

PLANT GENOME RESEARCH PROGRAM

**Interagency Program Announcement
NSF/DOE/USDA Joint Program**

***Arabidopsis thaliana* Genome Sequencing Project**

National Science Foundation

Department of Energy

U. S. Department of Agriculture

Deadline Date: April 15, 1998

Plant Genome Research Program

Interagency Program Announcement NSF/DOE/USDA Joint Program

***Arabidopsis thaliana* Genome Sequencing Project**

The purpose of this program announcement is to solicit proposals to continue the systematic sequencing of the genome of *Arabidopsis thaliana*. The ultimate goal of this project is to sequence the entire *Arabidopsis* genome within a reasonable time frame, maximally by the year 2000. This program represents a component of the new NSF Plant Genome Research Program. The overall objective of the Plant Genome Research Program is to support research on the structure, organization and function of plant genomes, and to accelerate the acquisition and utilization of new knowledge and innovative technologies that will aid in developing a more complete elucidation of basic biological processes in plants.

BACKGROUND

Rapid, automated sequencing technologies, with related computational advances in analysis and informatics, have transformed the nature of biological research. Complete DNA sequences already exist for several bacteria and for brewer's yeast, and within a few years the entire genome of the flowering plant *Arabidopsis thaliana* will be known. These huge amounts of data challenge the research and development community to understand and utilize this new knowledge effectively. One challenge, among many, is the revelation from complete sequencing of bacteria and yeast that nearly one-third of all the putative genes have no known function. Much about the structure, organization, and function of genomes remains uncharted.

In recognition of this unprecedented scientific opportunity, the National Science and Technology Council, at the request of Congress, established an Interagency Working Group on Plant Genomes (IWGPG) in May 1997. Representatives from the National Science Foundation (NSF), the Department of Energy (DOE), the United States Department of Agriculture (USDA) and the National Institutes of Health (NIH) were charged with developing a long-range, science-based plan for U.S. plant genome research. Subsequently, Congress appropriated funds to the NSF for Fiscal Year (FY) 1998 for "a comprehensive, peer-reviewed plant genome research program." It was further specified that the new NSF program should support research consistent with the IWGPG recommendations. In a preliminary report issued in June 1997, the IWGPG identified the following genome-based goals for advancing plant science with relevance to economically important plants: to support research and technology development in functional genomics; to develop the necessary physical and educational infrastructure to meet the needs of plant genomic research, including informatics tools and publicly available databases for Expressed Sequence Tags [ESTs] for major classes of economically important plants; in coordination with on-going international efforts, to initiate genomic studies of rice as a model plant for crop species in the grass family, such as corn, wheat and sorghum; and to accelerate efforts to complete the

sequencing of the genome of the model flowering plant, *Arabidopsis*.

PURPOSE

The purpose of this program announcement is to solicit proposals to continue at an accelerated rate the systematic sequencing of the genome of *Arabidopsis thaliana*. The ultimate goal of this project is to sequence the entire *Arabidopsis* genome within a reasonable time frame, maximally by the year 2000. It is anticipated that two to four three-year awards will be made in FY 1998, contingent upon the quality of proposals received and the availability of funds. Individual awards are expected to range from \$2.5M to \$6M per year.

INTRODUCTION

The Multinational Coordinated *Arabidopsis thaliana* Genome Research Project was established in 1990 to develop *Arabidopsis thaliana* as an experimental model system for flowering plants. Three groups were selected to initiate the genome sequencing project in 1996. The goals for numbers of bases sequenced which were initially stated by those groups have been met or exceeded. Given the rapid advances in the sequencing efforts and mapping technologies, the continued advances in research and genomic tools for both *Arabidopsis* and other model organisms and the expanded insights into the biology of the plant, it is now estimated that sequencing of the entire *Arabidopsis* genome could be completed by the year 2000. This time frame is based on current genome sequencing technology, the rate of sequencing of the plant genome to date and the available resources.

Recognizing the potential of an *Arabidopsis* genome sequencing effort to contribute to their missions, the Department of Energy (DOE), and the United States Department of Agriculture (USDA) have joined with the National Science Foundation to continue the U.S. *Arabidopsis thaliana* Genome Sequencing Project. This project is coordinated with other ongoing U.S. genome projects, including the human genome research project supported by the National Institutes of Health (NIH) and the DOE, the microbial genome project supported by the DOE and the plant genome project supported by the USDA, in order to minimize duplication of effort and to maximize efficient use of available resources. U.S. efforts to complete the sequence of the *Arabidopsis* genome will continue to be coordinated on an international level with other national and transnational programs.

