCGGCCACCGAGGTCCGAGCGC
GCTTGGGCAAGCTGATGGCTGAGA:TCA:CGCCGGGGGATATCGACGTCT
TCTTTTTCACCTGCAGCGGCGCCGAGGCCAACGAGAATGCCATCAAGGCG
GCGCGGTGGTTACGGGGCGTACAAGATCCTGAGTCGCTACCGGAGCTA
TCATGGCGCGACCCACGCCGCCCATGCTGACGGGTGACCCCGAAGGA
TCCCAACGAGCCCGGCGCGC:AGGGCTTTGTGAAGGTCATGGACCCGAG
GCCTTACACCTACAGCTTTGGCGAGACCGACGCCGAGAAGACCGAGCAA
ATCTCCGGTACCTCGAGGAGGTCATCCACTACGAGGGCCCGGAGCAAATC
GCGGCGATGTTTCATCGAGACCGTCAAGGCACCAATGGCGTCTTCCCGC
GCCGAGGGTTACCTCAAGGGCCTCCGAGCGCTCCTCGACCGCTACGGCA
TCTTACTCGTCTGCGATGAGGTCATGGCGGGCTTCGGGCGGACCGGCAAG
ATGTACGCCTTCGAGCACGCCGGGATCGTCCCGGACATCGTGACGATGGC
GAAGGGGCTCACGAGCTGCTACCAGTCGCTCGGGGCGATGGGGATGCGTC
AGAAGATCGCCGACCACTTCAAAGACAACGTCTTTACGGCGGGCTCACC
TACAACTCGCACCCAAGCGGGCTGGCGGCCGCTGAGGCCTGTATCCACGT
CCTCCGTGATGAGGGGCTCATCGAGAATGCTGCGCAGCTT:GAGCCCGTG
ATGCGTGAAGAGATGGAGCGCCTGAGGGCGACGCACCCCTCTGTCAAAGA
GGGGCGCGTTCATCGGCCTCTTTGGGATCCTCGACCTTCAGCGCGACAGCG
CAGGCAAACGGCTCGCTCCCTATGGGCACCACTCTGCCGTCCGGTGACGCC
TTTAAGGCGCGGCTCTTGGAGCTCGGGCTTTACACCTACGTGCGCTGGAG
CGAGTTCATGTGCATCCCGCCCTCTGCATCACCGAGGAGGAGCTTCGGC

ACGCTTTTGGGATCATCGACGAGGCGCTTGAGGTTGTCGACGCGGCCTTC
GAGGGGTGAGGCAGACCGAGGTTAATCCTCGGTCGAGTCATAGCTGTACC
AACCCCGATGAGCGCGTTTACCTCGCGCGCCTCGAGGTCCCCAAAGGCC
ATCGCATAGACGGTCTCCTCGTCCTCGGTTGTGGGGGCGTAGAGGTAGCC
CTGTATGCCGAGGTGACCTCCGCCTCGGTCGAAGATCACGACGCGCTCCG
CCTCGCTCGCCTGCAAGATCGTGATCCGCCCAGGGGGGCTGGCGACCAGC
GCCTCGCCCTCAAGGGTGACGCGCGCGCGCTGGCCAGGGGTCAAGGGGTA
GTGCATGGCGAGCTCCGCTTGGGGCTTGAGCGCCTCATAGGAGAGGGCGA
GGTTGAGGTAGCCAACGGAGATCGAGGCCCCCCAGAGCGTCGCAAGGGCG
AAGAAGCCAGGACGCATAGCACCACAAACCGGACCAAGATGCGCTGCGT
TTTCTTGAGGTGGGCTTTCTTCATGACGCCTTGGCTCCGGGTGCGTCTGT
CGACACCTGGATGGCGCTCCGACACTCGGCGCCTTGATGCTTGGGGCTAT
TGACCAAGGGGGAGACGGGCCAGGTCACGAGGATCTCGCTTGGGCAGGGT
CTTAGGAGGTTCAAGAGCGCTCTCGGGTTCGCTCTCTCGGCACAGCCAAGT
CTCATAGTCCTGGGGCGGGAGGATGACCGGCATGCGGTGGTGAAGGGGGG
CCAGCACATCGTTGGCCTCTGTGGTCAACACCGTGAAGGTGTCGAGGGTG
TCGCCGGCGACAGGGTTCGGTGTGGCGCTCCAGAGACCCGCCATCGCGAA
GGCGGACTCATCGCTGAGCCCGATATGATAGGCCTGCTTGGTCTTCGCC
CCTCGTCTCGCCTCCATTCGTAGAAGCCATCGGCGATGACGAGGCAACGT
CGTCGCTCAAAGGCGGCTCGAAAGGCGGGCTTGATATTCACGGTCTCCGA
GCGGGCGTTGATGAGCTTATGCCAATCTTGCGCGTCTTGGCCACGAAG
GGACGAGCCCCATCGGACCCAATGAAGCGCCCTGCGCTCGGCGCTGTTA
TGCGCACGATGGGCGCGAGCGTGCTCGGCGCGATGTTGAAGCGTGGGGC
CACCTCACCGACCTCACTCGGGGTTATCTCGAGGAGCAGGGCGAGCTGCT
CGGCGTTGGCGCTGATGCCAAATCGACCGCACATCGTGAGCCTTTCTCCG
TCTTGGGTTAGGGTCGCTCCGATGGCGACGCCTAAGA ACTATATCACCCC
ACGCGGCTTCAATCGGCTGCGCGCCGAGCATGACCAGCTTTGGTTCGTCG
AGCGGCCTCATATCGTTCGATGAGGTGGCGCGCGCCGCCGCGATGGGGGAT
CGCTCGGAGAACGCCGAATACATCTATGGGAAAAGCGCCTTCGAGAGAT
CGACCGTCGCTTGGCTCAATAAGCGCCTCGATATGACCG:AGATC
GTTGATCCAGCGATTGACCGTGGCGAGCGGATCTTCTTCGGGGCGACGGT
GACCCTGGGCTACATGGACGGCACCGAGAAGCAGGTGGCGCTGGTTCGGCG
TCGATGAAATAGAGGTTGATAAGCAGCGCATCAGCTGGCGCTCCCCTCTC
GGGCGCCTCTTGATCGGGAAGAGCGAGGGGGATGAGGTGACGCTCAACCA
CGCCGGGGAGGTGACGCCATCGAGATCCTTGAGGTGGCTTATCTGCCGC
AAGAGGCCGAATAAGCGGGCGCCCCACTCAAGGTGCGCGACGCCGCTCCA
CCACGATGGCGCCACGGAGGTCGCCCATCGCGTAGCCCGTGGCCTCGTCA
TCAGGGTAACGCGCCTTGAGCGCCTCAAGGACGGGCTTGGCGAGGGTGTC
TTGTGGACCATGGCAGCTCAAGCAGAGGGGCTTATCGATGAAGAGCCCAT
GCACCGCCGCTTGGGTGTCGTCGACGGTCCCGACGTAGACCGACGCGCCT
TTGCTGAGACCTTTAAGGACCTTGCGCATCGCCTCACTTGGGGTGTTCGT
CTCGCTGTTTCGGACCCGTGTCCCGACGCGACGCAGGCTCAGCGTCTCCG
TTGAGAGCTCGCGGCTCAGCGCTGGCGCCCTCTGCGCACAGACATCGATG
GCTCCAGCCAGCCCCTCGGCCTTGATGGCGGCGCCGAGTTCGCGGCTCAG

CGCCTCCTTGAAGCGCGTCACACGCGCGGAGGCGCGCTCTGGGAAGTTGT
CTTTCGGTGACGCATGGAGGCGCTCTGAGACGGCGCCCTTGAGCCCCTTG
AGTCCATGCGCGACGCCGTCGCGAGCGCCTCGCCGAAGTCAGCGTCGAG
GTGGCGCTGGGCTCGAAGGGCGAAGAGCGCGCCCACGCGGTTGCCGCTTG
CGCAGTGCACGAGGGTCGGGACCTCAGGGTTGAGGGCAGCGTGGAGGTCG
