Cloud-enabling an Evolutionary Genetics Tool and Computational Methods for Invigorating STEM Learning and Research

Bina Ramamurthy
CSE Department
University at Buffalo
345 Davis Hall
Buffalo, NY 14260
bina@buffalo.edu

Jessica Poulin
Department of Biology
University at Buffalo
109 Cooke Hall
Buffalo, NY 14260
jpoulin@buffalo.edu

Katharina Dittmar
Department of Biology
University at Buffalo
109 Cooke Hall
Buffalo, NY 14160
kd52@buffalo.edu

ABSTRACT
We present a cloud-enabled comprehensive platform called *Pop!World* for experiential learning, education, training and research in population genetics and evolutionary biology. The major goal of *Pop!World* is to leverage the advances in cyber-infrastructure to improve accessibility of important biological concepts to students at all levels. It is designed to empower a broad spectrum of users with access to cyber-enabled scientific resources, tools and platforms, thus, preparing the next generation of scientists. *Pop!World* offers a highly engaging alternative to currently prevalent textual environments that fail to captivate net-generation [11] audiences. It is also more mathematically focused than currently available tools, allowing it to be used as a basic teaching tool and expanded to higher education levels and collaborative research platforms. The project is a synergistic inter-disciplinary collaboration among investigators from Computer Science & Engineering and Biological Sciences. In this paper, we share our multi-disciplinary experience (CSE and BIO) in the design and deployment of the *Pop!World* platform and its successful integration into the introductory biological sciences course offerings over the past two years. We expect our project to serve as a model for creative use of advances in cyber-infrastructure for engaging the cyber-savvy net-generation students and invigorating STEM (Science, Technology, Engineering, and Mathematics) education.

Categories and Subject Descriptors
C.2.4 [Distributed Systems]: Client/server and Distributed applications; D. [Software]: D.1.3 [Concurrent Programming]: Distributed programming; Parallel programming; D.1.7 Visual Programming; H.3.4 [Systems and Software]: *Distributed systems;* H.5.1 [Multimedia Information Systems]: Animations; H.5.2 [User Interfaces] *Graphical user interfaces (GUI);* J. [Computer Applications]: *Education;* J.3 [Life and Medical Sciences] Biology and genetics.

General Terms
Algorithms, Performance, Design, Experimentation.

Keywords
Cloud computing, cyber-infrastructure, tools, genetics, evolutionary biology, net-generation, education.

1. INTRODUCTION
The primary motive behind the *Pop!World* project is to build a comprehensive tool by leveraging the emerging cyber-infrastructure to inspire learners at all levels to engage in Biological Science study and research. There are three primary questions that motivated us, all of which are relevant in any educational setting: (i) How can we address pedagogical (didactic) challenges in learner-centric teaching? (ii) How can we provide a sustainable and scalable technology infrastructure for a learning tool? and (iii) How can we use a cyber-infrastructure enabled delivery model to effectively teach students the mathematics behind evolutionary biology and engage them in critical and computational thinking?

Delivery models for education have undergone quite dramatic changes since the introduction of the Internet technology. One significant advance has been the move from teacher-centric to learner-centric models, and incorporation of learning styles and pedagogic models into teaching and learning [12,14]. Providing flexibility on how often and when students access educational interventions is especially important for the cyber-savvy, multi-tasking generation [11]. We need an education delivery model that is learner-centric, provides continual formative assessment, effectively monitors the students' progress, appeals to broad range of audiences, and is easy to implement for the user/educator. This need is all the more critical for the

STEM areas such as Biological Sciences and Computer Science and Engineering.
We wanted to create a learning environment that is usable by a wide variety of people without any technological barrier, while at the same time minimizing financial involvement. Specifically, we want to combat the extremely short cycle of obsolescence prevalent in technology products. In many cases, this is particularly detrimental in poor school districts, or situations of economic crisis, as it may not be possible to continuously buy new infrastructure, or to support labs or administrative staff to accommodate modern tools and technologies.

