

## Genome-wide expression analysis of genetic networks in *Neurospora crassa*

David A. Logan<sup>1</sup>, Allison L. Koch<sup>2</sup>, Wubei Dong<sup>2</sup>, James Griffith<sup>2,4</sup>, Roger Nilsen<sup>2</sup>, Mary E. Case<sup>2</sup>, Heinz-Bernd Schüttler<sup>3</sup> and Jonathan Arnold<sup>2\*</sup>

<sup>1</sup>Department of Biological Sciences, Clark Atlanta University, Atlanta, GA - 30314; <sup>2</sup>Department of Genetics, University of Georgia, Athens, GA -30602; <sup>3</sup>Department of Physics and Astronomy, University of Georgia, Athens, GA - 30602; <sup>4</sup>College of Agricultural and Environmental Sciences, University of Georgia, Athens, GA - 30602;

Jonathan Arnold\* - E-mail: arnold@uga.edu; Phone: +706 542 1449; Fax: +706 542 3910; \* Corresponding author

received December 09, 2006; accepted January 20, 2007; published online April 10, 2007

### Abstract:

The products of five structural genes and two regulatory genes of the *qa* gene cluster of *Neurospora crassa* control the metabolism of quinic acid (QA) as a carbon source. A detailed genetic network model of this metabolic process has been reported. This investigation is designed to expand the current model of the QA reaction network. The ensemble method of network identification was used to model RNA profiling data on the *qa* gene cluster. Through microarray and cluster analysis, genome-wide identification of RNA transcripts associated with quinic acid metabolism in *N. crassa* is described and suggests a connection to other metabolic circuits. More than 100 genes whose products include carbon metabolism, protein degradation and modification, amino acid metabolism and ribosome synthesis appear to be connected to quinic acid metabolism. The core of the *qa* gene cluster network is validated with respect to RNA profiling data obtained from microarrays.

**Keywords:** genetic networks; quinic acid; *qa* gene cluster; genome; microarray

### Background:

Advances in systems biology and the relative ease of manipulating eukaryotic microorganisms such as fungi have made it possible to elucidate a complex trait such as how a cell metabolizes a carbon source. The genetics of the metabolism of the carbon compound quinic acid in *Neurospora crassa* is one of the early paradigms for eukaryotic gene regulation with the support of over four decades of genetics and biochemistry. [1] A genetic network (Figure 1) has been reported for the quinic acid gene cluster of *N. crassa*. [2] The network involves 54 reactions and 38 macromolecules and small molecules and is based on a kinetics framework [3] and experimental data generated over a span of more than three decades. [4] In the diagram, circles, boxes and arrows depict, reactions, molecular species, and the flow of reactants and products for the reaction proceeding "forward". Reactions without outgoing arrows (lollipops), indicate decay reactions. Reactants include the seven genes of the *qa* gene cluster (*qa-x*, *qa-2*, *qa-3*, *qa-4*, *qa-y*, *qa-1S*, *qa-1F*) [1] with superscripts "0" or "1" indicating the presence or absence, respectively, of a bound transcriptional activator, encoded by the *qa-1F* gene. These genes are transcribed into RNA molecules (superscript r), which in turn are translated into protein products (in capital letters). Four out of seven proteins have been demonstrated by experimental means to participate in a known biochemical pathway at the bottom of the diagram. There are at least two cellular states for quinic acid, extracellular (depicted as QA<sup>e</sup>) or intracellular quinic acid (QA). One of the genes, *qa-y*, is thought to encode a carrier QA-Y, which is hypothesized to transport the quinic acid across the cell membrane. There is one hypothesized protein-protein interaction in the model between the repressor, QA-1S, and the transcriptional activator, QA-1F. Quinic acid in the cell is hypothesized to disrupt the bound complex of QA-1S/QA-1F to favor induction by QA-1F. [1]

Since the *aro* and *qa* gene clusters are biochemically coupled, the network outlined here is very likely incomplete. For example, the *qa-2* and *aro-9* genes have the same biochemical function, both being dehydroquinases, which can functionally substitute for each other. [5] Further, the *aro* cluster is known to encode products in the shikimate pathway, which in yeast is reported to be under cross-pathway or general control. [6] This implies that the *aro* and possibly *qa* gene clusters should be hypothesized to be under *GCN4* control. [7] Here we employ the ensemble modeling method [2] to fit RNA profiling data generated from microarray experiments to identify the *qa* cluster. Through the use of average linkage analysis [8] we have examined the relationship of *qa* responsive genes to all of the genes in the genome. This should provide the basis for a refined quinic acid genetic network model.

