OMB No. 0925-0001/0002 (Rev. 08/12 Approved Through 8/31/2015)

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Dennis Shasha

eRA COMMONS USER NAME (credential, e.g., agency login): DES1.NYU

POSITION TITLE: Professor of Computer Science

EDUCATION/TRAINING (Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)

| INSTITUTION AND LOCATION | DEGREE(if applicable) | Completion DateMM/YYYY | FIELD OF STUDY |
| --- | --- | --- | --- |
| Yale UniversitySyracuse UniversityHarvard University | B.S.M.S.Ph.D. | 197719801984 | Engineering and Applied ScienceComputer ScienceApplied Mathematics |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| **A. Personal Statement** |  |  |  |

Briefly describe why you are well-suited for your role(s) in the project described in this application. The relevant factors may include aspects of your training; your previous experimental work on this specific topic or related topics; your technical expertise; your collaborators or scientific environment; and your past performance in this or related fields (you may mention specific contributions to science that are not included in Section C). Also, you may identify up to four peer reviewed publications that specifically highlight your experience and qualifications for this project. If you wish to explain impediments to your past productivity, you may include a description of factors such as family care responsibilities, illness, disability, and active duty military service.

Dennis Shasha is a professor of computer science at the Courant Institute of New York University where he works with biologists on pattern discovery for network inference; with computational chemists on algorithms for protein design; and on computational reproducibility. Other areas of interest include database tuning as well as tree and graph matching. He has co-authored over seventy journal papers, seventy conference papers, and twenty patents. Because he likes to type, he has written six books of puzzles about a mathematical detective named Dr. Ecco, a biography about great computer scientists, and a book about the future of computing. He has also written four technical books about database tuning, network inference, time series, DNA computing, and statistics. The work in this proposal directly uses the work on graph matching, motif finding, and machine learning.

B. Positions and Honors

List in chronological order previous positions, concluding with the present position. List any honors. Include present membership on any Federal Government public advisory committee.

1984 - 1990 Assistant Professor, Courant Institute of Mathematical Sciences, New York University.

1990 - 1995 Associate Professor, Courant Institute of Mathematical Sciences, New York University.

1995 - Present Full Professor, Courant Institute of Mathematical Sciences, New York University.

1992 - 1993 Invited Professor, INRIA, Roquencourt, France.

1998 - 1999 Invited Professor, INRIA, Roquencourt, France.

2006 - 2007 Invited Professor, INRIA, Roquencourt, France.

1993 - Present Co-Editor-in-Chief (with Gottfried Vossen) of Information Systems.

Publisher: Elsevier North Holland

2003-Present: Distinguished Science Advisor, New York Hall of Science

2014: Fellow of the ACM (computing society)

2015: INRIA International Chair

**C. Contributions to Science \*\*\* PLEASE COMPLETE\*\*\***

Briefly describe up to five of your most significant contributions to science. For each contribution, indicate the historical background that frames the scientific problem; the central finding(s); the influence of the finding(s) on the progress of science or the application of those finding(s) to health or technology; and your specific role in the described work. For each of these contributions, reference up to four peer-reviewed publications or other non-publication research products (can include audio or video products; patents; data and research materials; databases; educational aids or curricula; instruments or equipment; models; protocols; and software or netware) that are relevant to the described contribution. The description of each contribution should be no longer than one half page including figures and citations. Also provide a URL to a full list of your published work as found in a publicly available digital database such as SciENcv or My Bibliography, which are maintained by the US National Library of Medicine.

 Here is a list of all my publications: http://cs.nyu.edu/cs/faculty/shasha/papers/resume.pdf

1. Understanding causality in transcriptional networks when there are few data points. The goal of this line of work is to find out which genes/proteins are influencing other genes and proteins and how. When there are few data points (e.g. short time series), this problem is very underdetermined. For this reason, it is important to make use of simple models (linear and perhaps quadratic) and find methods that are robust to noise. Our work in that area includes the following papers:

i) Krouk G, Lingeman J, Colon AM, Coruzzi G, Shasha D. Gene regulatory networks in plants: learning causality from time and perturbation. **Genome Biol.** 2013;14(6):123. PMCID: 3707030.

ii) Krouk G, Mirowski P, LeCun Y, Shasha DE, Coruzzi GM. Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate. **Genome Biol.** 2010;11(12):R123. PMCID: 3046483.

*iii) Network Inference in Molecular Biology -- a hands-on framework*

Jesse Lingeman and Dennis Shasha,

**Springer Verlag**, 2012, 109 pages, ISBN 978-1461431121.

iv) Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, et al. A gene expression map of the Arabidopsis root. **Science**. 2003;302(5652):1956-60.