WHO MAY SUBMIT

Proposals are solicited from a broad community of scientists at U.S. institutions. Consortia of eligible individuals or organizations may apply, but a single organization must accept overall management responsibility. The eligibility of applicants will be determined according to the guidelines in the "Grant Proposal Guide" (GPG), NSF 98-2, Chapter 1, Section D. The GPG is available on the NSF web site at the URL <http://www.nsf.gov/cgi-bin/getpub?nsf982/>. Proposals from the existing groups for continuation of their sequencing efforts are both expected and encouraged. Involvement of international collaborators is encouraged, although primary support for foreign participants/activities must be secured through their own national programs.

PRINCIPAL INVESTIGATOR AND OTHER SENIOR STAFF

The Principal Investigator (PI) and other senior staff responsible for the project are expected to have expertise and experience in large scale, high through-put, genomic DNA sequencing. If the application is submitted by a consortium of several groups from one or more institutions, the consortium must make a convincing case that it can function in an effective, efficient, timely and cost-conscious manner.

AWARD

The program anticipates supporting two to four three-year awards made as cooperative agreements. The exact amount of the award will depend on the advice of reviewers and on the availability of funds. The cost of the awards will be shared by the participating Federal agencies. It is anticipated that up to \$10 million per year will be made available for this program beginning in FY 1998.

PROPOSAL SUBMISSION

All participating agencies have agreed to use the NSF general guidelines as described in the GPG, NSF 98-2 and the NSF Proposal Forms Kit, NSF 98-3. The GPG, NSF 98-2, is available on the NSF web site at the URL <http://www.nsf.gov/cgi-bin/getpub?nsf982/>. The NSF Proposal Forms Kit, NSF 98-3 is available on the NSF web site at the URL <http://www.nsf.gov/bfa/cpo/gpg/kit/start.htm/>.

Proposals should be prepared following the GPG guidelines and the instructions below. Where the guidelines conflict between the GPG and this program announcement, the latter supersedes the GPG.

Each proposal must contain the following elements in the order indicated:

1. **NSF Cover Page (NSF Form 1207).** Clearly indicate that the proposal is submitted for consideration by the *Arabidopsis* Program in response to this program announcement (NSF 98-52) in the appropriate box.
2. **Project Summary.**
3. **Table of Contents (NSF Form 1359).**
4. **Project Description.** A description of the project must not exceed 25 pages inclusive of tables, diagrams and other visual material. The following points must be addressed in this section and integrated into the general format described in GPG, NSF 98-2.
 - **Progress Report** - Groups which have been supported to initiate sequencing of the genome should include the grant title and number, its funding agency and for each

grant a progress report which lists the number of bases sequenced and deposited in Genbank as finished sequences, the number of sequences which are in the process of being finished, the annotation which has been and will be provided, the level of accuracy achieved and how that determination was made.

- **Sequencing Strategies** - The proposal should include descriptions of:
 1. **DNA substrates to be sequenced:** Include source of the DNA (clones), map of the chromosomal region involved, the method of preparation and all other pertinent information. The selection strategies proposed must be applicable to efforts to sequence the entire *Arabidopsis* genome by the year 2000, and be justified on that basis.
 2. **Sequence quality and quantity:** The level of accuracy to be sought and how that will be measured as well as the numerical goal should be discussed. The numerical goal, defined as the number of bases to be sequenced per unit time, should be linked with a discussion of the finishing process and how that will be defined. Will gaps be closed? Will annotation be completed and what is meant by "complete annotation"? It is expected that the combined international effort will result in a complete sequence of the *Arabidopsis* genome at a minimal accuracy level of 1/10,000 bases by the year 2000. The submitted proposal should indicate the anticipated contribution to that overall goal.
 3. **Genome sequencing technologies and strategies:** Technologies/strategies that will be used should be described as well as plans for incorporating new developments and/or improvements in sequencing protocols, strategies and technologies as they become available. What is the goal for the applicant and what is the expected rate of sequencing which will achieve that goal? How will the applicants coordinate their efforts with other groups, both U.S. and international groups, sequencing other regions of the *Arabidopsis* genome?
 4. **Costs of production sequencing in relation to the product proposed:** The cost-effectiveness of the sequences generated will be a very important issue. An estimate of the dollars required to produce a specific number of bases (which should include the costs of generating clones, assembly and annotation) should be given. If investigators are proposing a strategy that will yield less than the complete genome sequence, they must provide an overall vision of how this strategy will contribute to the cost-effective completion of the entire *Arabidopsis* genome by the year 2000 or earlier.