### Methodology:

*Neurospora crassa* strain 74A-OR23-1A (#987, Fungal Genetics Stock Center (FGSC), Kansas City, KS) was used in all experiments and shifted from sucrose (1.5%) to quinic acid (0.3 %) to identify QA inducible genes by comparison to cultures grown on glucose (2%) over a 48 hour window. The control cultures, FGSC 1858 (*bdA*) grown on glucose (2%), were used because the array data were available. After shifting the cultures for the appropriate time, cells were placed at -70°C prior to RNA purification. Total RNA was isolated using the High Pure RNA isolation kit (Roche, Inc.).

The 12K oligonucleotide arrays were designed by Combimatrix, Inc. Arrays were constructed with Version 3 of the *Neurospora crassa* genome sequence [9] from a file

labeled *neurospora\_3\_gene\_dna\_3205.txt* downloaded from the Broad Institute Web site. From this sequence Combimatrix, Inc. electrochemically synthesized 12,000 oligonucleotides of ~35 nt on their chips including 633 negative control oligonucleotides as well as 4 positive control oligonucleotides from each of 4 distinct  $\lambda$ -phage sequences, each randomly distributed in 22 positions on the chip.

Total RNA (750 ng) was subjected to one round of amplification using the MessageAmp aRNA Amplification Kit (Ambion, Inc.), which uses an "Eberwine type" amplification. Biotin-11-UTP and CTP (Enzo Life Sciences, Inc.) were incorporated during an *in vitro* transcription reaction. Five micrograms of aRNA were fragmented, then a total of 10 pM each biotinylated spike-in positive control oligonucleotide ( $\lambda$ -phage) were added with hybridization solution and, the mixture, hybridized according to the manufacturer's protocol rev 2.03 (<http://www.combimatrix.com>). Hybridization was carried out at 45 °C for 24 hours using a 25 % formamide based solution. Washing of chips was done according to manufacturer's protocol rev 2.03. Streptavidin Alexa Fluor® 647 conjugate (Invitrogen) was employed at a final concentration of 1.0  $\mu$ g/ml to visualize hybridized target. Laser confocal scanning was performed on a GSI Lumonics ScanArray 5000 (Perkin Elmer, Inc.) using a laser power and a PMT gain setting adjusted less than 10% between arrays.

The median foreground (FG) counts were used on all 12K features. From each median foreground count on an oligonucleotide array a background subtraction was computed using the 5<sup>th</sup> percentile of the negative control oligonucleotide features. Median foreground counts were normalized within arrays by multiplying each feature's median FG count on the particular chip  $x$  (median FG across all chips in the cycle) / (median FG count on the particular chip). A MIPS functional classification was assigned to each feature on a chip. [10] Hierarchical clustering of genes was implemented using the methods developed in Eisen et al., [8] and implemented in Cluster 3.0 [11] available at <http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster> and visualized with Java TreeView 1.0.12 [12] available at <http://jtreeview.sourceforge.net>.

Putative QA-1F-binding sites were identified with the GCG program pattern (Accelrys, Inc.) operating on the 1000 nt upstream of each identified gene [9] from the file *neurospora\_crassa\_#10BD2C.fasta* from the Broad Institute Web site. The offset used was 1 and overhang 0. Patterns searched are given in. [13]

