

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

<b>1. REPORT DATE (DD-MM-YYYY)</b>		<b>2. REPORT TYPE</b> Final Technical		<b>3. DATES COVERED (From - To)</b> 7/1/99 - 9/30/01	
<b>4. TITLE AND SUBTITLE</b> Shotgun Sequencing of Plasmids from Marine Sediments  Bacteria - Genetic Exploration				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b> N00014-99-1-0860	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> Jonathan Eisen, Ph.D.				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>  The Institute for Genomic Research 9712 Medical Center Drive Rockville, MD 20850				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>  01PR06644-00	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> The Institute for Genomic Research 9712 Medical Center Drive Rockville, MD 20850				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b> TIGR	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b>  Unclassified Unlimited					
<b>13. SUPPLEMENTARY NOTES</b>  NONE					
<b>14. ABSTRACT</b>					
20020211 313					
We have generated DNA libraries from 10+ plasmids (some samples contained more than one plasmid) and have performed enough DNA sequencing reactions on these estimated to yield approximately 5x sequence coverage. Sequence assemblies were generated for each of these and gap closure performed so far on five plasmids. These five have been analyzed using TIGR's automated bioinformatics tools. All the results are available internally and to our collaborators through a web based system.					
<b>15. SUBJECT TERMS</b>					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b>
<b>a. REPORT</b>	<b>b. ABSTRACT</b>	<b>c. THIS PAGE</b>			Jonathan Eisen, Ph.D.
U	U	U	SAR	2	<b>19b. TELEPHONE NUMBER (include area code)</b> 301-838-3507

## ANNUAL PROGRESS REPORT

PROJECT TITLE: Shotgun Sequencing of Plasmids from Marine Sediment Bacteria - Genetic Exploration

PRINCIPAL INVESTIGATOR: Jonathan A. Eisen (jeisen@tigr.org)

INSTITUTION: The Institute for Genomic Research

AWARD #: N00014-99-1-0860

REPORTING PERIOD: Final

AWARD PERIOD: 7/1/99 - 9/30/01

OBJECTIVE: To use genome sequencing and analysis methods to characterize a set of cryptic plasmids from marine sediment bacteria.

APPROACH: Plasmids are sequenced using the shotgun strategy, in which plasmids are subcloned into a library of 2000-4000 base pair fragments and then these clones are sequenced randomly. Five-fold shotgun coverage (in which each base pair of the plasmid is sequenced on average 5 times) will be achieved for all plasmids. These shotgun sequences are then used to generate assemblies of plasmid molecules using computer programs that compare each sequence to all others. Gaps will be closed in as many plasmids as is possible given the financial constraints of the program, leading to the generation of complete or nearly complete plasmid assemblies. Gene and genome features of these plasmids are then characterized using a variety of bioinformatics tools.

ACCOMPLISHMENTS We have generated DNA libraries from 10+ plasmids (some samples contained more than one plasmid) and have performed enough DNA sequencing reactions on these estimated to yield approximately 5x sequence coverage. Sequence assemblies were generated for each of these and gap closure performed so far on five plasmids. These five have been analyzed using TIGR's automated bioinformatics tools and results are available internally and to our collaborators on the web.

Greater than 5x coverage was achieved for 9 plasmids. One additional plasmid appears to have been much larger than original estimates and current sequence coverage is only 3-4 fold. The sequences for each plasmid were assembled using TIGR assembler (a program designed at TIGR to assemble shotgun genome sequence data). The resulting assemblies were mostly quite robust and many contained very few gaps. The five plasmids with the best assemblies (least ambiguity and fewest gaps) were then sent to the TIGR closure teams to close as many gaps as possible (closing a gap involves determining the sequence of the region in the gap). All five have now been closed, meaning that we now have the

complete sequence of these plasmids. These five plasmids were then submitted to TIGR's automated annotation process, which involves analysis of gene and genome features including the identification of putative open reading frames, database searches to identify homologs and motifs, and assignment of tentative functions and role categories. All the results are available internally and to our collaborators through a web based system. Analysis of these is continuing with the goal of publishing papers on each or all of the plasmids.

SIGNIFICANCE: The sequence information will be useful for understanding the biology of these plasmids as well as marine plasmids in general. For example, the sequence of the plasmid that is similar to the F plasmid is only the fourth plasmid to be characterized in this plasmid family. The F plasmid is very important to many aspects of the biology of the pathogen *E. coli* and to studies of molecular biology of many species. Therefore this new sequence information will be very helpful in understanding the biology of this family of plasmids. We have also discovered a variety of genes on each plasmid that will be helpful in creating phenotypic screens to study the biology of these plasmids in marine sediment environments.

WORK PLAN (next 12 months): The results are being written up for publication.

PUBLICATIONS, AWARDS AND PATENTS (last 12 months):

None as of yet.