

SPORE* SERIES WINNER

Lessons from a Science Education Portal

David Micklos,† Susan Lauter, Amy Nisselle

When Cold Spring Harbor Laboratory’s DNA Learning Center (DNALC) launched its Web site in 1996, www.dnalc.org, we did not foresee that it would grow into a portal for 18 content sites reaching more than seven million visitors per year. The evolution of our multimedia efforts and the challenges along the way provide lessons for building learning resources or to attract larger audiences.

Our first major site, DNA from the Beginning, is a multimedia primer on 41 seminal concepts of modern genetics (1). This was followed by Your Genes, Your Health, a compendium of information on genetic disorders, and DNA Interactive, a companion to a Public Broadcasting Service series commemorating the 50th anniversary of the discovery of the DNA structure. The Image Archive on the American Eugenics Movement marked the online release of >2400 items from this dark period in science. Two sites focus on research insights into disease: Inside Cancer examines the genetic “hallmarks” of tumor cells; Genes to Cognition (G2C) Online explores how disorders of thinking entail multiple levels of biological complexity (2). These narrative sites are complemented by online notebooks of tested experiments in bacterial genetics, human and plant genomics, and RNA interference. Several experiments coordinate with purpose-built tools for bioinformatics analysis. We also partner with research groups and disease foundations to produce microsites and smartphone applications (apps) that focus on single topics.

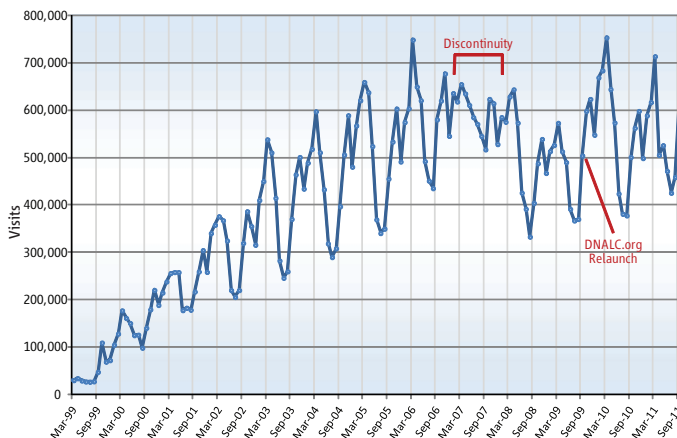
The DNALC benefited from early entry into the online world when there were only 10 to 25 million active Web sites (3). We rode a wave of increasing Internet speed and connectivity, with visitation increasing steadily each year and peaking at 7.1 million in 2007. Visitation then dipped to 6.0 million in 2008, when there were more than 100 million active Web sites (see the first figure). Search engines become defacto arbiters

DNA Learning Center, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.

*SPORE, Science Prize for Online Resources in Education; www.sciencemag.org/site/special/spore/.

†Author for correspondence. E-mail: micklos@cshl.edu

Doing your best on the Web requires attending to search engines, answering hard questions, and making cybertools accessible to a broad audience.



Visits to DNALC Web sites. Visits rose steadily through 2006 and followed an academic year cycle: lowest in August, rising to a preholiday high in November, and peaking in March. The cycle was disrupted in 2007, likely by changes in search algorithms, then reset in 2008 at about 100,000 fewer visits per month.

of an exponentially expanding Web, periodically directing “robots” or “spiders” to build a searchable index of a site. The engine then calculates site rankings. Our decline in visitation was almost certainly precipitated by changes in search algorithms. We therefore redesigned our Web sites to increase “visibility” to search engines, a process called search engine optimization (SEO).

A large part of our search problem stemmed from the advanced Flash software that allowed us to integrate text, video, and animation, into a file that could not be indexed. Our SEO solution was to develop an html content “mirror” to direct robots to text-based files. The first implementation of this fix, for the Your Genes, Your Health Web site increased search referrals from 10.6% to 36.8% of monthly visits. In late 2009, we revamped DNALC.org using cascading style sheets; providing rich keywords, title tags, and descriptions; and refreshing content with blogs and newsfeeds. This SEO makeover resulted in an 89.4% increase in average monthly visits by search engine spiders and a 24.5% increase in total visits in 2010.