- **Project Management** - The proposal should include descriptions of:
 1. Plans for establishing coordination with other existing or planned *Arabidopsis* sequencing projects, both nationally and internationally.
 2. Plans for establishing a close linkage to the plant biology research community at large in order to ensure a close collaboration between the sequencing project and the ultimate user community of the sequence information.
 3. Ways to assess progress of the project, including establishing milestones and

measuring progress toward them. A common advisory committee will be appointed based upon suggestions from all of the participants, including the agencies, which will serve as a means of advising all participants of problems or solutions which will benefit all of the participants. Details will be included in the Cooperative Agreements and will be agreed to by all participants prior to activation of the Cooperative Agreements.

4. Available facilities including a statement of institutional commitment for the successful completion of the project.

- **Information Management** - The proposal should include:

1. Data management plan, including: (1) mechanisms to assess validity and accuracy of data obtained which will augment or complement procedures to monitor accuracy which may be mandated by the agencies; (2) mechanisms for annotation of data and release of both raw and finished data into public databases -- creative, cost-effective strategies for annotating sequences are encouraged; and (3) community access to data mechanisms of data distribution and interactions with other community databases.
2. Data release policies including how rapidly sequence data will be publicly released after production. The sponsoring agencies encourage an immediate and continuous release of clearly labeled raw as well as finished genome sequence data as it is obtained by the groups supported by this Program.
3. A statement signed by an authorized institutional official should be included which clearly describes the institutional policy for sharing information and materials resulting from this work with other researchers of the community of scientists.

5. **References cited.** Citations must be complete and arranged in alphabetical order by author.

6. **Budget (NSF Form 1030).** Provide a budget for each year of support requested as well as a summary budget for all three years. If there are subcontracts, signed separate budget(s) must be included for each subcontractor. Funds for facility construction or renovation may not be requested.

7. **Budget justification.** A brief explanation for funds in each budget category should be provided. For major equipment or software materials, a particular model or source and the current or expected price should be specified whenever possible. A brief explanation of the need for each item whose costs exceeds \$10,000 should be provided. This section should also include a summary of institutional cost sharing, if any, and other sources of support for the project, such as government, industry, or private foundations. Appropriate documentation of any such commitments should be provided in the appendix. Although cost-sharing is not required, any such commitment specified in the proposal will be made a condition of an award, and the dollar amount of the proposed cost sharing must be indicated on line M of the Budget Form 1030.

8. **Facilities and Equipment and Other Resources (NSF Form 1363).** Include a brief description of available facilities, including space and relevant equipment available for the project. Where requested equipment or materials duplicate existing items, explain the need for duplication. This section is limited to 2 pages.
9. **Biographical Sketches.** For each of the key personnel, provide a curriculum vitae or short biographical sketch. Briefly describe relevant experience, list up to 10 publications (to include the individual's 5 most important and up to 5 other, relevant publications). The section is limited to 2 pages for each individual.
10. **Current and Pending Support (NSF Form 1239).** Provide a complete list of current and pending research support for each of the key personnel.
11. **Appendices.** Those individuals, institutions, or programs that are participating in the project in a significant way, but have not endorsed the cover page as official co-PIs, must submit letters describing the nature of their collaboration and commitment to the proposed project. They can be included as appendices. General letters of endorsement may not be included.
12. **Additional Information.** The following items must be provided:
 - One complete copy of NSF Form 1225 (Information about Principal Investigators/Project Directors).
 - A list of current and past collaborators of senior personnel whose CVs are included, and of other staff mentioned by name in the proposal. This should include names of their graduate and postdoctoral mentors, all graduate students and postdoctoral fellows who have trained with them, all co-authors of papers within the last 4 years, and individuals who have served as paid consultants to the investigators named in the proposal.

Attach these items to the copy of the proposal that bears the original signatures, with the Form 1225 and the lists of collaborators on top. These items are for the agency's internal use only and will not be available to reviewers. Do not provide additional copies of these items with the other proposal copies.

WHEN AND WHERE TO SUBMIT

**PROPOSALS MUST BE RECEIVED AT NSF BY 5:00 P.M. EASTERN TIME (ET)
APRIL 15, 1998.**

You are encouraged to use NSF FastLane to prepare and submit your *Arabidopsis thaliana* Genome Sequencing Project. To access FastLane, go to the NSF Web Site at the URL <http://www.nsf.gov>, then select "FastLane" or go directly to the FastLane Home page located at <http://www.fastlane.nsf.gov/>.