TCAATCGCCGACTCTGTGAGGCCTTTGGCGCCTGGCACTGGGATCATTG
GTAGGACATCTTCAGCTCTTTGACGACCTTCTCCTCATCAAAGTCGAGCT
CGCTGGGCGTCCCTCAGGTTGATGACGTGCTTCAGTCCTTGCGTGGCGAGC
TCCTCGAGCGCGGCCTTCGTCGGCTGACCGCCGATCACCAAGGTGGGTGA
GATCACTGTTCCCCCGGGATAGGCGCGCTACCAGCGCCGCCACCTCCG
CAGCGTGAACATCTGCCTGGGCCTGCTCTCCCGGCGGTGCAGGTTTCTCG
CACCCGGTCACGAGGACGAGGGAGCTGAGCATCAAGGCGCGACGTATCAC
GGTAACTCCTGGCGCACGCGGGGTCCATCGGCAGCTTACCACGCTTGCC
GCGTGGATGGCGCGCATGGCACGTTGCCTCGATGAAGCACACCCTAAAGC
GCGATAGCTACGCCCCCTGCGCGTCGAGGGGGTTACCTTCAAGCCCGTG
ATCCGCTTGCCCCAGACCTATGAGGTCTACGATTTAGCGAGGGGTATGA
CCCTGAACGAACGCTCACGAGCCCTATGGCATCGGGCGCTACAACGAGC
GGCGCCCTGGGATGTACCTCGGTGAGCAGTTCAGCGAAGGCCGTGAGAC
ATCCACGTCGGCATCGACATCGCCGCGCCCGCAGGCGAGGCCGTTGATGC
CTTTTACGCGGGGAGCATCTTCAAGCTGGGAGACAACGCCCTGCCCTACG
ACTACGGCCCGACGATCATCACGCGCCACAGGTGGCTTGACCAAGAGGTC
TTCGCGCTTCATGGGCACCTGAGCCGAGGGAGCCTCTCTCGCTGGTCTGA
AAGGGGAGCGCTTTGAGGCTGGAGAGCGCCTCGCCGAGGTGGGGTTCGCG
GAGGTCAA:CGGGGGCTGGAACCCGCACCTTCATTTCCAGCTGAGCTTGG
TCGTCCAGAGACCCACGATCTCCCGGGGGCGGTTCGGTGACGCGGATATG
CCGTGGGCTCTCAGCGCTTTCCCGACCCGAGGTGTGTCTCGGACCGCT
TACTGAGGCTGTTTTCAGCTGTAGCGCGCGATGTACGCGCCGACGCGAACC
TTTCGGGCGGTGTTTTCGCTCGACATGTAACGGATCTTGCCGAAGATCGG
ACGCTCTGGGCCCCATGCGCGGTGCAAGCGACCTAGGCACCAGAAGATGC
CGGAGTACGAGTTGGGGTCTCTGCCATCGAGGGCGTAGCGGTTGTTGAGC
TCGATAAGGGTGTGAGCGCGACCCGAGGGTGCTCGGTCCACTCAAGGAT
CTTCTTGGCCAGAGCATCCTCAGGTAGTTGTGGATCATCCCCTCACGGC
GGAGCTGTCGTTGCGCCGCGTTCAGAGCATGTCGTGCGTCTTGCCGGCG
TCAAGCGCCTCGAAATCGTAGAGGACAGGGCGACGGTCTCGGGCGTGGGC
CTCGAGGGTTCGTGAGCGCCACGGGGGGAGGCTCTCGAAGCGATCATAGT
CGCGCCACCTGACGTGACAGAAGGTGTAGCCGAGTTCCCGCCACGTCAGC
ACTTGATCGAGGAAGCTCTCAGCGCCCTCGCTCATCCCCACCAACCCGC
GCGGCTCCCCGTCGGCTTGGGAGCCACCTGGTCAGGGCTCCAACGCTCGC
GCTCGATGAGCTCCCGGAGGATCGTGTGTGCGCCGATGTGCCGAAGTGG
AGGTGGGGTGAGAGGCCAGTCCCAATGTCACGCTCGACTTGGTTCCGCTC
GCCATAGGCGTTGAGCTGTCGCCCCAAGAAGGCGTCGAGTCGGCGCCAGG
CTTCTTGCTCCCTCCGCGTGTGCTGACGGGCGTACGCTGTGGTTCGATC
GGGAGCGCGCTTAACGTCACGTTGAGGGCGCCTTCGCCCCGTGTAGGCTGC
CCATGGGGCGATATTGAGCGCGTCACGCGCCGGCGCAGGGGGCAGCTCGG

CAGACGCGAGCGGTTTCGGCCTCCGGCATCTCCGCGACGTGCTCCAGCACG
TTTTTATGGACGTGGCGTCGGAAGGAGTGGGCGGTGGTGAACCCCTCGG
GGCCGCGAAGAGCGGCAAGATCCCATTGGAGTCGACCACATCCACCTGCA
CGTCCAGGCG:CTTGGCCGCTTGTGCGACCATTCGAGGGACAAAGAAGCA
GGGGAAGTGGTCGCTGACGACCATGCAGGCCTGCGCGGCGAGGGCCTCGA
GGAGCCCGCTCCCCGCCCCCGCCTCCGGCTCGACGTAGGGGAGATAGGTG
ACACCAGCCACCCCGTCGAAGGCCTCCGCGTTGTCCCGCATCCCCTCGAG
GACGAAGGCATGGAACCGCTCTGACGCCATCGATAGCCACACGCAGGG
GCTCAAAGACCACGAGGGGCACGCCGAGGGTCTCCGCGTGATAGAGGGCA
TGCTGGAGCCCGAAATTATCGGTCGTGCGGCGCGCTGAGACCATCCAGTA
GAGGACAAAGCGCCTCTCGGGTCGTATGGGGGCCCGTTGGATGGTCTGGA
CGCGGCTCTGGGGGATGGGTGAGGTCATGATGTCTCCGTTTTTGGCGTCT
CGTCTCGGCTGAGACGCCTCAAGTGAGCTCCGAAGCGGACCATCACCCAG
CGGTGAAAGGTCCCTTGGGTGACGCGGTAATCCACCATGGCAAGCGCGG
TACAAAATCGTGGAGCGCCGCGATGAGCGTTCGGTCTCCGAGAACGTTTC
GGAATTCGAGACGCCCGCGCTCGGTGCTGCGTGCAGGAGCCCTCGGTC
ACGAGCATCACCTCGCGGTCAGGGCCGCTTCGGTGCGCCAAGCGGCGCAG
CCCAAGAGGGGTTGACCTGTGAGGCGAAAGACGAACTGGATCTCTTGGC
CGTCTTCCGAGCGCGTGATGTGGATGAGCCCTCGCATTTTACGCGCGAGC
CAGCGGAAGTACTCGTCGGCGACCCAGCGGATGTCTCGCCCCGGGGGCGT
CGTCATGCGCTGCACACTGCAGACCGTCGAGGGCCCCTTGGGGCGCGTGG
CGCGACGGAAGGCGCGTGGGACGATCGTGACATCACC:TTGCTCGTCTGC
GGCCTCGCGCAACATCTCGACGAGGGAGGTGGTTCGGGCGCTCTGTGCGCA
GCTTGGCGTCTGGGCGCGCGAGCATGTGCGTGGCGGAGGCTCTGGATAAGC
GGCGCCACCAGCGCCTTGGGCGCGCCGGTTCGTGAGCGTGACCCAGAGCCG
AGAGAGTCCCGGGGAGAGGACAGGGACCGAGATGAAGCGCCGCGTCAGTC
CGAAGGAACGCGCCGTCGCCGCCATCAGCTCCCTATAGCTCACGGCGTCC
GCTGCCCGAGATCAAAGACGCGGCTCTCATCGGGGACATCGTCCGGTGAC
CTGTGTGATCGCCGCGACCACGTCTCCGAGCGCAACGGGCTGCATGCGGC
TTGAGGTCC:AGCTTGGGCAGACCATCAGCGGCAGGCGCCTCACCAGCCG
CGCGAGCAGCTGATAGGAGGAGCCTCGGGCGCCGACGACCAGGCCACAGC