Students and teachers often turn to the Internet to answer specific questions or to illustrate key points. Results are harder to find when deeply embedded within a Web site. Thus, we disaggregated our Web sites into searchable content “atoms” that can be

accessed individually. Aided by our participation in the National Science Digital Library Project, we cataloged >5300 animations, videos, photos, and illustrations. We also organized these content atoms into thematic collections. These strategies worked, with 79.8% of visits to DNALC.org now arriving at a content atom identified by a Web search.

In parallel, we aggressively distributed content through other channels. We established a DNALC channel on YouTube (4), adding 188 podcasts, animations, and videos. We also retooled the immersive three-dimensional (3D) Brain module from G2C Online as a smartphone app. Within months of release, 3D Brain rose to number 7 of 7100 education iPhone apps and number 1 among 250 iPad education apps, with >900,000 downloads to date. In 2010, the Web version of 3D Brain received 54,868 visits, whereas the app version was downloaded 413,874 times. As a result of SEO makeovers, disaggregation, and aggressive moves into apps, YouTube, and blogs, total visitation rose 9.6% to 7.0 million in 2010, where we were before the 2007–08 discontinuity. Additional efforts in spring 2011 increased visitation 20.5% in June to November.

All Web site developers are challenged to answer the question, “Does this program actually help students to learn?” In the context of the 2007 America COMPETES Act (5), Congress asked the same question of the educational grant portfolios of the National Science Foundation (NSF) and National Institutes of Health (NIH), which led to increased pressure on principal investigators to move from anecdotal reports of teacher satisfaction to sophisticated studies of resources’ impact on students. This is a big task. NSF and NIH investigators seldom have direct access to students,

CREDIT: DNA LEARNING CENTER

Downloaded from www.sciencemag.org on January 4, 2012

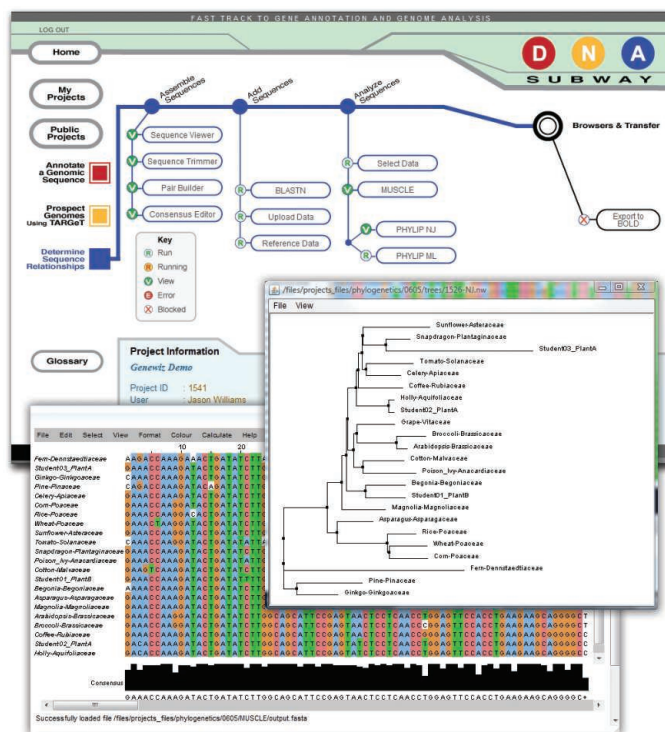
and learning takes place in complex environments with many interacting institutional and personal variables (6).

We conducted experiments from 2010–11 to test whether G2C Online and Inside Cancer improve student learning (7). The experiments involved 626 students in 28 high school and college classrooms across 10 states. To control for differences between teachers and students, we used a crossover repeated-measures design, in which each student participated as both an experimental and a control subject in a single repetition of the protocol (8). Participating teachers taught two topics to students separated into classes A and B (average class size 24 students). For the first topic, class A used a DNALC Web site for classwork, and class B used lectures, textbooks, or other Web sites. The classes then switched conditions for the second topic, so that each student learned one topic using a DNALC Web site and one topic using another resource. Students completed a quiz after each topic, which allowed comparison of how well each student learned with and without the use of a DNALC Web site. Students' quiz scores were significantly higher when using either G2C Online (mean 81.2 ± 19.5 versus 70.7 ± 20.2 , $t_{328}=7.789$, $P < 0.001$) or Inside Cancer (85.0 ± 20.8 versus 73.8 ± 21.3 , $t_{296}=7.361$, $P < 0.001$). Thus, we have an answer to that difficult question: an engaging Web site can potentially increase student learning by about one letter grade!