Another important motive is to foster the student's understanding of the interdependency of disciplines (e.g. to do biology, one needs to know math), and thus increase their grasp on interdisciplinary research. A comprehensive solution was developed to address all the issues discussed above.

2. BACKGROUND

2.1 Evolutionary biology: is the central guiding principle of all biological sciences [6]. Evolution shapes the living world we see in both adaptive and non-adaptive ways [5]. Population genetics is the field of biology that allows scientists and students to make predictions and trace the outcomes of evolutionary forces [9]. Yet, as population genetics is highly mathematical, these concepts are difficult to convey, even to advanced students. As evolution is rarely suited to a traditional bench lab environment, these topics may be best demonstrated through computer simulation.

2.2 Cloud computing: Cloud computing is commonly defined [13] as a cyber-infrastructure that provides dynamically scalable, and often virtualized resources as a service. Powerful working examples for this approach are, for instance, GoogleApps (software) [7], Microsoft Azure (infrastructure) [15] and Amazon's EC2 (platform) [1]. In each of the cases, resources are provisioned entirely by an external platform (the cloud), which provides the service on demand and at the configuration specified by the user. Most importantly, the user need not have expert knowledge in, or control over the technology infrastructure in the cloud. Apart from commercial clouds, there are a few emerging clouds [3, 4, 10], which are mostly used to address computer-intensive problems. This project will demonstrate an innovative use of cloud computing for supporting scientific learning environments. Specifically, we will explore the flexibility of the cloud computing approach to enable convenient and on-demand access to a shared pool of configurable learning/research applications that can be released with minimal management effort and learning curve for the end-user.

3. Project *Pop!World*

The *Pop!World* project was initiated by the Biological Sciences department primarily to address the attrition in their introductory Biology courses. The specific goal of the project is to develop a professional education and research platform for high-fidelity simulations of evolutionary processes (e.g.: selection, mutation, migration, genetic drift, and non-random mating) to engage net-generation students and improve their interest and understanding of these concepts. Specific technical goals involved design and development of a comprehensive and end-to-end set of features, including simple interaction, high fidelity simulation, appealing visualizations, pop-up annotations as explanation for results and processes, parameterization and equation editor, split screen comparison of real world and simulated outcomes, automatic testing, personal preference for students, system configuration for teachers, experimental setup for hypothetical research, innovative collaboration setup for research and storage of historical and experimental results and data with load and playback facilities.

The project began with a simple "proof of concept" prototype of the tool in the Fall 2009. This preliminary stage helped immensely in building the truly diverse team of researchers and students, and in clearly understanding the requirements from the Biological Sciences (BIO) end, and the limitations on the Computer Science and Engineering (CSE) end. Rapid prototyping methods were used to complete the concept tool in about 6 months. The tool was used in a classroom setting in the summer of 2010. Cyber-savvy students were not only the users of *Pop!World* but also served as the creative force behind the design and implementation of the tool.

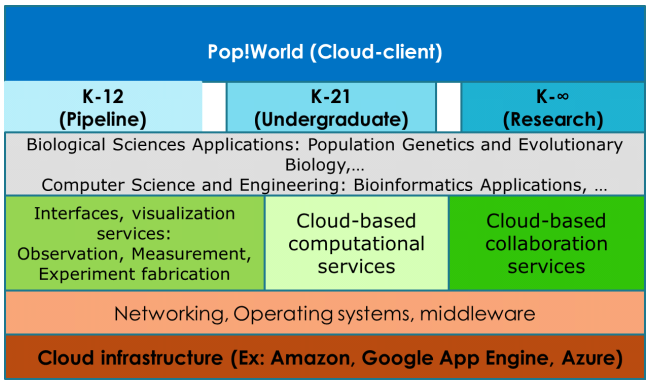

Figure 1: A High-Level Architecture of Pop!World

The student feedback from the use of the prototype and the lessons learned during its development and use helped us design a completed revamped version of the tool with three levels: K-12, undergraduate (teaching K-21), graduate (research K-∞) for use in formal and informal settings. The design and development of *Pop!World* is guided by the cyber-infrastructure approach explained in the report advisory panel on cyber-infrastructure. This report titled

"*Revolutionizing Science and Engineering through Cyberinfrastructure*" is frequently referenced in the context of cyber-infrastructure research as the Atkins Report [2]. We modeled our infrastructure on the core diagram given in the report, and modernized/updated it to reflect the recent advances such as cloud computing, and service-orientation to offer a truly state-of-the art environment in *Pop!World*.