GCAGCGTCGTGACCGGGATGCCGGTCCGCCGAGCGCGCGCTCCATCTCC
TGACGACTCCGAAGGTGTTGCTCAACGCGTTCCTCGCTCGGTAGGAGGCC
GCCGACATAGACGATGTGCTCGACACCGACCTCGGCGGCGCCCGCGCGA
AGTTGTCCGCGCAGAGGACATCGAGGTCTCGAAGTGGCCTTGAACGAGC
CTCGCGGCGGGCATCATGCTGTGAACCAGGTAGACGGCGAAGCGCGCGCC
CCGGAGCGCCTCTGTGCGCCTGAGAGCGAAAAGAGGTGACCTGACGGC
AGCTCCGGTAGCCTCCCCGGGCTCTCTGAGCGAGCGCGAGAGCGAGATG
GGCGCGAAGCTCTCGGAGAGGTTGGGTCCAGGGCTTGTCCAATGAATCC
CGATGCTCCCGCGACGGCGACCATCACGTACCCCTTTGGTGGCGTCTT
GTGTCATCGCTTGTCTACGCGTCGCGTCATCGCCGTCTGCCCTTGTGAA
CCCTGAGGGGCGGCCCTCAGCTCTCCTTAACACCCGCTGGTCCCTGCGC
TCTTCCCGTCTTTGACCACCTCGGATGAGCGCAAGGGGAGGCTTCATGC
GTAGCACGCAGTGGCGCGCCGTGCGACCTCGGCGGTGGATCTCCGTCGT

CATCCTCGGAGCGCTCGTGCGGGTCATATCGTGCGGTCTCTGCGGGTGA
TGCCATGTGGCCCTCTTTAGATGATGTGTACCCTCTGATTCGTCTTGAGC
GAGGCTAGATCGTAGCGCGTGCCGCGCCAGGAGACCGTGCCATGGATGAG
GGCCTTGAACGTCGAGCGCATCAAGATCAGCGCCATCACGCAGGGCGCCA
AGGGGCTCAAGAGGTAGGCTAAGGGTCGGGCGCCCGCAGGGTTCGCGAAA
AAGCAGAGCAGCCCATTGAGCCCCAGGGTCAGCCCGCCGAGCCAGGCGTA
GGGCCC GAAGAGGAGGAGCAGAGGGCTCAGCGGCATCATCAAGGCGCCCA
CGGTGATGACGCACGCCCTGAGGGGGCTAAAGCGCGCGACGACCGGAAAC
ATGTTCTTCTCAAGG:CCGCGTATGAAC:CCCCG:AG:GCT:CGTATAC
CAGCGGAGGCTTAAGAGCCGCGGCGCAAGCGCGATCCGTCCCCGTGCCCC
GGCGCGGTGCATCATGAGGGCCAGGCCAAG:ATCGTCGGCGATCTCCAGC
CTCAGCCACTCCACCCTGGGCTCGCCTCAAAGACAGAGCGCCTGACCAT
GCTGAAGGCCCGACACCCGCGTAGGCCTCCTCAGAGTCTCGGTTGACCG
CCCGCGCGTTGGTGCCCGCGAGAAAGAAACCCCGAAGCCCC:AAGCGT
CGCATT CAGCAGGACGCCCTCAGCGTG CATCTTTGGGATGAGCGCCATCA
TGTCGATGCGCTCGCGCTCGGCCAGGCGACGGCGGCTTGAAGTGCCCC
GGGGCGAGGTGCACATCGGCGTTCGGCGTAGATGACGTAGTCGCCCTTCGC
CTCCCTGCTCGCGATGTGCATCGCGTGAACCTTCCCGAGCCACCCCTCAG
GCAGCGCCTCGATGTGGTGGACGACGACGCGCGCGTGCCTCTCGCCACG
TCGTCTATCATCGCCCCGGTGGCGTTCGGTGGAGCGGTTCGTTGACGACCA
GACCTCGAGCGCGGGGTACGCGCAAGCGAGGGTGCTCCTGAGCGCGGCTT
CGATGGTCGACGCTTCATTGCATGCAGGGACAATGAGCGACACCGTGGGC
CAGGTCGCGCAGGGCGTGACTGTGGTCTCTATGAGCTCGGGCACGCCGCG
CGCTCGGTAGAAGACGACGAGATAGACCATCAGGCTCAGGAGCGAGATGG
CGCAGAGGCTCAGGCTTATCATGGGTCCAATCGGTCGTGCGTGGCGCCGC
TCTTATGAGGCGTTGAGGCCGCGATAGCCCCTGACGCGCCAGAGTCGCCA
CCAGGGCAGATCGGGCCGCGCGTGGTGCTCGAAGTGATAGCCGAAGTGAT
AGCACGCGAGGAAGGAGAGCCAAGGAGG:CAGGTCGACGCTCCGCGCCCC
CAGCGGCCCGTCCCCAAGTAGGCTCCTGGGCGGTGCGGCAACCATGTCC
CAACGATGAAGAGCTGAAAGGTGCTCAGGATGAGCGGTGTCGCCAAAAC
AACGCGACGCGGGCGACGGACATCTCAAGGACGCCGATCATCACGGCTAA
GAGGAGACAGAGCCATGCAA ACTGAGCCCAGCTGAGGTA CTGGCGCATGA
AGCTGATGTACCACGCGACGAACCTGGGTCCGACGGTGGAAAGTCGGGA
TCGTCCGCTCGCTGGTTTGGCGTGGTGCTCGTGATGAGCGCGCTCGAG
CGCGGCGTAGTTGAGGCCCGCGTAGCTCCACGTCGCGAACCAACCTATCG
CGCGTTGAGAGCGGCG:T:T:GCCCGAGCGATCAGGCCATGCATGGCG
TCATGCGCTGTGATGAAGACGCCGGTGGAGAGGAAGGTCGTCAGCGCGAT
CCAGAGCCCATCGGCGATGATATGGCCTGCTCCGGGGTGTTACCGCGGA
GCCCCAGGAACATGGCGAGCGCCAAGCGCTGAGAACCGCCGCGGCCGT
AAGAGCCCAGGGGTGCTCCGAGTGGGCGCCTGAGGGGCCTTCGCCTCCGA
GAGGCGTCCGAGGACCGCTGGCACCCCTCCGATAAAGGGTCCCCAT:GGG
AGACGCAGAAAGAGGCTCATCTCATCGATGAGGCGGGTCTCTCATCCAG
GAAGGTCAAGATCGGCGCGATGCCTTTTATGCCAAAGAGCCTCTGGAAGA
TGCTCGCGCCTCGCTCGGGCTCCTCGTGGAGGATCCTGAGG:AGCAGCAG

GTCATAGAAGCGAAAGCGCGTCGCTGCCTCTGGGACCTGAGAGAGTCGAC
CCCTCTCGACAAAGCCCTCGGCGAGGGCGCGCGTGTGGCGCTGGATACGC
TCAAATGCGTAGCCGGTGCTGGGCTTCATCTGCCACCGACGATCCCGAT
ATTGAAGACGTGTGGGCTGCTTCGTTGGGCAGGGTAGCGATCGTCCATCG
GCAGCGACCCATATTCCTCACGCTCGATGGCGTACTCGGTCTCGTATTCG
CGCGCGATATAGGCGAGCGCGGCCCTCTTTGTGGACCTCAGGCGGCAGCGG
GTCGGTTGAGAACATCGTATGCTCGATAAGCGCTTTGTTTCGGCGCGAAGG
GGAGCACGTAGACGAAGGCGACCCCGCCTCGTTGGGGAACGCTAAAATCC
ATCATGGTGAAGCGCTCTGGGTCTGAAGACGGGCTCCTCCGTCACGATCTC
TACGCCTCCGAAGTGCTGCTTGACCGGGTACTTGAGAGGTTGGGCTTCGT
CGTCGGCCCCGTAGCGGACGCTCTGCGCGACCCACGCGGCTCGAAGCTCT
CTCTCTGGGGTCCGAACATAGGCGCCATTGTCTGTCCTTCGCCGATGGCGAG
GATGTCCCCCTCGATCAAGGTGATGGTATCGCTCTCCTCGATGCGCCTGA