1. Querying graphs for topological pattens and finding over-represented patterns in networks. The goal of this work is to locate portions of graphs that reflect known patterns or to describe new patterns. Because most interesting versions of this problem are NP-Complete, performance is a main concern.
2. Di Natale R, Ferro A, Giugno R, Mongiovi M, Pulvirenti A, Shasha D. SING: subgraph search in non-homogeneous graphs. . BMC Bioinformatics. 2010;11:96. PMCID: 2850364.
3. Giugno R, Bonnici V, Bombieri N, Pulvirenti A, Ferro A, Shasha D. GRAPES: a software for parallel searching on Biological Graphs Targeting Multi-Core Architectures. **PLOS ONE.**
4. A. Ferro, R. Giugno, M. Mongiovi, A. Pulvirenti, D. Skripin, D. Shasha, "GraphFind: Enhancing Graph Searching by Low Support Data Mining Techniques" **BMC Bioinformatics**, vol. 8 ISSN: 1471-2105, 2007.
5. A. Ferro, R. Giugno, G. Pigola, A. Pulvirenti, D. Skripin, G. D. Bader, D. Shasha "NetMatch: a Cytoscape Plugin for Searching Biological Networks" **Bioinformatics**, 2007 23(7):910-912; doi:10.1093/bioinformatics/btm032

**D. Research Support**

List both selected ongoing and completed research projects for the past three years (Federal or non-Federally-supported). Begin with the projects that are most relevant to the research proposed in the application. Briefly indicate the overall goals of the projects and responsibilities of the key person identified on the Biographical Sketch. Do not include number of person months or direct costs.

Ongoing Research Support

MCB-1158273 Coruzzi (PI) 04/01/12 - 03/31/16 (in a NCE)

A Systems Approach to the NPK Nutriome and its Effect on Biomass

This proposal explores the molecular underpinnings of the nitrate, phosphate, potassium (NPK) effect and tests the hypothesis that the enhancement of biomass under low-N conditions is the result of NPK interactions at the *signaling* level.

Role: Co-Investigator

IOS-1339362 Coruzzi (PI) 09/01/2014 – 08/31/2018

NutriNet: A network inspired approach to improving Nutrient Use Efficiency (NUE) in crop plants

\*\*Brief Description Needed\*\*

MCB-1412232 Coruzzi (PI) 07/01/2014 – 06/30/2017

Prospecting for Resources: A systems integration of local and systemic Nutrient signaling

\*\*Brief Description Needed\*\*

N/A Shasha (PI) 05/01/2012 – 08/31/2017

WICAT Center and NYU WIRELESS Research Distribution Award

The goal of this work is to provide a platform for querying and analyzing power-delay experiments for millimeter wave wireless communication. This will continue to the design of better layout for wireless base stations.

Role: PI

MCB-1355462 Christine Vogel (PI) 11/15/13 – 10/31/16

EAGER: Modeling Protein Degradation - Evaluation of Strategies and Targets

This project attempts to distinguish both experimentally and computationally among different mechanisms for protein degradation. The computational part will primarily use stochastic gradient descent.

Role: Co-Investigator

**Completed Research Support**

5R01 GM032877-28 Coruzzi (PI) 05/01/09 - 04/30/15

A Systems Approach to Regulatory Networks Controlling N-assimilation

Our goal is to develop Systems Biology approaches to identify transcriptional networks regulating N-assimilation into amino acids, as a model metabolic regulatory network in plants.

Role: Co-Investigator

MCB-0929338 Coruzzi (PI) 07/15/09 - 06/30/15

Arabidopsis 2010: Nitrogen Networks in Plants

This grant concerns the mechanisms by which nitrogen signaling mediates genome-wide changes affecting plant growth and development, with a special emphasis on root development explored using ecotypes.

Role: Co-Investigator

IOS-0922738 Coruzzi (PI) 08/01/10 - 07/31/14

Genomics of Comparative Seed Evolution

The objectives of this project are to develop datasets (deep-transcriptome), bioinformatic resources, and informatic pipelines that will enable functional trait-to-gene predictions based on genome-wide phylogenies and/or machine learning approaches. This project largely focuses on plants from the Gymnosperm clade of the seed plants and is performed is in collaboration with the NYBG, AMNH and CSHL.

Role: Co-Investigator

N/A Shasha (PI) 09/01/2013 – 12/31/15

Building reliable genetic devices using unreliable ones

The goal of this work has shifted away from the bench. Now the problem is to determine how to adjust machine learning problems when the context/environment changes. When that happens previous prediction methods may no longer work.

Role: PI