3.1 Three Levels of Pop!World

Pop!World was conceptualized at three different levels of complexity: K-12 (***pipeline or gateway***: middle school – high school instruction), K-21 (***teaching***: college level undergraduate courses, esp. Biological Sciences), K-∞ (***research:*** undergraduate and graduate level research and experimentation) as shown in figure 1. These three levels are available as three different tools and the focus of this paper is on the teaching (K-21) version of the *Pop!World* tool. It has a highly visual presentation with capabilities for entering the experimental data (parameters) representing the various evolutionary forces (Selection, mutation, etc.) working through generations of evolution, engaging display of hypothetical red and blue lizards to present the population's changing traits, graph output that summarizes the evolutionary trend, and a computation windows that display the computed numbers. A representative screen shot of the *Pop!World Teaching* environment is shown in figure 2. The *Pop!World Gateway* environment is meant for high school students to teach them the fundamentals of Mendelian and population genetics. A screen shot of this interface is shown in figure 3.

Figure 2: Pop!World Discovery Interface

3.2 Deployment on the cloud

Pop!World was developed using Adobe Flash and the action script language strictly following software engineering principles in the design and object-oriented programming. The K-21 Teaching module has about 10000 lines of modular code understandable by BIO professors and students. The .swf (shockwave file) generated is then deployed on the Google App Engine (GAE) [8]. GAE is a comprehensive cloud platform that allows for deployment of applications irrespective of their size. GAE was chosen since the basic quota on this environment is free and more resources (processing power, bandwidth, storage etc.) beyond the free usage tier can be added by enabling billing. It offers the same reliability, availability and scalability at par with Google's own applications. Also GAE cloud offers excellent monitoring features for observing the load and for load balancing on demand.

GAE allows for multiple instances of the tool to be deployed. We deployed about 10 instances and load balanced the incoming requests among the instances of the tool. We expected about 200-300 simultaneous access request across the 10 instances.

Figure 3: *Pop!World* Gateway (K-12)

4. *Pop!World* IN BIO COURSES

Entering undergraduates use the *Pop!World* K-21 tool to have an immersive experience of each evolutionary force (migration, non-random mating, genetic drift, mutation, and selection). By entering the parameters that affect the population, students will "derive" the mathematical laws governing evolution by each force, allowing them to have a more intuitive understanding of how each microevolutionary force works to change a population. This can be paired with a more concrete mathematical presentation of the ideas, which is typical in a standard lecture format. Our experience has been thatthe graphical interface of *Pop!World* leads to a visceral understanding of these processes even if students do not have the mathematical background to use the actual formulas that govern these forces. Early simulation experiences aim to be directive about how to manipulate the simulation outcomes, but the format of the program allows students to be increasingly independent, leading to self-discovery of evolutionary patterns and trajectories.

When students control the parameters that govern the simulation, they are more inclined to engage with the program and understand the impact of the results. *Pop!World* is an experiment the student is doing – in much the same way an evolutionary modeler would use more advanced versions of the software. Such experiences are invaluable in encouraging students to remain in the STEM disciplines.

The *Pop!World* K-21 module has been fully integrated into the instruction and curriculum of the introductory course (BIO 200: Introduction to Evolutionary Biology) in the Biological Sciences department of University at Buffalo. The course is typically offered in the Summer and in the Fall semesters. The course is taught by the co-author on this paper, Dr. Jessica Poulin. About 1200 students were enrolled in the course in the Summer and Fall 2010. In general the use of the tool was well received and liked by the students. For example, when a lab work was assigned to the class of more 1200 students, the due dates were staggered so that we could manage the traffic and load. However when we monitored the load on the first day of the assignment we found that more than half the students in the class were on *Pop!World* working on their assignment! The bandwidth quota on GAE was exceeded and we had to increase the bandwidth capacity of GAE by enabling billing as well as by deploying more instances of the tool. It was good to observe the eagerness of the students in learning science through *Pop!World*. The cloud deployment of *Pop!World* helped in a quick response to the surge in the number of users (students) and to deploy additional instances of the resources on demand.