GCACCTCCCTCTTGAAGTGTTGCTGCGGACGCAATGATAGCGTGTGTCC
TCGAGCGGCTCATCACGCACCTC:TGTGTCGAAGGCGACGCGCGCCTTTT
GCCAGGCGTGATCGGCCAAGGCGCTGAA:GGGAGCGTCATCGGGGCCCA
GAAGCACCAGAGGCGGTCCGCTTGGGCCTCAAAGCTCCGATCCACGACCG
CGACGCGCTGGCCTTCAAGGCCGGCGTCGAGCAGGTGCCACGCGAGGCTT
AAGCCCGCGCAGCCCGCTCCGGCGATAAGCAAGTCGTAGCGCTCCGTCAT
CGAGCGAGCGCCTTACCCTCGTCCCCATCGGGGGTTCGAGCGCGTCCGGT
GCCAGGGCGGCCTCGCAGCGCAGTGTGGAGCGCGGGGTCATGCGCTTTGC
GTGCGCTCAGCCCGCAGAGGCTCGCTCGAACACCAGCGCGGAGGACCCAA
GAGACCTTCGCGGCGCCCGGACAACAGCGCGCCCCTGGGTGACATCATA
GGCTCGGTTCGCGGAGGACAAGGCCGATGGCCCGATAGACCCTGCTCGCCG
ACTGGATGGCGAGCCGAGGGCGCCACGGGATGGCGCGCATGCCCTCATCA
CCGCTCGCATAGTAGCGTTCGGCGAGCTCCAAGACGTCCCTGAGGACGGG
CGCGAAGGCCTCCCCGTCGATACTCCGACGCGGACTTGCTCAGGGTCTA
TCCCCGCCGCCACCAGGCGCCGAGTTGGGATGTAGATGCGCCCTCGGCC
GCGTCTCGACCACATCCCCTGAGATGTTGGTGTATCTGCATCGCGACCCC
CAAGTCGATGGCGTTCGCGCCAGCGCCCATGGCTCTCTACCCCGATGACCC
CACACATCATGAGCCCCACGGTCCCAGCGACCGCATAACCGTATTGGATG
AGGTCTCGGTTCGTCGACAAGGGTCACTTCCCCACGTTCGCTCAGGACCCC
CTGCATCAAGTCATGGGCGGCCTGAAGTGGAAGCCCCGGCGCTCGGCGA
CCTCAAGGAACGCCTCCGCGAGCGGGCTGGCGGGCCTCTACCGCGCAGC
ATCTCAGCCAGGTGCCCGATGTCGCGCTCGCCCTCCTCGGGGTGCCGCGC
CTCGTCGACCAAGTCATCGACGGTTCGACAGAAGGCGTAGACGATCGCCG
CGTCGTCCCGTTGGTTCGCGCGGCAGGAAGACGCTCGCGAGGTTAAAGGTC
CGTGCGTTCGAGGCGAGGACCTGACGGCTCTCTGCAGCCATCGTCCGGGT
TGCCAGCGCGCTCGTCATCTCAGCTCGCCTTTCGGAGCGCGCCCTTCTTG
GTGGCCTCGTGGCGCTCGCC:GGG:CTC:TCGTCCTGGACGCTGCACCTG
GACCTCGGGGTAGTCCCTCGGCGACGCACCGTGCGGTTCATCTTGGCGCTCA
TCATACCGACGGGGTCCCAGCGCCGGGTTGCGTGTTTCGCCCCGACAAGG
TAGAGCCCCTCGATCGACTTGACCCGGTTGTGCGGACGGAACCAAGCGGA
CTGCGTCAAGACAGGTTTCGAGCCCGAAGCCGTTGCCAAGGTGGCTCCGGA

GGGTGTTCTGGAAATAATCCGGGGTGACCCAGCTGAGGTGCGTCAGGCGC
TCCTTCAAGTTGGGGATGTAGCCCTCCTCATCGAGCATCGTGAGGACGCG
GTCAACGAACCCCTCTCCGAGGGCCTCCCAGTCGAGCCCACTCCCGTTGT
GGGGGACAGGGACCAAGGTGTAGGCCGCGTGGTGCCCAGCGGGCGCCATA
GCGGGGTCCGTTAAGGTAGGGACGTGGAGGTATTGGCTGAAGTCTTCCCC
GATGACCTTGGTCTGGAAAATCTCGCGCAGCAGCGCTTCGTACCTCGGCC
CGAGGATGATGTTGTGGTGGCGCAGGTCGAGGGGAACCTCGGGGTTCATCG
CTGAACCCGAAGTAGATGACGACCAAGCTCATCGAGTAGCGCTTGGCGCG
CACCACCCAGTCTGGCTGGGTAGGTCGCACCTTCGAGGGCAGCATCTTCA
TCATCGTCGTCGCGTAGTCGGCGTTGGAGACCACCACGTCGGCATAGAAC
TCTTGGCCATCCTCGAGCCTGACGCCCCGCGCGACTTTGCTCGTCAGACG
CCCGGGCTCCTCGCCGTCCCCTGTGGTGACGAGGATCTCATCGACCGGCG
CGTCCAAGAGCACGGTGCCCCCATCTCCTCAAACCTTCGAACGAACCCG
GAGACGAGCGCCCCGGTGCCGCCATGCAGTAGTGTACCCCCAGGTCTT
CTAACGAAGTGGATCATGGCGTAGATCGCGGGCACGCTCATGGGGTTGC
C:GCCACGAGCAGGGTCTCGAAGGAGAAGACCTGCCTCAGCTTGTGCT
CTTAAAGTAGCGGCTGGAGAAGCGGAAGAGGTTCCCTGACCGCGTCGAGCT
TGAGCAGGTCTGGCAGCACCTTGAGCATCGAACCGATGTCTAAGAACCGG
GTATAGCCGAGCTCCAAGAACCCGCGCTCAAAGATGGCGCCAGCGGCTTT
ATGGAAGCGCTCGTACCCCTCAAGGTCTTCTTCGCAGATCTCCGCGATCT
GTCGCCGCGTGTTGTCCCGATCGCCGTCGTAGTCGAAGTAGGTCCCGTCA
TC:AAAGTAGATGCGATAG:AAGGGGCTCACCGGACGGAGCTCGCAGTAC
TCGCTGGTCCGCGGACCACCGCTCACCCAGAGGTCACGCGGGCATCGGG
TCCTAGGGTCTCCGGGGGAAAGTCGGGCTGGTCGAGGCCGTGGACCCCGC
GCTCAAGCGCGAAGAGTTCCTCGATAAAATGGGGGACGGTGATGACGGTC
GGGCCATGTGGAAGATGTATCCGTCTGCTTACGGACGTAGGCTCGGCC
GCCCGGCTTGTGAGCGCCTCGATGATCGTCGTCGTAAGGCCATGCTCT
GAAGCCGCATTGCCATGGCGAGCCCGCCAAAGCCGGCGCCAATCACGATG
GCCTCTTTGCCTTCATTTGCCTCGATATCTGTTCATCATTATTAACCTCCTC
AACCTGTCAGAGGCGCCGCGATGGTTTGCGCCGTCGCGCTCACGAGGGC
CCGGAGGCCCGGCTCCATCGCGAGCTCGGCGGACGCGACGAGGTCCCGCA
GGAGTCATCTGCCAGGCGCACGTGTCCTCGAGCGCGCCACTGAGAACC
AGGCGCTCAGACCACTCGGCCACGTGTCGTCGTCATGGGTCTCGTCGCGGTC
GCGTTTCAAGAGCGCGACCAAGCTCTTCGGTGTCCCTCAGGGCGCCGCTCCA
GGTGAGTCACGGTGAGCGCGCTCACCTTGCCTTACGCAGGTCGTTGCCG
CGTTGCCCGCGCCCCTTGTCCCCGAAGAGGTCGAGCACGTCATCTCGGAT
CTGAAAGAGCGCCCCAAGCGCCAAGAGCGCGTCGCCAATGGCGCGCGCTC
GCGTCGGTTGCCGCCCCGCGGCGAGCGCCGCGGTCTCGATCGGCAGAGAG
AAGAAGGGCCCGGACTTCCCGAGGGCCGCCCTCAGGTAAGGGCCACGGCC
