December 5, 2005

Bioinformatics Search Committee

Dear Search Committee,

Dr. Jiajian (JJ) Liu asked that I send you a letter of recommendation regarding the bioinformatics position that he has applied for. I have known JJ for about four years, during which time he has been a post-doc in my group. He came to my group with a background and a Ph.D. degree in microbiology, but with a desire to become proficient in bioinformatics. At the time he had some programming experience but not really any in bioinformatics. However, he had very strong letters of recommendation so I decided to give him a chance in my lab, especially since I was interested in someone who was willing to be involved in both experimental and computational research, and JJ was happy to do both things.

I have been quite happy with his performance in both areas. We have now published three papers and another paper is nearly completed that covers his work on protein-DNA interactions using neural networks. The four papers are quite different from each other, and indicate his diverse expertise. The first, published at the end of 2003 (Nucl Acids Res 31:6891-6903), he initiated on his own, based on his prior interest and knowledge about sporulation in B. subtilis. Given the recent genome sequences of several Bacilis species and microarry data from B. subtilis related to gene expression during sporulation, comparing wild type and mutant profiles, JJ decided there was enough new information to try and fill out the regulon for the spo0A gene. He was helped a bit by a graduate student, Kai Tan, but JJ did the majority of the analysis. A paper from Rich Losick's group came out almost simultaneously with ours that used an experimental approach to get more genes in the spo0A regulon, and it was gratifying to see the large overlap with our predicted set. I think this is a testament to his insight in how to combine such diverse data sets to learn more about Bacilis biology, and with his long experience in that field I think it also indicates the kind of work that he intends to continue doing in a combined microbiology and computational biology endeavor.

His second paper (BMC Bioinfomratics, 6:176) is about DNA-protein interactions measured in vitro with a new method that we've developed that can give very accurate quantitative measurements of affinity to a variety of different DNA sequences simultaneously. The original description of the method was published just before JJ entered the lab (Nucl Acids Res, 29: 2471-2478), but he really applied it in an effective way on a larger problem. Using the zinc-finger protein EGR (zif268) he made five variants with one or two mutations in one of the zinc fingers. For each of the six proteins he measured quantitatively their affinity to six different DNA targets, to get a complete 6x6 table of affinities for all combinations. The goal was to ask about the inter-dependence of mutations, in both the DNA and the protein, on the affinity. This is an important issue, especially for the computational modeling of such binding sites and the prediction of regulated genes in genomic studies. This is the first analysis I know of with such a thorough analysis of the question, and the results are interesting. While there is clearly some inter-dependence of the interactions, most of them can be well approximated by an additive model. But there are a couple of special cases where that approximation works considerably worse, but we show that a fairly minor modifcation of the model fits the data much better. The third paper related to this (NAR 33:e141) shows how a combined experimental

approach, using SELEX and quantitative binding assays, can obtain very accurate models of protein-DNA affinities in a fairly high throughput manner. We think this approach will be very useful in obtaining large, quantitative data sets for use in genomic analysis of transcription factors and their cognate binding sites.

A fourth paper that we are close to submitting is a computational approach to take advantage of the inter-dependence I described above. In those cases where additivity does not hold well, one needs more complex models than those typically used to identify important DNA-protein interactions. His program will provide a method to solve that problem This emphasizes his feeling, which I fully endorse, that computational approaches to studying biology should be firmly grounded in experimental data, and the groups closest to the data are in the best position to develop computational methods that exploit it.

In the last few months JJ has also begun work on a DOE sponsored grant that we were just awarded to study gene regulation in Shewanella. This is a collaboration with Dr. Chip Lawrence where we will be working to figure out the regulatory connections between the transcription factors and the genes they regulate.

In summary, JJ has been a very good post-doc and I think he will make a good research scientist in a more independent position. He is also good at working with people and so can be an effective member of a research team. He has shown the ability to learn a lot in a short time and a very strong motivation to succeed. He has been a very good colleague in the lab, helping several other post-docs and students with their projects. He is now very interested in taking his new skills in bioinformatics analysis and applying them to interesting problems in microbiology, his original interest. I think he has the appropriate expertise to be successful in this field. I think you will find him a bright and amiable colleague and I recommend him highly. If you have any questions that I can answer, feel free to contact me at the email address or phone number below.

Sincerely,

Gary D. Stormo
Professor, Department of Genetics
Director, Computational Biology
Program
stormo@genetics.wustl.edu
(314)747-5534