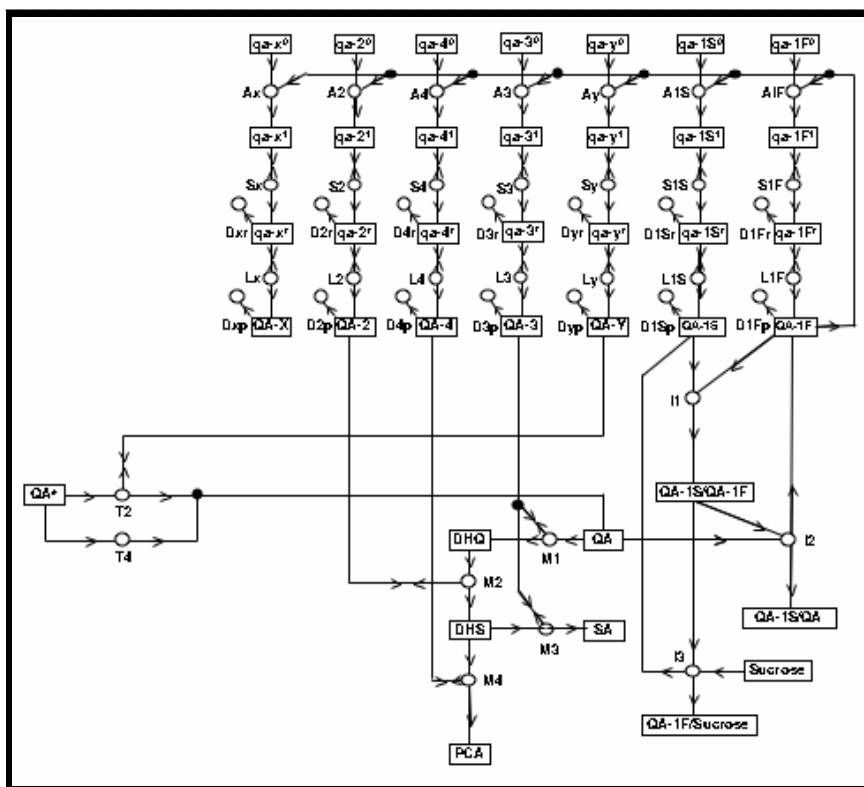
**The ensemble method of circuit identification was used [2].** An ensemble method of circuit identification was used to fit genetic networks of the structure in Figure 1 to RNA profiling data on genes in the *qa* cluster. With a sweep being a visit once on average to each parameter in the model (i.e., initial species concentrations and rate constants in Figure 1), 40,000 sweeps were used to equilibrate a Markov Chain in the parameter space. Transitions between points in the parameter space were governed by the Metropolis acceptance probability

$$r = \min(1, Q(\theta')/Q(\theta)),$$

where  $\theta'$  is the proposed update in the parameter space and  $\theta$ , the current state of the Markov Chain. Since the acceptance probability affects the efficiency of the Metropolis algorithm [14], a 50:50 mixture of global and local updating rules were used in such a way as to keep the acceptance probability between .33 and .67 by adjusting the step width of the Markov Chain. A local move involved updating only one element of  $\theta$ ; a global move involved updating all elements of  $\theta$ . For each proposed update, the LSODES method of solving ordinary differential equations was used to solve the kinetic equations specified by Figure 1. [15] The chain was initialized with a best fitting ensemble using the data in Battogtokh et al. [2] and a minimum chi-squared statistic of 14.

### Results and Discussion:

The microarray experiment was designed to determine what genes of *N. crassa* respond significantly to a shift from sucrose to quinic acid as the sole carbon source. The fungus was grown in glucose (2%) for 48 hours (hrs) (columns 1 to 13, Figure 2), and then compared with cultures shifted from sucrose (2%) to quinic acid (columns 14 to 21, Figure 2). The transcriptional profile at 0 hrs, 0.5 hrs, 1 hrs, 1.5 hrs, 2 hrs, 4 hrs, 6 hrs, 8 hrs (columns 14 to 21) after shift from sucrose to quinic acid can be seen in Figure 2. The bright green (-3) and bright red (+3) is gene expression level on a decadic log scale. The microarray data were produced from Combimatrix chips probed with biotin labeled aRNA as the target. The microarray chip contained 11,000 *N. crassa* genes (45 genes used as positive controls) and 633 negative control oligonucleotide sequences including those derived from plant, bacterial, bacteriophage, and *N. crassa* rDNA sequences. The empirical false positive and false negative rates are 0.16 and 0.38, respectively. The data were normalized within the array and background subtracted using the average spot intensity of the negative controls on the chip, logging, and clustering with UPGMA using Euclidean distance between mRNA profiles of different genes (i.e., average linkage). [8] Only the transcriptional profiles of approximately 895 genes with QA-1F binding sites were analyzed. [13]