CAGCCATCCTCGACCAAGAGCTCGAGCTCAAGCGCCTGACCGCAGGCGG
TCTCCGTCGCGCGACGCGTGAACGCCGTGGCGATGTACCAACGCACAGAG
GGCATGAGCAGCATCTGCTCGAGCGCGAG:GGTCGGGAGCATCAACATGG
CG:T:CGCCGACGTTGATGGCTTGCGCCTACCATGGCGCACCCATACGG
CAGGTTACCGCGTCGCACGGTGTGCCATCCTGCAGGTCGTCGTGAACC

AGCGTGGCGTTATGGAGGAGCTCACAGGCCGCCGCGATGGGCACCAGAGT
CATCGGCTCCTCTCCGAGGGCGCGGGCCGCCTCGAGCACGATGCGTGCCC
TGACCTGCTTGCCGCCCGTTTGAAGGTGCTGGACGACCATCGAGCGCAAG
GGCTCATTGAGGTGCGCGTGCGCGGTCTGAAGCGCGTGGTCCCTCACCGC
CTCGAGGCCACGCTCGAGGGCGTGGGCCTGAGTGCCTCTTACGGCGCC
TGAGGATGCTCTGCGACGTCGGTTCCGCGACGAGTCGGAGCGAGGCAGAC
TGTTTCGGGACCATGCGAACTCCTCGGACGTCTCGCTGGCGTAAGCGTGC
GTGGCCGCTCCGCAGACCCAGCGGTTGCTGACGTCTTACAGGCGCTAGGA
TACGATCTGTTGAACATGTTTTGCAATGACTTTGGGTGTTCAAAGTCAT
TCATTACGAACGTGTTCCCTCCACAT:GGGGCTCTCGTCCGTCGAATACG
CCCATCATGGATGCGCTTTCAAGCGCGCACCCGCGGCGTCCGCGGTTG
GCGCTCGGGGGCCAGACCTTGCATTGAGCGCCGATCGGGTGCGGTCTG
CTGCGCGCTTTCGTGCGTTGATGGAGCGGACCGAGA:GGGGGGAGCT:CC
TCGAAGGAG:CG:CT:GCCG:G:GGAGGCTTAACCTTCGGGGTTCA:AGC
CGA:TCGGACAAGAGATGCCGGTGCCGCCGATGCCACAGTACCCTGAAGG
GTTTTTGTGCAGGTATTGCTGATGGTAATCCTCGGCGTAGTAGAGGGTCG
GCGCGTCTAGGATCTCGGTGGTGAATTGGACCGAAACCTGCGCTCGCCAGG
CGCGCTTCGTAGCGCGCCTTACTCTCGAGGGCCTCTGCACGCTGTGACGG
CGTGGTGACATAGAGCCCGGAGCGATATTGCGTCCCTCGGTCCATGCCTT
GACGCATGCCTTGGGTTGGATCGTGGTTCTCCCAAAGTGCGCGATCAGC
GCCTCGTTGCTCAAGACCGAGGGGTCAAACACGATGAGGACCACCTCGTT
GTGTCCCGTGAGCCCCGTGCACACCTCTTCATAGGTGGGGTTCGGGGTGT
GGCCTGCGGCGTAGCCGACCGCGGTTCGTCACACGCCAGGGAGGCTGTAA
AAGAGGCGCTCAGCGCCCCAAAAGCAGCCCATTCCGACCAGGATCACCTC
GAGCTGGTCGCCGAAGGGCCCCACCATGGGGTTTTGGGTGACGATGTTCT
TGGGTGCCACCGCCACCGCTTGGTTCGCGGCCGGGAAGGCATTCAGACGCG
GGCGGGATGGTCATCGGGCGGCGTGAGAAGAACACGGGCTTACCTCCAGG
ATGAGGACACGACGATGCGTCCGTCGAGGTCAGGATTGTGCGCGATGAAC
AGGCGCAGTCAAGCGCCCGTCAAGGTGTTAGCGCGCCTTGCATCGTCAT
GAGGATCCAGATGGCCACGCCGCTCCGCAAGGCTCCAGCCCCATGTCCTCA
CCCAGTTGGAGCGCACCAAACGCCGGTGGGTTCGCTTGGTGAAGCCGGCG
GCGAGGCGCCCGTGAGCCGGACAGAGTAGAGCGCCGTCGCTGCCAGAC
GACGACCAAGGTGATAAGCCCGATGATCGCGGGCAGGGCCGCGGCGTCAT
TGGGCGGGGCAAGGGCGAGCAGCGCGGCGCTCATCGCCTCAATCAACATT
GGGGGACCGACCACCCACCTGTCCGGGTGACGTGGCGCGCCTGATAGTC
GCGAAAGGTGGCGTCCCCGACCGATGACTTGAGTGGGTAATGGACGACCT
GGACGAACCAGATCAGCCCCGTCATGAAGAGGGTGGCGCTGAGGTGCGCG
AGCACCAGGGCCCAGGAGAGCGCTTCCGGGGTTCATGCGCTCACCATGGGA
TCCGAGCGCCATCCCAGGCGAAGAGCGCCCCGGAGTCCGAGGGAGTGAGG
CCACCTATCACCTCGCAGAGATGTCGACAGCTCTCCTCGGGCGAGAAGAC
GACGTGACCGCGCGGAGGCGCTTGGTGAAGGGCTCGCTGAGCTCGGTGC
CCACGGTGCCAGGGTGCATCGCGCAGATGATGAGCTTGGGGTATCGGTTCC
GCCGCTCAATGCTCGCCGTCTTACGATCATGTTTTGAGCGGCCTTGA
GGCCCGATAGCTGTACCAGCCGCCAGGCGGTTGTCGGTGAATGCTCCCGA

CCCTGGCCGAGAGGGTGATAAAGAGCGCGTCCTGCCTCCTCGGGACCGCT
TTGACGAGGTGTTTGTATGGCGAGCGCCGGCCGATGGCGTTGACCGCGAA
GGAGCGGGCCATGTGCTCGGCGTTGAGCTCTCGCCAGGTGCGCTCTGGGC
GGAGCTACCATCGTGGAGTAATCCACTGCAGTTCAGGACGACGCGGGGC
GCGCGCTCTGCGCGCGAGCCCGTCGGCAAAGGCGCTCAGGCTCGTCTC
ATCGCTGATGTCCAGGAGTGC GCGCTCCACTTTGGGGT GCGATGAGGACC
CCTGTGCCAGCGCGCATCTCGACTCGTTCGCGTAGACGCTGCGGACAGAC
TCTGCCTGAGCGAGCGCTTCGGTGAACGCCTCACCGATGCCGCCACGCGC
GCCGATGACAACCGCGTCGACGGGGCCCTCCATAGTTAACCACAGC:ATC
CGAAGCTCCTGGTCACGTGACCGACTGTAGGGACGCTAGAACACCTCAGC
TCTTGGGTCTACTCCCTGCCCTGAGCGCTTGTGTGCG:TTCGTCTATGTC
CCCATCTCCTCGACCCCGATGATGAGGCGCCTGAGGGATGCCTCGACCTC
AGGTCTCCGCCTCGAATACTTCCCAGGGTTCATGATCCCCGATGGATCG
AACACCGCTTTGATGCCCTGCATGA:GGCGGACTTCCTCTGCGCCTCGGC
TGAA:GGTCAGGTGCGAGCGCTTAAGCAGCCCGACCCCGTGCTCCGCCGA
GATGGATCCCTCGAAGCGCGGACGGCCCCGTAGACCTGATCGCTGATGC
GCTCGCATCGCGCTTTGAACTCGCTCGCCTCCATCTCCTCTGGCTTGAGG
ATGTTGAGGTGCAGGTTGCCATCGCCGATGTGCCCCAACCCAGATGATCTC
GAAGCTTGGGTACTCCCGTTGGACGATCTCCTCGATGGCGGCGAGAAAGG
CTGGAGCTTTTGAGACCTTCACGCTGATGTGCTTTTTGTAGGGCGTCCAG
