Ralph Haygood, Ph.D.

Software developer • Data scientist • Population biologist

Skills

- Familiarity with a broad range of topics in computer science, statistics, applied mathematics, evolution, ecology, genetics, genomics, and physics.
- Designing, building, and maintaining web applications and other software, particularly for biological laboratories and core facilities.
- Mapping business processes and laboratory workflows into information architectures and software-user experiences and interfaces.
- Programming in languages including C, FORTRAN, JavaScript, Prolog, Python, R, Ruby, and SQL, with frameworks including Ruby on Rails, Flask, Vue.js, and Ext JS, and integrating subsystems including PostgreSQL, MySQL, Redis, and Memcached.
- Deploying web applications using tools such as Nginx, Apache, and Phusion Passenger and managing them using tools such as Docker, Docker Swarm, and Kubernetes, typically on virtual private servers running some version of Linux.
- Translating business and scientific issues into statistical and mathematical terms and interpreting quantitative results for nonspecialist audiences.
- Assembling and curating large data sets and applying a wide variety of statistical and machine-learning methods to them.
- Analyzing data using tools including Jupyter, SciPy, pandas, scikit-learn, and R.
- Formulating, analyzing, and interpreting mathematical, statistical, and computational models of evolutionary, ecological, and genetic processes.
- Applying evolutionary perspectives to diverse biological phenomena.

Examples

- As a freelance software developer, I've designed, built, and maintained laboratory information management systems (LIMS) for Duke University's Sequencing and Genomic Technologies facility and Proteomics and Metabolomics facility, which serve customers both at Duke and around the world. These systems, known as DUGSIM, SeqLIMS, and PAMLIMS, enable customers to get self-serve estimates, request quotes from staff, place and track orders, and receive invoices for services such as high-throughput DNA sequencing and mass spectrometry. They enable staff members to prepare quotes, process orders, and issue invoices. Since 2013, they’ve handled over 10,000 orders. They’re built with Ruby on Rails, Ext JS, Vue.js, MySQL, PostgreSQL, Memcached, Redis, Sphinx, Nginx, Phusion Passenger, and Ubuntu Linux.

- As the Data Science Developer at ReverbNation, I collaborated with executives, product managers, and marketers to understand and predict our users' responses to our communications and services. Usually, my work began with my colleagues' questions and intuitions, which I translated into descriptive statistics, graphs, hypotheses, and statistical models. For each analysis, I prepared a suitable data set, often from multiple data sources. Simple, well chosen descriptive statistics and graphs were often highly instructive, but when appropriate, I also used more elaborate statistical and machine-learning methods. In every case, I strove to present the results in lucid, practical terms. My primary computational tools were SQL, including PostgreSQL and MySQL; Python, including Jupyter, SciPy, pandas, scikit-learn, and Luigi; and Docker, Kubernetes, and Spark.
- As a Postdoctoral Fellow in the Duke University Biology Department, I conducted research in evolutionary genetics and genomics. For example, colleagues and I performed the first survey of promoter regions of human genes for evidence of adaptive evolution since the most recent common ancestor of humans and chimpanzees. We fitted (MLE, MCMC) statistical models to DNA sequences of these non-protein-coding, putatively gene-regulatory regions of the human, chimpanzee, and macaque genomes, and we found evidence for many adaptive changes in the human lineage, particularly in promoter regions of genes for proteins involved in neural development and function (Haygood et al., 2007). Subsequently, we performed a meta-analysis of surveys for adaptive changes in the human lineage, and we found that neural-related genes were prominent in surveys of noncoding regions but not in surveys of coding regions (Haygood et al., 2010). These findings affirm a long-standing conjecture that human cognition evolved mainly through changes in gene regulation. My primary computational tools were Ruby, R, and C
- As a Quantitative Analyst at Hydrologic Consultants, Inc. (of Sacramento, CA, later acquired by Bookman-Edmonston Engineering, Inc., later acquired by GEI Consultants, Inc.) and Timothy J. Durbin, Inc., I analyzed hydrologic data and situations for several clients. For example, I applied statistical methods (ANCOVA, MAP estimation) to streamflow measurements in order to reveal trends in water use within the North Platte River watershed, despite climatic fluctuations. Other

projects were less statistical and more mathematical. For example, I extended and applied proprietary numerical software (PDE solution via FEM) for modeling groundwater flow and solute transport in order to elucidate salt-water intrusion into an aquifer beneath Lompoc, CA. These analyses were implemented using FORTRAN, Excel, and Access.

Experience

Freelance software developer, 2012 –present.

Web application design, construction, and maintenance.

Data Science Developer, ReverbNation, 2014 –2017.

Applied statistics and machine learning in support of online services used by over four million musicians.

Founder, CardVine, 2009 –2011.

Development, promotion, and operation of a web application replacing business cards.

Postdoctoral Fellow, Biology Department, Duke University, 2005–2009.

Research in evolution, ecology, genetics, and genomics.

National Science Foundation Postdoctoral Fellowship in Biological Informatics, 2005–2006.

Postdoctoral Fellow, Department of Zoology, University of Wisconsin – Madison, 2002–2004.

Research in evolution, ecology, and genetics.

Graduate Student, Section of Evolution and Ecology, University of California, Davis, 1997–2002.