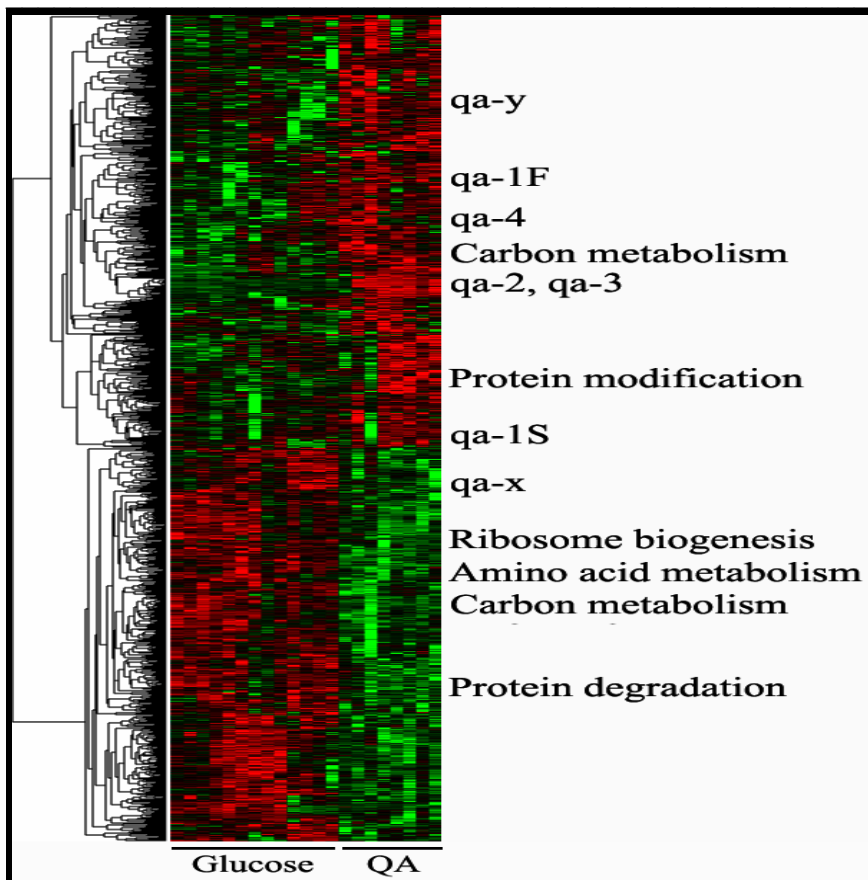


**Figure 1:** Genetic network for the *qa* gene cluster (redrawn from Kochut *et al.*, [16]). Circles denote reactions, and squares denote reactants or products. Boxes linked with arrows pointing into a circle are reactants, and boxes pointed to by arrows emanating from a circle are products. Boxes linked by double arrows to circles appear on both the left and right side of a reaction as catalysts. Circles with no outgoing arrows are decay reactions. The boxes and circles in the top part of the diagram constitute the Central Dogma. Boxes in the bottom of the diagram constitute the biochemical pathway converting quinic acid to energy and carbon for the cell. The QA-1F protein feeds back to turn on all seven genes in the *qa* cluster. The QA-Y protein is a carrier protein, moving quinic acid (QA<sup>e</sup>) external to the cell into the cell (QA)

The clustering of transcriptional profiles of individual quinic acid responsive genes revealed groups of functional categories of carbon compound and carbon utilization, (1.05), protein modification (14.07) and degradation (14.13), amino acid metabolism (1.01), and ribosome biogenesis (12.01). [10] The results from this transcriptosome dynamics (Figure 2) indicate that protein degradation, ribosome biogenesis and amino acid metabolism are turned off during the shift from glucose to quinic acid. Protein degradation functions which are switched off in the presence of quinic acid include 20S proteasome, ubiquitin-protein ligase and acid proteinase. A number of putative mannosyl transferase enzymes which function in protein glycosylation are turned off and epoxide hydrolase, aldose epimerase and alcohol dehydrogenase enzymes which function in carbohydrate and energy metabolism are switched on after the change in carbon source. The gene which encodes a probable fructose 1, 6 biphosphatase was turned on during the shift from glucose to quinic acid, suggesting that the gluconeogenesis pathway may become essential in the metabolic shift in carbon

metabolism. However, the gene encoding phosphoenolpyruvate carboxykinase which supplies the gluconeogenesis sequence with metabolites was not detectable under these conditions (either because it did not have a QA-1F binding site or because its RNA level did not change significantly under shift from sucrose to quinic acid). Induction of the genes encoding two distinct formate dehydrogenases were induced on quinic acid as the carbon source but the genes encoding isocitrate lyase and malate synthase, key enzymes of the glyoxylate cycle, were not detected.