CGCGAGATGGTCTCGCTTATCTCCTCGCGACACCTCCAGAGCCGTTTGGC
TTGCGCGTTGTCTTGGCTCATGACGCCATCGGAGACGTCACCGCTCTCGA
CGCAGGCCTCGAAGAGCTCGAGGGCCACCGCTTCGTCCTCCTCCGTGACG
CACTCGAGCTCCAAGAGCGCGTAATAAGGATGGTCTCCAG:CCATCGGGC
GCGGCACCTGGGCATGCTCGATGACCTTCGTCAGCGCCTCGTGAGAAAAG
AATTCAAAGGCGCTTAAGACCACGCGGTTTTGAAAGACGTCTAAGATGTT
GAGGACGGAGCCCATGCTTGGGGCCCCGAGCACCAAGACAGCGCTCTGAG
CTGGCGCGCGCTCAACGAGAGCGTCGCTTCGACGATGAAGCCGAGCGTG
CCCTCTGAGCCGATGAAGAGGTGCCGAAGGTCATAGCCC GCGTTGTTTTT
CAGCAGGCCTCGTTTCAGGTCCAGAAGCTCACCTTGCCCCGTTGACGACCT
TGAGACCCGCGACCCACTGACGGGTCATGCCATAGCGGATGACGTTGATC
CCCCAGCGTTGGTCGCGATGTTCCCGCCGATCTGAGACGACCCCGCCGC
CGCGAAGTCAACGGGGTAGAAGAGGCCTTCTTCGCGCGCGTAGGCCTGCA
GCTGCTCGGTGATGACCCCGGGCTGGCAGACGACCAGACGGTCGACGGGG
TCAAAGGAGAGGATCTTGTTCATCCGCGCAAAGGAGACCACAACCTCGCC
GGCGAGCGCGCAGGCGCCCGCTGAGCCCCGTGCGCCCCCGGAGGGCA
CGAGGCCGATCATGGACTIONTGGCCACGCGACCCAGCGCTTGCACCTCT
TCGACGCTCCGAGGGAACGCGATGGCGAGCGGATTGGGCGTGTAGAAGCG
CGTCCA:ATCGAGACCATAGTGCGTGAGATCGTCCGGCTGGGTGGATAAC
TCCAGCGCGGGGCAGCGCTCGCGGAGCGCGTCGAGGGCGTTGTCTGGATC
GCGTGTGATGGCTCACGTTTGGGTGCGTCTTAGGTGTTTCATGACAGCGC
CAGCGGGGAATCACAACCCCAAAGGCACGCGGCCTGCTACAAGGCCCCG
GTCGGGGTGGCGTGACGCGGGCGACGAGTCGACGGGGACCCAAGGACCG
AGACCTTCTAGGCCACAGACAAGGCCCTCCCGGCCAGGGATGCGTTGACG

TCGCGCTCGGGTCGGATTAATCATCGGGCTCCATTGTAAGGAGTCCCCCA
TGAAGCTCACCTCCGTGTCATCGCCCTCGCTGCCGCTCTCAGCCTCGTC
GCCTGTGAGAATAAGGCCCTGAGCCCGCGGCTCAAGCGCCCGCCGCTAA
GGTGGAAGCGGCTCAAGAGGAGCCGGCTGCGCCGACCTCGCTGCCAGCGC
CCAAGGGGGCCAAGGTCTTCTTCGACAACCTCAAAAACGGTGACGAGGTC
ACGAGCCCTGTGACGGTCTGCATCGGCTCGGAAGGGTTGGAGCTTGAGGC
GTCTGGCCCCGGTGAACCCCGGTAAGGGTCAACCACCATCCTCGTCGACA
TTCCGGTCCCCGATGACCTGAGCAAAGATATCCCGAAGGACGCTCAGAAC
ATCCACCTTGAGACGGCTCCTCGTGC GCGGACGTGACCCTGAGCCCAGG
CGCTCACACCCTGCGCTTGCTCTTCGCGGACGGCTCACACCGCCCCTATG
AGCCCGCGATCACGGCCTCGATCGACATCAAGGTCGTCGAGAAGAAAGAG
GCGCCGAAGCCGGCGGCCGCGCCGCAAGGCGCCAGCTGAGCCCAC
GACGGCAAGGTCTTGCTGAAGCCCAACATGGCCCCTCCAAAAGCAGGGA
TGGCCAAGCCCCAAGTCATGCTCAAGCCTGGCGCGGCCCAAGGTGATC
CTCAAGCCCAACCAAGCCCCTGTGGTTCGTGAAGCCTGCGGGTGATAAGCC
CAAGGTGGTGGTCAAGCCCAAGCAGGACAAGCCCAAGGTGGTGGTCAAGC
CCAAGCAGGACAAGCCCAAGGTGGTGGTCAAGCCGAAGCAGGCCAAGCCC
AAGGTGGTGGTCAAGCCGAAGCAGGCCAAGCCCAAGGTGGTGGTCAAGCC
GAAGCAGGCCAAGCCCAAGGTGATCAAGCCCAAGCAGTAGCATCGCT
GCGCGACCTTTCGTGCGGAGACTGAGCGCCGATCGTCGGCGAGCTCTCCC
TCGCGTGCGCCACGCCCTCTTTCGGGTGGCCTTAGCGCGCGCCCAGCTGA
CGATGATCCGGTTCAGGAGGCCGATTGGGCGCTCTCGATGCGCGAGGTGC
GGTCCCCCTGAGCGCTGAATACCGTGATCACCCCTCTCTTCCGACGTGA
AAGGGCAGGGTGCATGGGACCCCTGCGCGGCGCGTCTTCGTGATGCTGCA
ACGGGTTTGTGGAGGGCTCCGGCCGCTTACCCATCGGGAGAGGGGCAGA
GATTCGCGTCGGTTCGGGATTACGGTCTCGTGAAACGCATCGAACCCCTG
TATGCTCTCGTCCTCGTGACGGGCCAAGCGGAGGTGTTTCGGTGGATTGG
CGTGATTCAGTACGTTAATGGCATTCTCATGGGTTGCGTTGGGGTGCCA
GGCCGAGCAGAGCACAGCGGATAAGGTGCAACTCGAGACCGCGCAGCTCG
CGATTAAGGGCGGCTACAACCTCGACGGACAAGGCCGTTGTCGGCCTCGGG
ATGGCGAGCCAGGGAGGCTTCGGCACCTGCAGCGGCACCATCATCAGCCC
GAACGTCATCCTGACGGCTCAGCATTGCATCGCGAGCATGTCGACGAACA
ACTACGTGGACTGCAACCAAAGCACCTTGGGCCTGCGCAGACCCCTGAG
TATGCTTTCGTGACCAACCGATCGCCGATGCCCTGGGATGACAGCGCTA
TATCCGATGCCAGCAGATCGTCGTCCCGCAGGCCTCCA ACTGGGCCTGTG
GGTACGATATTGCGCTCCTCATCCTGGAGGAGCCTATCCCTGAGTGGTGG
GAGGCCTCACCGATGATAACCGCGCGTCGACACCGGTCATAACCGCCAACGA
CCCCTACTCCGCCATCGGCTACGGCTCGACCAATGGACAGGGACAAGGCT
CTGGGACTCGGCGTCGACGAGACGGTCTCTACACCAACTGTGGCCCGGGC
GAGTGCCCCGGCTGGTATCCTCTGACGCCGTCGGAGTGGATCGGCGACAC
CGGGGTTTGTAGGGCGACAGCGGCGGGCCCGCCATCGATGCCTTCAATC
GGATCATTGGTGTGCGGTCGCGCGGGGGCGCGGAGTGCTCAAGCCCCGTG
TACTCGGGGGTCAACACCTGGGGCGCCTGGATCAAAGAGACCGTGGTTTA
CGCCGCGGAGCTCGGCGGCTTGATGTGCCGCTTGGGCGCTAGGCTGGC

CGACGCACCCTGATTATTCCTACCCCATCGGCGGCTGGTGCCTGGCGGG
AGTGACTGTGAGTCTGGTACTGTGTGGGCGCCATCTGCACGCGGCGCTG