4.1 Outcome Assessment

The most valuable outcome of the project is the formation of a highly productive multi-disciplinary team of Computer Scientists and Biologists.

We designed survey instruments to assess the effect of *Pop!World* and were approved by the Internal Review Board at University at Buffalo. The evaluations discussed here are from Fall 2010.

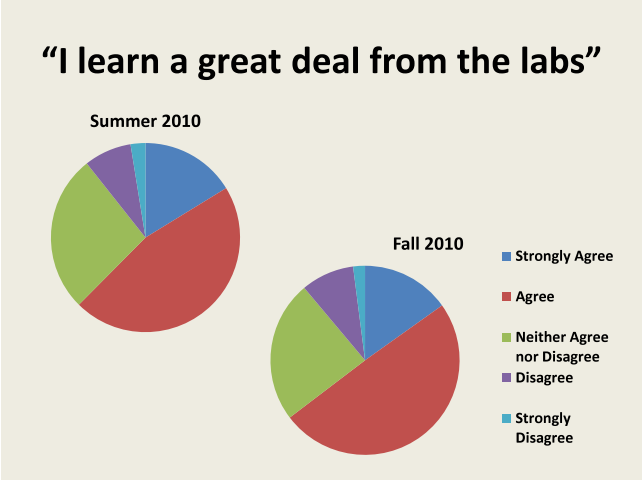

Figure 4: Learning through Pop!World

Figures 4-7 provide a series of representative evaluations from the first year of use of the *Pop!World* learning environment. Figure 4 shows about 65% of students agree that they learned a great deal from the lab. There is a slight improvement from the Summer 2010 to Fall 2010. It is in this Fall offering that we had the bandwidth overload for the cloud access. Since that time we have added a load balancer and we did not experience a problem with load during the Fall 2011 offering of the course.

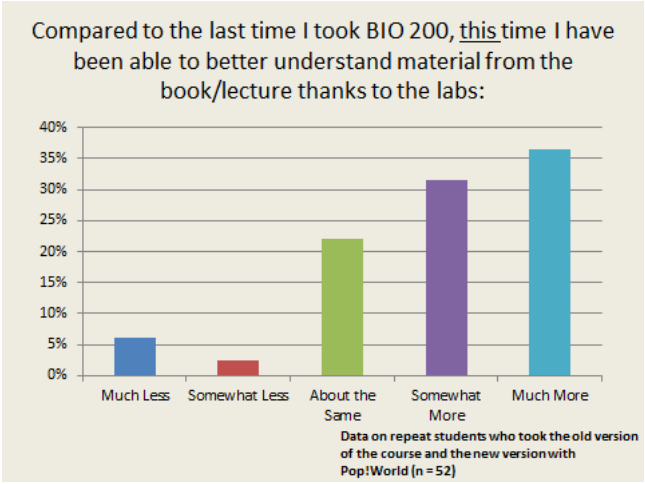

Figure 5: Understanding of Course Material

The introductory biology course is an important course for many science (STEM) majors including pre-med students. The attrition rate in the traditional offering of the course is as high as 20%. It is not unusual to find many students repeating the course for various reasons. We used this fact to evaluate the "understanding of the course material". The majority of repeat students clearly agreed that *Pop!World* (which they had not used in the first offering of the course) helped them understand the course material (Figure 5). This statistic is all the more significant since this is on the students repeating the course.

