GeniQuest:
Using Computational Biology for Guided Inquiry

Randy Von Smith, Chad Dorsey, Stephen Bannasch, Aubrey Smith, Francis Eberle, Jon Geiger, Gary A. Churchill

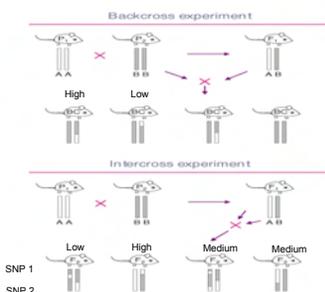**The Jackson Laboratory , Bar Harbor, Maine 04609 USA**

Abstract

GeniQuest is a team effort combining the talents of the Jackson Laboratory, the Maine Mathematics and Science Alliance, and the Concord Consortium to develop classroom modules to teach computational biology. The project introduces students to genetics, Quantitative Trait Loci (QTL) analysis, and the relationship between phenotypes and genotypes. The project uses a simplified model of Dragons and Drakes (small dragons) to allow the students to breed animals inside a virtual computer-generated world to answer genetic questions. Students learn to use software to perform QTL analysis and to use Genome Browsers to explore target genomic regions for potential genes of interest. Ultimately the students are guided though a search for the genetic cause of a disease in the model organism.

Introduction

The Jackson Laboratory has successfully conducted summer student programs for over 78 years. This program gives talented high school and college students the opportunity to conduct cutting edge scientific research in the laboratories working on mouse models for human disease. The summer program is modeled on the way research is conducted in the modern laboratory. The students work with a mentor and design a written proposal for their work. The research is conducted in a team environment with close supervision by their mentor. The summer concludes with the students completing a written summary of their work and a public oral presentation at the laboratory.

Three years ago the Center for Genome Dynamics designed an independent study course in computational biology (ISCB) staring with the design of the JAX summer program and incorporating Journal Clubs and more frequent presentations. The course expanded to a two semester program with the first semester focused on surveying the field of computational biology and the second semester dedicated to conducting research.

We gained experience at JAX conducting the ISCB course with students from residential magnet schools and students that started with a strong focus in science and mathematics. The next challenge was to adapt this technique to teach the core principles to a wider audience of students and to attempt to reduce the instruction to a class module that can be completed in around 3 weeks. We decided to build the module around a dramatic story, a search for a cure….

Independent Studies in Computational Biology

In this course students are prepared for conducting team oriented independent research by a lecture series that covers the core processes of:

- Introductory Statistics
- The R computer language
- Asking meaningful Scientific Questions
- Reading the scientific literature
- Writing a NIH grant proposal
- Genetics of the laboratory mouse
- Quantitative Trait Loci
- Microarray analysis
- Cluster analysis

All of this material is currently on a web based course management system (moodle) accessible to this year's class

How can we bring the same dynamic experience to a larger school audience in a class module?

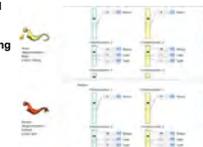## Figure 1 : Introduce a model organism and controlled crosses:

QTL analysis uses statistics to locate regions of the genome that are associated with the phenotype. Looking at the patterns of SNP1 and SNP2 and the known phenotype of the parent strains, we can see that SNP1 is consistent with the phenotype but SNP2 is not.

Figure 2 : QTL scan to narrow genomic region

Use QTL software to isolate a region of interest (Confidence Interval 95%) and show the Effect Plot for the phenotype:

```
Marker     chr  pos   lod
D4Mit237    4   17.5  6.557876
D4Mit164    4   29.5  8.093661
D4Mit178    4   30.6  6.374922
```

Figure 3: On Mouse and Human, Drake and Dragon

While the mouse is an excellent model for human disease, with 98% of the genes and the ability to map homologous regions between human and mouse, for the introduction of the techniques of computational biology this genome is too complex. There are over 25,000 genes with over 6 million SNPs for the mouse and for most QTL peaks the gene list is in the hundreds. To simplify this model we created fictitious animals:

Dragon (*Draco sapiens*): an ancient and intelligent species that is long lived and have associated with humans for many years. Much valued and integrated into the culture, dragons have become a keystone in the realm's economy.

Drake (*Draco familiaris, Draco igneus*): a small (50 gm) pet size dragon common to most regions of the realm with a wide span of phenotypes (including wings, fancy coloring, fire breathing, etc.). Long domesticated and bred for show, these are now the basis of continuing research on Dragons because of their shorter generation time (3 months compared to 100 years) and small size.

In designing the genome of the Drake we used the mouse genome, cut it up into slices and then reassembled it into two autosomal chromosomes and one sex chromosome. We included only 450 genes and 2,000 SNPs (most taken from the mouse {except wings and fire breathing…}).

Figure 4 : The Drama

Students are hired as Dragon Breeder apprentices and start by using an interactive software to explore the genetics of Dragons.
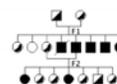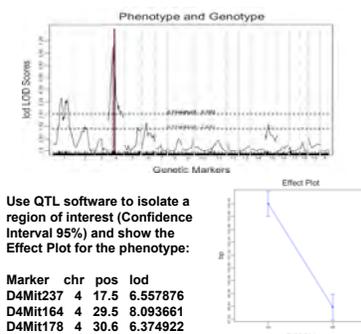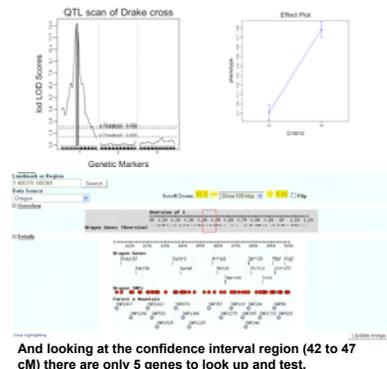
After learning about the genes and several traits the students are informed that there is a new and dangerous disease that is spreading through the Dragon and Drake population and they need to help find a cure.

Scaleblanche

A disease of the dragon identified by a progressive degeneration of the scales, starting at the tip of the tail and at extremities, and gradually working its way up the body towards the head. There is a model for this in the drake (draco familiaris), which progresses at a much faster rate; further study may explain the reason behind the disease, which seems to only affect those drakes in this region imported from Mountain stock.

Using strains of Drakes that are susceptible and resistant to *Scaleblanche*, the students design crosses to isolate regions of the genome to discover the causal gene.

Figure 5 : From QTL to Gene

And looking at the confidence interval region (42 to 47 cM) there are only 5 genes to look up and test.

In checking the database for the genes the students discover "Tmprss6" that is involved with iron metabolism. They find that the Dragon's diet has changed and they require supplements.

Conclusions

- With a proper back story, we can introduce the fundamentals of computational biology to the high school audience and give them the real world tools to solve interesting and challenging problem.

- By simplifying the genome the students have a problem that they can solve within the time limits of the class module.

- The module can be expanded in the future to include mouse QTL data sets allowing students to explore real world data. This is facilitated by the introduction of the same tools utilized when they were exploring the simplified genome of the drakes.