**TIMELINE: Year 1:** Aim 1. Implement InferNet Inference using a variety of machine learning methods. Cross-validate on the expression experiments from the 21 species (Fig. X). Aim 2A. Test methods for trait-network-gene using seed data and mutants. **Years 2-3:** Aim 2B. Apply trait-network-gene approach to nitrogen regulatory networks. Select nodes for validation in transient protoplast assays. Aim 3. Deploy the first version of the X-Net software platform to collaborators (R. Gutierrez, Chile), and other beta-testers.

**PLAN TO INTEGRATE RESEARCH AND EDUCATION**:

**Cross-training of Biologists and Computer Scientist in Systems Biology**. We accomplish this goal in several ways. **R-boot camp**: Each year, Dr. Katari (a computer scientist with a PhD in Genetics) teaches an R-boot camp (weekly during the Fall semester), to train the Biologists in using R to analyze genomic data. This consists of a group of biologists at all levels from students to professors. Recent “students” have included faculty on sabbatical Mary Lou Guerinot and Rob McClung of Dartmouth. As a complement, Computer Scientists are taught Biology through a Molecular & Cell Biology Class (taught by Dr. Coruzzi) and during the weekly joint lab meetings between the Coruzzi Lab ( NYU Biology) and NYU Courant (Shasha and Tranchina Labs). This is in additional to formal courses taught at NYU’s Center for Genomics & Systems Biology: G23.1128 Systems Biology; G23.1130 Applied Genomics & Network Modeling; G23.1127 Bioinformatics & Genomes. Graduate students are co-advised by a Biology and Computer Science faculty: In the last year, we have trained two PhD students, two interns and two MS students from Courant in this environment. For a complete listing of students trained in the past 4.5 years, see Education and Training section in Results from Prior support. Computational students will be involved in constructing the pipeline and making it perform through the use of parallelization. Such students will also help develop and optimize machine-learning algorithms for network inference. [Don’t forget your intel winner]

**PLAN TO INTEGRATE DIVERSITY**: We are committed to training scientists at the graduate and postdoctoral levels across diversity.  Students trained as part of the parent NSF grant include Hispanic and African-American scientists. Damion Nero an African American PhD student, has written programs contributing to the Virtual Plant project. Roberto Jimenez (Systems Admin) associated with this project is of Hispanic origin as is our collaborator Rodrigo Gutierrez. Unusual for a computational grant, we have numerous female scientists are associated with this project: Coruzzi (co-PI); Rebecca Davidson (Programmer); Varuni Prabhakar (UG Programmer); Ana Arroja (MS); Ranjita Iyer (MS Courant).

**SHARING OF RESULTS**: **Publications:** The results of our analysis of the data we generate will be made available through peer- reviewed literature as it is the most appropriate way to make this information available.

**MANAGEMENT PLAN**: To coordinate and facilitate interactions between individuals, Dennis Shasha, the [Gloria: the project summary currently has you as the PI] PI (NYU Computer Science) will also serve as the overall Project Manager. Gloria Coruzzi (NYU Biology) will serve as a biological advisor and conduit to a working lab and the wider plant community.

**Bioinformatics manager: Dr.** **Manpreet Katari** (NYU Biology) will be in charge of the bioinformatics data. To enable efficient information exchange of raw and processed data, a file server has been set up at the NYU to store and distribute data and its analysis among users at NYU Biology and NYU Courant. This will be maintained by **Dr. Roberto Jimenez**, the Systems Administrator for this project, who will also maintain the web server, database server, and update the multinetwork databases.

**Senior Programmer: Dr. Arthur Goldberg** (NYU Courant, current affiliation- Memorial Sloan Kettering) will manage the development of new software analysis tools and pipelines to enable Neighborly Network Inference (NNI) which will support the different species and inference, and also new pipelines for cross species analysis, especially as they relate to crop species in coordination with the PI, and a computer science doctoral student.

**Principal Investigators:** Shasha and Coruzzi will each supervise personnel, organization, intellectual developments and contributions.

**Role of Participants:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Institution** | **Role** |
| ***Dennis Shasha***-PI | NYU Courant | Project Leader: Computational |
| ***Gloria Coruzzi***-Co-PI | NYU Biology | Co-leader: Biological |
| ***Manpreet Katari***-Co-PI | NYU Biology | Bioinformatics Manager |
| ***Arthur Goldberg***-Senior Programmer | NYU Courant | Senior Programmer |
| ***Rodrigo Gutierrez***-Consultant | UCatolica Chile | Assembling validated networks for target species  |

**COORDINATION WITH OUTSIDE GROUPS:**

**Please see attached letter of collaboration:**

**Rodrigo Gutierrez (U Catolica, Chile)** Dr. Gutierrez, the creator of the Arabidopsis multinetwork (Gutierrez et al 2007), will assist in the assembly of multi-networks for crop species in the list of 21 species including Vitis (Grape), Corn, and Medicago.