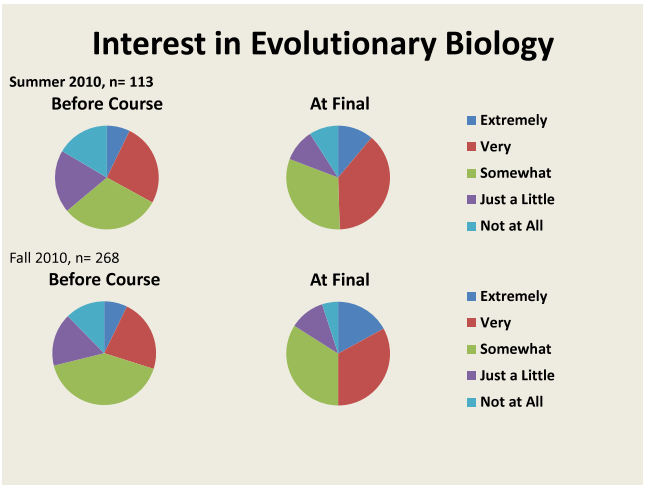

Figure 6: Interest in Evolutionary Biology

We evaluated the improvement in students' interest in Evolutionary Biology before and after the course, using *Pop!World* (Figure 6). The majority of the students responded positively. Clearly more than 80% agree that their interest increased though by different degrees (extremely, very, etc.). The significance of this outcome is that we have been able to engage the majority of the students in the course, whereas before many of them were dropping out of the course.

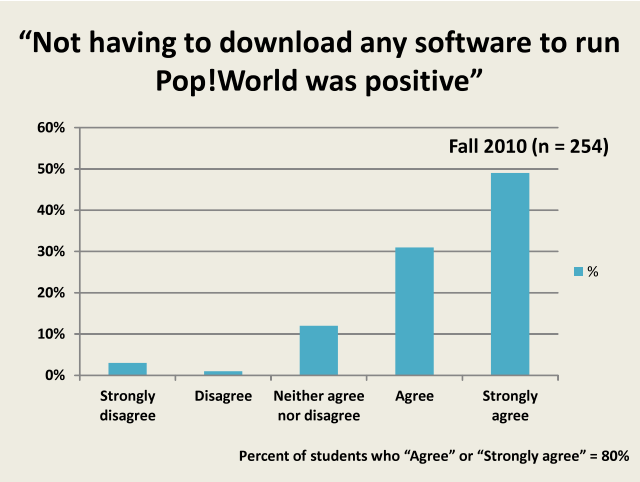

Figure 7: Pop!World Access Model (Cloud)

Figure 7 shows the students' evaluation of the cloud access model. Students simply access the cloud-deployed *Pop!World* without any need for downloading or installation. Students really like this aspect of *Pop!World*. Indirectly cloud-enabling also provided the 24X7 availability and access from anywhere that may not be given with departmental or university-wide servers.

Figures 4-7 provides validation for the introduction of *Pop!World* into the introductory Biology course BIO 200. Besides these, attrition in the course is down to 5-10%. We expect further improvement in the recent offering (2011) of the courses. There is a lot of interest in the *Pop!World* environment from students who are not in the course. The Chronicle of Higher Education had a report on the software. We experienced quite a response from the readers (general public, probably educators) since many of them accessed the cloud-enabled version when the article was published. Figure 8 shows the cloud monitored for the access patterns of *Pop!World*. The load curve at the top of the screen shows the intense activity we experienced after the publication of the *Pop!World* article. We attribute this traffic to the interest in the readers of the newspaper since there was no *Pop!World*-related work assigned to the students during that time frame.

5. SIGNIFICANT CONTRIBUTIONS

A technical success of the project is that of rapid prototyping and cloud-deployment of a highly usable Biology learning environment to invigorate STEM learning and research. This project is a success model for synergistic inter-disciplinary collaboration between Computer Science and Biological Sciences. Students from the CSE and BIO department worked together, learned from each other and accomplished the goals of the project. The *Pop!World* environment is available for anyone to use at http://popfrontpage.appspot.com/.