CAATGACGCCGCGCCCTGTCCGACGGGTCTCTTCTGTAAACGGTGACGAGC
TCTGTGAGGTCGTGCCCGTCCGGTGGCCGTTGCCTTGAGCCTTCGGAGTGC
CTCGGTGGGTGGTGCCTTGATGGGACGTGCTCCCGCACCTGCTCTCAGGA
TATGCCCTGCCCAACGGGCTTCGAGTGCGACCCGGTCTCAGTCGGCTGTC
GTCCGGAGATCGGCGTCACCTGTCTCGACAATAGTGAGTGCTCCTCTGAT
ATCTGTCTCGGCGACTTATGCACCCGTGAGTGCACGCGCCGATCTCGACTG
CCCTGAGGGCTGGAATTGCACCAACGGCGCGTGCTTGCTCGACAAAGTTG
GCATGGGCTGCAAAGACGACCTGGCGTGTCCGGGTGGTACTGTTTGGAC
GGGATTTGTAGCCGTGCGTGCGCCGACAACACGCCCTGCCAGCGGGCTA
CCTCTGTGACGCTGCGGTGGGGATCTGTGAGCTGGTCCCGGCCGGGAGCC
CTTGCGAGAGTGACCTCGACTGCAGCTACACCGTCTGTGTGGAGGGGATC
TGCTCTCGGACATGCTCCGATGAGGCGCCCTGTCCGCCGGGGCTCTTCTG
CGCCCTCGAGTCTCAGATGTGTGAGGTCGCGAGCGTCGGAGGAGGCTGCA
CGCTCGACGACGAGTGCGACTCGGGGATGTGCGCCGACGGGGTTTGTACG
CGGGGCTGCGGAGGCTCGGTGGGCTGCCAGCGGGCTACCGCTGCGATGT
GCAAATCAACCAATGTCTACTCGTCGAGGCCGGCGGGCGTGTGCACCGAGG
ACGCGGATTGCAGCTACAGCGTCTGTATTGACGGCATCTGCACGCGTGCA
TGCTCGGCGGAGGCGCCTTGTAATGACGGCCTTGTCTGTGACAGCGCCTC
CGGCAAGTGCACCCTTGGCTCGCAGAGCAGCTCTGGCGACGGCGACGCGA
GCACGGCGAGCGTCCCTGAGTGTGAGCTCAGCGAGGGATGCTCGGCCTCA
GGGTGCTCGGATCCAAGCTTCACTGGGAGCGCTGACTGCGTCCGCTGCTGG
CGCCACCTGGGGTGCCTCACGACCCGAGATGCCTGCCTCGCCGCGGGTC
ACAGCTGGGGCCCGCTTGAGACGCGAGGATGCCTGCATCGCGGCCGGGAGC
GTCTGGGTTGAAGGACGTAAGTGGATGTCATGCAGGTCGGCACCCATCGCT
CCCCTTGGGGGCTGGGCTGGCCATCTTGGCCCTTCTCTTCTTACGATGGA
GACGTTCCCTCCCTTCTCTCACCTAGCCGCCAGCCATGCTAAAGCTCAGG
ACCCGGACCCTCGGGTCCGCCGATAGTCCCCCGCTCCTCTTGGGGGTGCT
CCACCTCGCGCCTCTGCCGGGGAGCCCTCTTTCTCCGGGGACATGGACG
CCGTGCTTAAGGCGGCTATCCACGACGCTGAGGCCCTTATCGCCGGGGGC
ATGGACGGCTATGTCGTGGAGA ACTTTGGTGACGCCCTTTTTACCCGGA
GCAGGTTGGGCCCGAGACGCTGGCCGCCATGGCCCGGGTCTGGCCCTGC
TCCCTCGGGGTGATGTGCTCGTGGGCTGCAATGTGTTGCGCAATGATGCT
CAAGCCGCCTTGGGCTTGGCCTCGGCGTTCGACCTCGACTTCATCCGTGT
GAACGTCCACATCGGCGCGGCGTTGACCGATCAAGGGATCATCGAGGGTC
GCGCCGCCACGACGCTCCGCACAAGGGCGAGGCTTGGCGCCAGGGCCGCT
CTCTTCGCCGATGTTGACGTCAAGCACGCGAGCCCGCTCGGTGAGCCTCG
CCCCCTGGGGCTCCTCGCCGAGGAGACCGCCGCGCGTGGGCGCGCAGACG
CGCTCATCGTCTCGGGTTCACAGACAGGCGCGCCGACGTCGATAACAGAC
ATGCAGGCCGTGAAGGCGGCTGCCAGAGGTCCTCTGCTGATCGGGAG
CGGCATGACCCTTGACGGGGCCGACAGAGGCGCTCGCCATCGCGGA:TGGG
GCGATCGTGGGGACGGCCCTCAAGCGGGGAGGGCAGGTATACGCGCCGGT
CTGCGCAGAGCGCGTCGCGGCCTTTGTAGGAGCGTGCCGATGAGTGAAGG

GACAGCAGCGCTTCGCGCCGATGCGCGCGCGATCTATGACGC:GGCGGTC
GACGCCGTCGAGCCCGCGAAGGCCGGTGCGGGCGGCGCTTGAGAGGCGTCC
CTTAGAGGGAGACCCCTACGTGGTCGCGATCGGTAAAGCCGCGAGCGCCA
TGGCGGGGGCGGCCCGGATAGGGGCGCGCGTCGCGGGCTCATCGTCACG
AACATGGCCAATGCGCGGCCTCGGGCTGGCTTCCGCGTCTCTGGTGGGTGA
CCACCCCGAGCCTGGGAGTGGGAGCATCGCGGGTGGTGACGCGGTTCGAGC
AAGCCTTCCAAGAGCATGAGGAGCTCTTGGTGCTGGTCTCGGGCGGCGCG
AGCGCCATGGTCTGCGCGCCAGCCCTCGGCTTGGAGCTTCATGATATCGT
GAACGCAACCCAGCTCTTGATGCGTCGCGGCGCGGACATTGAGGAGCTGA
ACACGGTTCGTCGCCACCTCTCCCGGATCAAAGGTGGTCAGGCGCTTCAA
GAGGGACAGGGGGTGCAGCGCGCTCTTGCTCAGTGACGTGTTAGGCAATGC
TCCCGAGGCGATCGGCTCTGGGATCGCGGCGCCTGATCCCACCACCTATC
GGGACGCCCTCCGCGTCTTGAGCGCCATGGGCTCTTAGGGGCGGTGACG
CAGGCGGTGCGGAATCACTTGGAGCTCGGAGACGCAGGGGCCTTCCCTGA
CACGCCGGATGCGCGAGACCCTCGCATCGCGGGCCTCGACATGACCATCG
TCGCTGGGATCGACCAGGCGCTCGAGGGCGCGCGGGCCTGCGCCGAGGCG
CTCGTTATCGGGTGATCCTCTTGAGTCCTTATCTGGAAGGCGAGGCGCG
GGAGGCGGCGCGATGGATCGCTGACCAAGCTGCGCCCCAGGGGCGAATCG
CGCTCATCGCCGGTGGCGAGA:CCACGGTCACCATCACGGGTCGTGGTTCG
A::GGGGGCCGAAGCCAAGAGCTGGCCTTGGCCTTTGCTCAGCTTCGACG
CGCTGAGCCCTCGGGCAGGCCCTGGGCCCTCTTGAGCGCTGGGACCGATG
GCAGAGATGGGCCACAGACGCCGCGGGCGCGATCATTGACGCGGAGCGC
GTCACCGGGCTTGACGCCATCGACCAGGCCCTCGCAGATAACAACCTCTCA
CCCTGCGCTCGATGCCGTCGGTGCCTCATCAGGACCGGTCCCTCGGGGA
CCAATGTCGGCGACATCGTGACGCTACTTTATGGCGAGTAGCTGCGCGGT
GAGGTGATGTTTTGGAGCTTCCCGAGCGCATTCTCGGGAGCTCTGTGA
GCGGGTAGTAC::G:::CGCGTGGGCACTTGAAGCGCGCGAGGTGGGCGT
TGCTGAGCGCGTCGAGGCGCTCGCTGTCGATCGTCTCGCCTGGCGTCAGC