Science takes place on a continuum between research and education. Traditionally, access to limited data kept most good science far to the research end of the spectrum. Now, the availability of nearly unlimited data from high-throughput DNA sequencers, plus powerful bioinformatics analyses from shared servers, offers the promise of merging research and education into a single endeavor.

About the Authors

David Micklos founded the DNA Learning Center in 1988 as the nation's first science center devoted to public genetics education. Three facilities in the New York metro area provide hands-on lab experiences to 30,000 students per year. Sue Lauter leads the DNALC team of multimedia designers. Amy Nisselle produces and evaluates DNALC Web sites and apps. An evolving team of designers, programmers, writers, and producers contributed to our Web sites, including A. Ava, S. Chan, M. Christensen, S. Conova, J. Connolly, C. Ghiban, U. Hilgert, E.-S. Jeong, M. Khalfan, J. Kruper, B. Nash, B. Terrill, J. Williams, J. Witkowski, and C.-H. Yang.



DNA Subway. The Blue Line of DNA Subway allows students to analyze DNA barcodes and construct phylogenetic trees from DNA sequence data.

For the first time in the history of science, students and teachers can work with the same data, at the same time, and with the same tools as elite-level researchers (2).

We have devoted considerable effort to developing educational resources to help students generate, share, and analyze genome data. In 1998, we developed the first cyber-experiment to allow students to analyze a small portion of their own genome. At their schools, students isolate DNA from cheek cells and then use PCR to amplify the mitochondrial control region. After free or low-cost processing, student DNA sequences are uploaded to our BioServers Web site. There, students use bioinformatics tools to compare their sequences to those of classmates, world populations, and extinct hominids. Over 56,000 samples have been sequenced to date, and the BioServers site has received 1.4 million visits.

Our interest in community workspaces and bioinformatics culminated in our involve-

ment in the iPlant Collaborative, a project to develop a cyberinfrastructure for plant science research (9). Drawing on the computers and storage of TeraGrid (now XSEDE), iPlant's Discovery Environment enables scientists to build and analyze phylogenetic trees with thousands of species and to correlate plant phenotypes with variation in large-scale genotype data sets. As educational outreach for this project, we developed a parallel bioinformatics workflow, DNA Subway (see the second figure). Using the metaphor of a familiar subway map, this simple, intuitive interface allows nonspecialists to extract information from DNA sequence data. "Riding" on different "lines," students can predict and annotate genes in genome sequences, prospect genomes for related sequences, build phylogenetic trees, and analyze DNA barcodes.

An important battle for cyber-literacy takes place midcontinuum, where scientist educators in colleges and universities can invite students as coinvestigators to explore abundant genome data. Here, intuitive bioinformatics work flows must work in classrooms and teaching laboratories, without the need of high-level computational support. By doing their best on the Web, multimedia producers can equip teachers, students, and even citizen scientists to actively participate in the genome age.

References and Notes

1. Web sites at www.dnalc.org were funded by the NIH; NSF; Howard Hughes Medical Institute; National Human Genome Research Institute; Department of Energy; Dana Foundation; Josiah Macy, Jr. Foundation; Roche Molecular Systems; Biogen Idec Foundation; William A. Haseltine Foundation; and Applied Biosystems.
2. A. Nisselle et al., in *Proceedings of World Conference on Educational Multimedia, Hypermedia, and Telecommunications 2011*, T. Bastiaens and M. Ebner, Eds. (Association for the Advancement of Computing in Education, Chesapeake, VA, 2011), pp. 1403–1412; www.editlib.org/p/38050.
3. Netcraft, news.netcraft.com/.
4. DNA Learning Center, www.youtube.com/DNALearningCenter.
5. U.S. Congress, America COMPETES Act, (2007); commdocs.house.gov/reports/110/h2272.pdf.
6. L. Cohen, L. Manion, K. Morrison, *Research Methods in Education* (Routledge, Oxford, ed. 6, 2007).
7. These studies were supported by the Hewlett Foundation and an NIH Science Education Partnership Award. We thank C. Connolly for her assistance.
8. B. Jones, M. Kenward, *Design and Analysis of Cross-Over Trials* (Chapman and Hall, London, 2003).
9. iPlant Collaborative, www.iplantcollaborative.org, funded by the NSF.