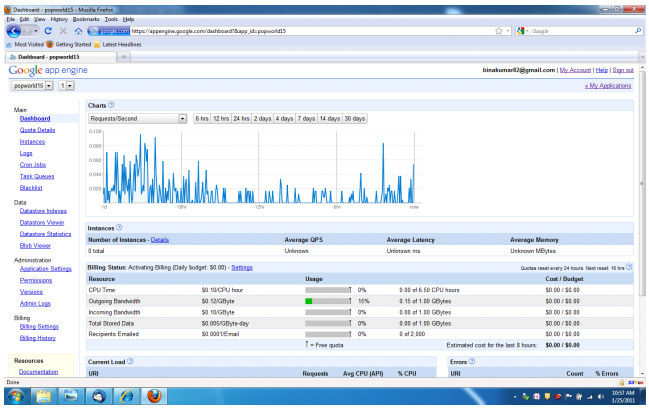

Figure 8: Pop!World Monitoring

6. FUTURE DIRECTION

We have nearly completed work on a K-12 model of *Pop!World* (Pipeline/Gateway). By summer 2013 we will have a *Pop!World* Research version available. The three versions each have distinct contents without much overlap, but they feed into each other. We plan to filter out the experiences from this project to publish general guidelines and best practices for cloud-enabling STEM tools and methods.

7. ACKNOWLEDGMENTS

This project is funded by the NSF OCI-CI-TEAM 1041280 awarded to B. Ramamurthy, J. Poulin, and K. Dittmar. Current version of (K-21) Pop!World was developed with the help of Tyler Taylor (undergraduate, Biological Sciences), and Timothy J. Hartnett, and Shuang Wu (graduates, Computer Science and Engineering).

8. REFERENCES

[1] Amazon Elastic Computing Cloud. Amazon EC2, http://aws.amazon.com/ec2/, last visited January 2012.

[2] D. E. Atkins, K. K. Droegemeier, S. I. Feldman, H. Garcia-Molina, M. L. Klein, D. G. Messerschmitt, P. Messina, J. P. Ostriker, M. H. Wright. Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, January 2003, http://www.nsf.gov/od/oci/reports/atkins.pdf, last viewed January 2012.

[3] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katzh et al, "Above the Clouds: A Berkeley View of Cloud Computing" http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf

[4] K. A. Delic and M. A. Walker. 2008. Emergence of the Academic Computing Clouds. Ubiquity 2008, August, Article 1 (August 2008), 1 pages.

DOI=10.1145/1414663.1414664
http://doi.acm.org/10.1145/1414663.1414664

[5] J.S. Freeman and J.C. Herron. Evolutionary Analysis, 4th Edition. Pearson/Benjamin Cummings, San Francisco, CA, 2007.

[6] D.J. Futuyma. Evolutionary Biology, 3rd Edition. Sinauer Associates, Inc, Sunderland, MA, 1998.

[7] Google apps. http://net.educause.edu/ir/library/pdf/ELI7035.pdf and http://www.google.com/apps, last viewed January 2012.

[8] Google App Engine. http://code.google.com/appengine/, Last viewed January 2012.

[9] P.W. Hedrick. Genetics of Populations, 4th Edition. Jones and Bartlett, Sudbury, MA, 2011.

[10] N. Leavitt, "Is Cloud Computing Really Ready for Prime Time?",Computer Volume 42, Issue 1, Jan. 2009 Page(s):15 – 20, http://www.computer.org/portal/web/computingnow/0209/theme/computer

[11] D. G. Oblinger and J. L. Oblinger. *Educating the Net Generation*. Educause Publication, 2005, http://www.educause.edu/EducatingtheNetGeneration/5989, last visited January 2012.

[12] The Project Kaleidoscope (PKAL), http://www.pkal.org/, last viewed January 2012.

[13] B. P. Rimal, E. Choi and I. Lumb, "A Taxonomy and Survey of Cloud Computing Systems" ncm, pp.44-51, 2009 Fifth International Joint Conference on INC, IMS and IDC, 2009 http://www.computer.org/portal/web/csdl/doi/10.1109/NCM.2009.218

[14] E. Soloway, "How the Nintendo Generation Learns," Communications of the ACM, 34(9), pp. 23-26, 1991.

[15] Windows Azure Platform. http://www.microsoft.com/windowsazure/, last visited January 2012.