For proposals not submitted via FastLane the completed application, including the original proposal and 18 copies should be sent to (see the GPG, Chapter I, Section E):

Program Announcement Number: NSF 98-52
National Science Foundation PPU
Arabidopsis thaliana Genome Sequencing Project
4201 Wilson Boulevard, Room P60
Arlington, VA 22230

INSTRUCTIONS FOR SUBMISSION OF *ARABIDOPSIS THALIANA* GENOME SEQUENCING PROJECT PROPOSALS USING NSF FASTLANE

- If you are using FastLane to prepare your *Arabidopsis* proposal it **must be received electronically no later than 5:00 p.m. ET April 15, 1998**. The signed Cover Sheet and signed Certification Page must be received at NSF no later than 5 p.m. ET April 22, 1998. To access FastLane, go to the NSF Web Site at the URL <http://www.nsf.gov>, then select "FastLane" or go directly to the FastLane Home page located at <http://www.fastlane.nsf.gov/>.
- For **instructions to prepare and submit** your proposal via Fastlane please see Instructions for Preparing and Submitting a Standard Proposal via FastLane located at <http://www.fastlane.nsf.gov/a1/newstan.htm>. Additionally, read the PI Tipsheet for Proposal Preparation, and the Frequently Asked Questions about FastLane Proposal Preparation. The information can be found under Proposal Preparation on the FastLane website at <http://www.fastlane.nsf.gov/>.
- **Project Description.** The Project Description should include the following sections, appropriately labeled in the following order, as specified in the Proposal Submission section of this announcement:
 1. Progress Report
 2. Sequencing Strategies
 3. Project Management
 4. Information Management
 5. List of current and past collaborators (Note: This list does not count against the 25 page limit of the Project Description.)
- **Appendices.** Do not include appendices as part of the FastLane submission. Instead, **mail 18 copies of appendices** with the **one required signed cover sheet and certification form** to be received by no later than 5:00 p.m. ET April 22, 1998. Mail to: Division of Biological Infrastructure, National Science Foundation, 4201 Wilson Boulevard, Room 615, Arlington, VA 22230. Label package "Appendices – Announcement 98-52." It is the proposer's responsibility to ensure that materials are received by the deadline.

EVALUATION OF PROPOSALS

Selection of awards will be based on merit review by experts using established peer review systems as described in GPG, Chapter III. A special emphasis panel will be formed to review the applications and site visits may be used as needed. All participating agencies use similar general criteria in the evaluation of proposals submitted to their respective competitive research grants programs. The new NSF review criteria as outlined in the GPG (NSF 98-2) will be interpreted in light of the objective of this solicitation as follows:

- 1. Performance competence:** This criterion addresses the technical soundness of the proposed approach, the capabilities of the proposed personnel, including those of the PI and other senior staff as discussed above, the adequacy of the resources available or proposed, and the likelihood that this project will lead to a successful, timely, cost-effective completion of *Arabidopsis* genome sequencing by the year 2000.
- 2. Project management:** This criterion addresses the overall quality of the technical and managerial aspects of the proposal, including plans for the release of the data and the sharing of the information and resources resulting from the project to the scientific community as noted below, and for management oversight and long-range planning.
- 3. Effect of the activity on the scientific infrastructure:** This criterion addresses the potential of the proposed activity to contribute to better understanding or improvement of the quality and effectiveness of the Nation's scientific research, education, and human resources capabilities. An important issue is a likelihood of national impact and widespread, appropriate dissemination and use of results in strengthening the scientific infrastructure of this nation.
- 4. Scientific collaboration and information sharing:** Sequencing of the genome of a model organism is a community activity. As such, a close collaboration among the scientists and organizations involved in sequencing activities and effective dissemination to the users of the information (the scientific community) are important components of this criterion.

AWARD ADMINISTRATION

The *Arabidopsis thaliana* Genome Sequencing Project will be administered and managed as an interagency program involving all participating agencies throughout the entire process from the development of the program announcement to the review, selection and administration of the awards. Awards will be administered in accordance with the terms and conditions of NSF GC-1, "Grant General Conditions," (12/97) and NSF CA-1, "Cooperative Agreements General Conditions" (12/95). This information can be obtained from the NSF OnLine Document System (<http://www.nsf.gov/>). Copies of these documents are available at no cost from the NSF Clearinghouse, P.O. Box 218, Jessup, MD 20794-0218, phone (301) 947-2722, or via e-mail at pubs@nsf.gov. More comprehensive information is contained in the NSF Grant Policy Manual

(NSF 95-26), for sale through the Superintendent of Documents, Government Printing Office (GPO), Washington, D.C. 20402. The telephone number at GPO is (202) 783-3238 for subscription information. The NSF Grant Policy Manual can also be accessed online at the above URL.