Protein modification genes which encode products that function as kinases and phosphatases are turned on with quinic acid as a sole carbon source. It is worth mentioning that various kinase and phosphatase proteins have been clearly emphasized as key factors in metabolic processes such as regulatory cascades. However, the exact role that the quinic acid responsive genes play in metabolism remains to be determined.

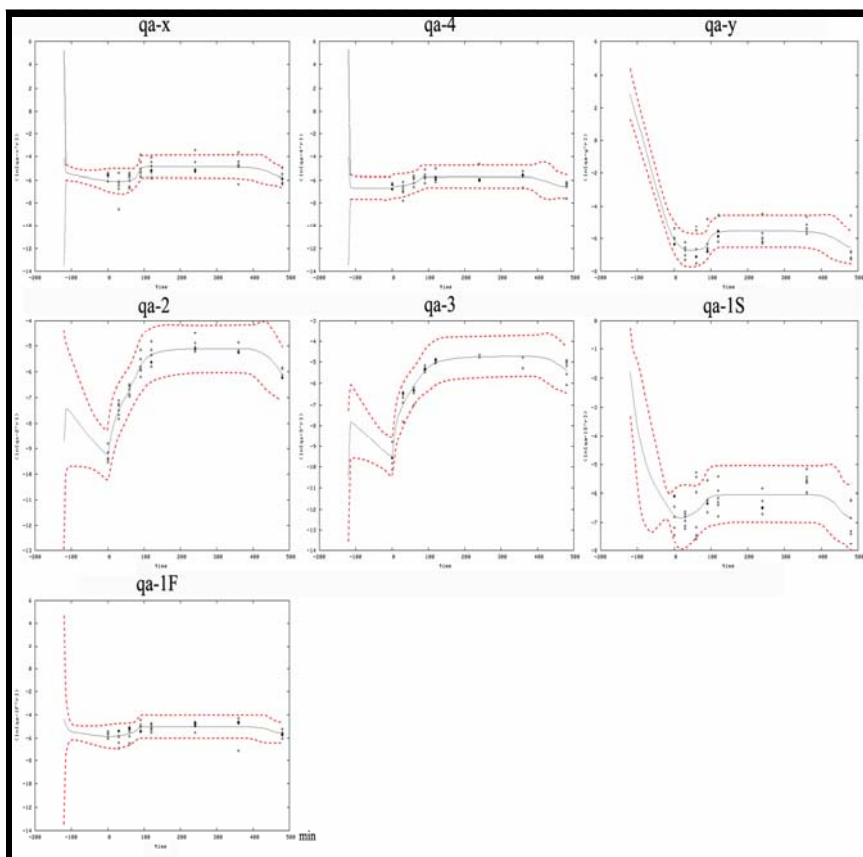


**Figure 2:** Clustering of the RNA profiles of 895 genes with a QA-1F binding site. Only genes with a QA-1F binding site responding to a shift from 1.5% sucrose to 0.3% quinic acid (QA) in their mRNA profiles from strain OR-74A are displayed. The first 13 time points were control cultures on 2% glucose from the *bda* strain. The last 8 time points were OR74A cultures shifted from sucrose to quinic acid. Times of the last 8 time points after shift from sucrose to quinic acid were 0, 0.5 hrs, 1 hrs, 1.5 hrs, 2 hrs, 4 hrs, 6 hrs, and 8 hrs. After background subtraction, normalization within arrays relative to grand median of each chip, logging, the mRNA profiles were clustered with UPGMA using Euclidean distance between mRNA profiles of different genes (*i.e.*, average linkage). [8] The bright green is  $-3$  and the bright red is  $+3$  is expression level on a decadic log scale. Data below arose from 21 chips probed with a biotin labeled aRNA. The *qa* cluster genes and cluster centers of MIPS functional categories appear in the right margin of this microarray experiment. Genes in Figure 2 are represented at least 5 times on each chip. The ~895 genes were selected by t-test comparing the mean of the first 13 time points on glucose (2%) with the mean of the last 8 time points on quinic acid with those having  $t_{19} > 2.093$  ( $\alpha=0.05$ ) displayed