ACGAGCGCCGCGACGATGCGCTCTCCCCACTCGGGATCCGGTAGCCCGAC
GACGGCCACCTCGGCGACCGCGGGGTCTCTTAGTAGCACGTCTTCGATCT
CCAAGGCGGAGATCTTAAAGCCCCGCTCTTGATGATGTCCGCAGAGCGC
CGTCCAAGGATATGGTGGTAGCCGTCCTCATCGACGCGGGCGACGTCACC
TGTCCGAAACCAGCCCTCTGTGAAGGCGCTTGCCTAGCCTCTGGCTGGC
CCCAGTAGCCGAGCATCAGTCCATCGCCTCGTACCCAGATCTCGCCAGCG
TCCCCAGGCCCGCAGTCCTCGCCCTCGTACGTGACCACCCGGACCTCGTA
GCCAGGGACCGCGAAGCCGACGCTGCCGGCGCGTCGCTACCCGGCATAGG
GGTTGGAGAGGGTGAAGCCGGTCTCGCTCATCCCGTAAACGCTCGAGGATG
CGCTTGCCAGTCAGCCGCTCGAAGGCCTCATGGTCCGACGCGGGGAGCGC
GGCGGACCCCGAGGTAAAGAGGCGCGCGGCGCGCAGGGCCTCTCGGGCTT
GGCCCTCGCCCTTCAGGTGCGCGATGAGGCGCGCGTACATGGTGGGGACG
CCCATGAAGATGCTGCCCCCTGCGGCGACGGCCTCGATGACCCTCTCCGC
GTCAAAGCGAGGGTGAAGCGTGCGCTCATCTGAGTGAGGAGGGTCCCGA
GGACACCCAGACCAAGTCCATGCACGTGGAAGAGCGGCAGAGCGAGCACA
AGGTGATCGTCTGTGACCAATGCCAGAGGGTGGTTCGTCGCGCCGACATT

CGCGCTGAGCGCGTTGAGGCTTAGGGCGACGCCTTTGGAGCGCCCGGTTCG
TCCCAGAGGTGTAGATGAGCATGGCTGGGGTGTGCTCTCCCAGCTCTCC
TCAAAGCCTGCAGGGTCGCCAATGAGGGAGTCCACAGGGACCCATGGCGT
CCCACGCGCCTCCGCGACATCCGCGGGCGGGCGCCCCGTCATCGACCAAGA
CGAGGCTCGCTTGCGAGTCCTCCAGGATGTGGTTCGATCTCATCGGCCTGA
TAGCGGGTGTGATAGGCACATGGATCACGCCGAGCCGGAGGTGAGCCAC
GATGACGCCAGCAACGAGGCTGAGGTCTCGGACTGAACAGCGACCCTCT
CACCAGGCCGTATCCCTTTGGCTTTAAGGGCCTGTGCGACCGACGCGGAG
CGCCGATCAAGGTCGCGATAGCTGAGATCGACGCCCCCGACGGAGAAGGC
AAGGCGATCCGGGCAGGCGTGAAAGCCATCGAAGAGCCGCGTCATCGCCC
AACACGCTCCAGAGCTCGGGTAAGAGGCGCCCTGAAGGGCCGGCTAAAT
GCGCGTCCGCGCGCCCGGAGAGCTCGGTTCGCCCTCGGGGTTGATGTGATG
AGAAAGGCCCCAGAGGCTGCGCATGGAGCGCCAAGGCGACCGGGGGCCA
GACCTGCGCAGAGGTGCCGACGACGAGGCAAACGTCCGCCTCCTCTGCGG
CCTGCTGCGCGCGCTCAAGGATCCCTGGCCAATAGGTCTCCCCGAACCAA
AGAATGTGCGGCCGATGCGCTCGCCGACAGCGCGCATCGCGGCACGGG
GTCTGAGTCTCTGTGCCCTCGGGAAGGGCGCCAAGGTGTGCGAGGCGG
TGCAACGCATGTCAGCGATCGAGCCGTGAAGTTCGACGACCGCTGGGCTG
CCTGCGCGGCGATGGAGGCCGTCGACGTTTTGGGTGCGGACCAAAAAGTCG
CCTATAGCGAGAGGCCATCGCCGCGATGGTGTGCTGCCCGCGTTTGGCG
CGCAATCGCCGTATTGCTCGGCGCGCCATCGGTACCATCGCCACACGCGC
TCGGGGTCTGCAGCGAAGGCCTGTGGGGTTCGCGAGCTCCTCGGCTCTGAA
CCCCCTCCAGAGTCCCCCTGAACCTCGAAAGGTCGGGATGCC:GACTCG
GCGCTTATCCCCGCGCCGCTGACGACGAAGACCCGCTCCGCCCGCCTCAG
TCGGGCGACGCTGCGCTCAAAGGCTTGGCTATCCATGGG:CTCACCTCT
TCAGACCG:CCAGCCTAGCATAGCGCGCCCCACGGTGGCACGGGCTTGGT
ACTGAACGGATGATCCTATAAGATGCCTTCATGAGCACGGAGAAGAGCAT
CCGTCGGGTGCCAACC CGCCGAACCCTTGGCAAGCCTCCGAGCTTGAGT
TTGATGGCATGCCTCCGCCGACGGCGCTGAAGGTCTTTGAGGATCACAGC
CGCCAGATCCTCTCCAAGAACGCGAGCCCCGATCTCCCCTTCCGCTTACG
CCTGAACCCCTATCGTGGGTGCTATCATGGATGCGCCTACTGCTACGCGC
GACCCTCCCATCACTATCTGGACCTTGGCGCTGGGACCGACTTTGAGCGG
AGCATCGTCATTAAGCCTCAGGCTGCTGAGCTGCTTCGACAGGCCTTCGA
TCGCCCTCATGGTCCGGTGATGCCATCCTCTTCTCAGGCAACACCGATT
GCTATCAGCCCCTTGAGCTCAGCTACGGGTTGACCCGCGCCTGCTTAGAG
ATCTGCCGCGACTACAGAAACCAGTGACCATCATCACGAAGTCCACGCT
CATCGAGCGAGACATCGATCTCC:TCAGGGACCTGGATGCGCGCGCAGCG
TGTACCGTGACCGTCAGCCTCCCTTTTATGGATCCGGGGATGAGTCGAGC
TATCGAGCCTTATGTTCCAAGCCCCAGGCGTCGGTTGAAGACCATCAGGC
GTCTGAGTGACGCGGGCCTCAACGTGCGGGTCAGCATCGCGCCGGTCATC
CCGGGGCTCGCCGATGCGCAGATCCCTCAGATCTTAGAGGCCGCCACGA
CGCAGGCGCTCGCGCCGCGAGCTTCGTCTCCTTCGGCTCTCGCCGGTCG
TCAGCGAGGTCTTTGAGGCCCGCTTGAGGGACGCGCTCCCCGAGCGCGCC
GATCGGGTGATGAATCAGCTTCGAGCGTGTGCAATGGAGAGACCTACGA

TCACCGTTTTGGTCGTCGCCTCGGAGGCGTTGGTGCGCGCTACGAGGCGA
TTGAGGCTCTCTTTGTGAGCACGGCGCGTCGCCTCGGCATGCTCTCAAAG
GAAGACCCCTGGAGACTGACACGTTACACGGCCCCATGAGGGGCGGCA
AATCGCGCTCTTATAGGTACGCTCCGTCAAGGGTCGCCATGTCCAAAGCC
GTCAGCCTGAGGGACCCTGCCTTCGCGCTCGAGCGGGCGCTCGCGACCTT
GCTCGCCCCAGGGATGGGGAGCATGACCTTGGACGAGGCGATGAGCCAGG
CGAGCGCCACCTCGTAGGGACTTGCGTCATGTCGCCGCGCGACCTCACCC
AGGGTGCTGTCCTGGCCGATGTCCTCTTTGTCGAAGCCCCGACTGGCGA
GTAAGGCAAAAAGGTCATCCCGTCCGCCTCACAGCGCGAGAGCACACCGT
CGCTGAACGCCTGGG