OTHER INFORMATION

Potential applicants are strongly encouraged to contact project officers and discuss their plans. Inquiries regarding the announcement can be directed to any one of the following agency representatives:

DeLill Nasser
Division of Molecular and Cellular Biosciences
National Science Foundation
4201 Wilson Boulevard
Arlington, VA 22230
Phone: (703) 306-1439
Fax: (703) 306-0355
E-mail: dnasser@nsf.gov

Edward K. Kaleikau
Plant Genome Program
NRJ Competitive Grants Program
CSREES, United States Department of Agriculture
STOP 2241, Washington, D.C. 20250-2241
Phone: (202) 401-1901
Fax: (202) 401-6488
E-mail: ekaleikau@reeusda.gov

Gregory L. Dilworth
Division of Energy Biosciences
Office of Basic Energy Sciences
United States Department of Energy
ER - 17, GTN, Washington, D.C. 20585
Phone: (301) 903-2873
Fax: (301) 903-1003
E-mail: greg.dilworth@mailgw.er.doe.gov

Henry Shands
Agricultural Research Service
United States Department of Agriculture
Jamie L. Whitten Federal Building, Room 319-A
Washington, D.C. 20250-0300

Phone: (202) 205-7835
Fax: (202) 690-1434
E-mail: shands@ars-grin.gov

The Foundation provides awards for research in the sciences and engineering. The awardee is wholly responsible for the conduct of such research and preparation of the results for publication. The Foundation, therefore, does not assume responsibility for such findings or their interpretation.

The Foundation welcomes proposals on behalf of all qualified scientists and engineers, and strongly encourages women, minorities, and persons with disabilities to compete fully in any of the research and research-related Programs described in this document. In accordance with Federal statutes and regulations and NSF policies, no person on grounds of race, color, age, sex, national origin, or disability shall be excluded from participation in, denied the benefits of, or be subject to discrimination under any program or activity receiving financial assistance from the National Science Foundation.

Facilitation Awards for Scientists and Engineers with Disabilities provides funding for special assistance or equipment to enable persons with disabilities (investigators and other staff; including student research assistants) to work on NSF projects. See the program announcement or contact the program coordinator at (703) 306-1636.

Privacy Act. The information requested on proposal forms is solicited under the authority of the National Science Foundation Act of 1950, as amended. It will be used in connection with the selection of qualified proposals and may be disclosed to qualified reviewers and staff assistants as part of the review process; to applicant institutions/grantees; to provide or obtain data regarding the application review process, award decisions or the administration of awards; to government contractors, experts, volunteers and researchers as necessary to complete assigned work; and to other government agencies in order to coordinate programs. See Systems of Records, NSF 50, "Principal Investigators/Proposal File and Associated Records," and NSF 51, 60 Federal Register 4449 (January 23, 1995), "Reviewer/Proposal File and Associated Records," 59 Federal Register 8031 (February 17, 1994).

Public Burden Statement. Submission of the information is voluntary. Failure to provide full and complete information, however, may reduce the possibility of your receiving an award.

The public reporting burden for this collection of information is estimated to average 120 hours per response, including the time for reviewing instructions. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Gail A. McHenry, Reports Clearance Officer, Information Dissemination Branch, National Science Foundation, 4201 Wilson Boulevard, Suite 245, Arlington, VA 22230.

NSF has TDD (Telephonic Device for the Deaf) capability which enables individuals with hearing impairments to communicate with the Foundation about NSF programs, employment, or general information. To access NSF TDD, dial (703) 306-0090; for FIRS,

1-800-877-8339.

The program described in this publication is in the Catalog of Federal Domestic Assistance:

NSF - Number 47.074, Biological Sciences

DOE - Number 81.049, Basic Energy Sciences, High Energy and Nuclear Physics,
Magnetic Fusion Energy, Health and Environmental Research, Program
Analysis, Field Operations Management.

USDA - Number 10.206, Competitive Research Grants Program

OMB# 3145-0058

P.T. 34

K.W. 1002037

NSF 98-52 (*Electronic Dissemination Only*)

(Replaces NSF 95-159)