Thus, the results suggest that numerous gene products are somehow related to quinic acid metabolism. Among the 895 genes with QA-1F binding sites are 105 genes that are clustered into the four functional categories indicated above and 412 genes (46 %) with unknown function. As expected, the quinic acid gene cluster was turned on by quinic acid. The remaining genes show a scattered pattern in the MIPS classification scheme. [10]

The development of a series of new computational biology tools allows the identification of genetic networks from RNA profiling data. [2] One approach known as the ensemble method for identifying regulatory circuits was

applied to the *qa* gene cluster (see methodology) with results displayed in Figure 3. The dots are values measured by the microarray experiment and the space between the broken lines ( $\pm 2$  standard errors) in red are predictions by the fitted model ensemble. All seven of the *qa* gene mRNA levels are predicted well by fitting the microarray data to the model ensemble with topology of that in Figure 1, and in almost all cases the areas between the broken lines (in red) cover the experimental data. The final minimum chi-squared obtained was 5212 with 486 data points and with the model having 38 species, 67 rate constants, and 3 scale factors (converting observations from Northern/microarrays to model units).



**Figure 3:** Predicted mRNA levels from the model ensemble fitting the data agree well with the observed mRNA levels. The mean of the ensemble is denoted by a solid line (in black), and the dotted lines (in red) are used to indicate  $\pm 2$  ensemble standard errors about the ensemble mean

### Conclusion:

The ensemble network identification was successfully applied to a portion of a whole-genome array analysis of *N. crassa* shifted from sucrose to quinic acid carbon source (relative to control cultures on glucose). More than 100 gene products identified by microarray and cluster analysis are possibly linked to quinic acid metabolism. The results reported here suggest that the pattern of gene expression in *N. crassa* is dramatically altered when shifted from sucrose to quinic acid, thus providing the fungus with the metabolic capability to use a less preferred carbon source for energy and growth. Future studies will seek to integrate these genes and their products into the QA reaction network model and thus provide an expanded view of the metabolic circuit of this model organism.

### Acknowledgement:

This work was supported by National Science Foundation grant MCB-0542915 to D.A.L. and J.A. as well as by NSF DBI-0243754, NSF BES-0425762, and the UGA College of Agricultural and Environmental Sciences.

### References:

- [01] R. F. Geever, *et al.*, *J. Mol. Biol.*, 207:15 (1989) [PMID:2525625]
- [02] D. Battogtokh, *et al.*, *PNAS*, 99:16904 (2002) [PMID: 12477937]
- [03] U. S. Bhalla, *Science*, 283:381 (1999) [PMID: 9888852]
- [04] N. H. Giles, *et al.*, *Microbiol. Rev.*, 49:338 (1985) [PMID: 2931582]
- [05] N. H. Giles, *et al.*, *PNAS*, 58:1930 (1967) [PMID: 4966264]
- [06] G. H. Braus, *Microbiol. Rev.*, 55:349 (1991)[PMID: 1943992]
- [07] E. Sattlegger, *et al.*, *J. Biol. Chem.*, 273:20404 (1998) [PMID: 9685394]
- [08] M. B. Eisen, *et al.*, *PNAS*, 95:14863 (1998) [PMID: 9843981]
- [09] J. E. Galagan, *et al.*, *Nature*, 422:859 (2003) [PMID: 12712197]
- [10] G. Mannhaupt, *et al.*, *Nucl. Acids. Res.*, 31:1944 (2003) [PMID: 12655011]

- 
- [11] M. J. De Hoon, *et al.*, *Bioinformatics*, 20:1453 (2004) [PMID: 14871861]
- [12] A. J. Saldanha, *Bioinformatics*, 20:3246 (2004) [PMID: 15180930]
- [13] N. H. Giles, *et al.*, *J. Heredity*, 82:1 (1991) [PMID: 1825499]
- [14] G. O. Roberts, *et al.*, *Ann. Appl. Prob.*, 7:110 (1997)
- [15] B. Aleman-Meza, *et al.*, *Theory and Practice*, (2007)
- [16] K. J. Kochut, *et al.*, *Parallel and Distributed Databases*, 13:43 (2003)

Edited by Susmita Datta

Citation: Logan *et al.*, *Bioinformatics* 1(10): 390-395 (2007)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.