Coursework, research, and teaching in evolution, ecology, and genetics.

Merton Love Award for best dissertation on ecology, ethology, or evolution at UC Davis in 2002.

Quantitative Analyst, Hydrologic Consultants, Inc. / Timothy J. Durbin, Inc., 1996–2000.

Statistical and numerical analyses of surface-water and groundwater flows.

(This position was part-time, supplementing my graduate-student stipend.)

Graduate Student, Department of Mathematics, University of California, Davis, 1994–1997.

Coursework and teaching in mathematics.

(I fulfilled all the requirements for a Ph.D. in mathematics except the dissertation before transferring into population biology.)

Guest Researcher, Swedish Institute of Computer Science, 1992–1994.

Research and development in compilation techniques for logic programming languages.

Consulting Programmer, Department of Electrical Engineering, University of Southern California, 1991–1992.

Research and development in compilation techniques for logic programming languages.

Programmer/Analyst II, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, 1988–1991.

Research and development in compilation techniques for logic programming languages.

Graduate Student, Department of Physics, University of California, Santa Barbara, 1986 –1988.

Coursework and teaching in physics.

Education

Ph.D., Population Biology, University of California, Davis, 2002.

Merton Love Award for best dissertation on ecology, ethology, or evolution at UC Davis in 2002.

M.A., Physics, University of California, Santa Barbara, 1988.

B.S., Physics and Mathematics, University of California, Irvine, 1986.

Publications

A. Berrio, R. Haygood, and G. A. Wray, 2020. Identifying branch-specific positive selection throughout the regulatory genome using an appropriate proxy neutral. *BMC Genomics* **21**:359.

C. C. Babbitt, R. Haygood, W. J. Nielsen, and G. A. Wray, 2017. Gene expression and adaptive noncoding changes during human evolution. *BMC Genomics* **18**:435–445.

A. R. Ives, C. Paull, A. Hulthen, S. Downes, D. A. Andow, R. Haygood, M. P. Zalucki, and N. A. Schellhorn, 2017. Spatio-temporal variation in landscape composition may speed resistance evolution of pests to Bt crops. *PLOS ONE* **12**:e0169167.

D. A. Garfield, D. E. Runcie, C. C. Babbitt, R. Haygood, W. J. Nielsen, and G. A. Wray, 2013. The impact of gene expression variation on the robustness and evolvability of a developmental gene regulatory network. *PLOS Biology* **11**:e1001696.

D. Garfield, R. Haygood, W. J. Nielsen, and G. A. Wray, 2012. Population genetics of *cis*-regulatory sequences that operate during embryonic development in the sea urchin *Strongylocentrotus purpuratus*. *Evolution and Development* **14**:152–167.

O. Fedrigo, A. D. Pfefferle, C. C. Babbitt, R. Haygood, C. E. Wall, and G. A. Wray, 2011. A potential role for glucose transporters in the evolution of human brain size. *Brain, Behavior and Evolution* **78**:315–326.

T. A. Oliver, D. A. Garfield, M. K. Manier, R. Haygood, G. A. Wray, and S. R. Palumbi, 2010. Whole-genome positive selection and habitat-driven evolution in a shallow and a deep-sea urchin. *Genome Biology and Evolution* **2**:800–814.

R. Haygood, C. C. Babbitt, O. Fedrigo, and G. A. Wray, 2010. Contrasts between adaptive coding and noncoding changes during human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **107**:7853–7857.

C. C. Babbitt, J. S. Silverman, R. Haygood, J. M. Reininga, M. V. Rockman, and G. A. Wray, 2010. Multiple functional variants in *cis* modulate *PDYN* expression. *Molecular Biology and Evolution* **27**:465–479.

L. R. Warner, C. C. Babbitt, A. E. Primus, T. F. Severson, R. Haygood, and G. A. Wray, 2009. Functional consequences of genetic variation in primates on tyrosine hydroxylase (*TH*|) expression *in vitro*. *Brain Research* **1288**:1–8.

J. Tung, O. Fedrigo, R. Haygood, S. Mukherjee, and G. A. Wray, 2009. Genomic features that predict allelic imbalance in humans suggest patterns of constraint on gene expression variation. *Molecular Biology and Evolution* **26**:2047–2059.

R. Haygood and M. Turelli, 2009. Evolution of incompatibility-inducing microbes in subdivided host populations. *Evolution* **63**:432–447.

J. L. Walters, E. M. Binkley, R. Haygood, and L. A. Romano, 2008. Evolutionary analysis of the cis-regulatory region of the spicule matrix gene *SM50* in strongylocentrotid sea urchins. *Developmental Biology* **315**:567–578.

C. C. Babbitt, R. Haygood, and G. A. Wray, 2007. When two is better than one. *Cell* **131**:225–227.

R. Haygood, O. Fedrigo, B. Hanson, K.-D. Yokoyama, and G. A. Wray, 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nature Genetics* **39**:1140–1144.

B. W. Spitzer and R. Haygood, 2007. Migration load and the coexistence of ecologically similar sexuals and asexuals. *American Naturalist* **170**:567–572.

Sea Urchin Genome Sequencing Consortium, 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **314**:941–952.

R. Haygood, 2006. Mutation rate and the cost of complexity. *Molecular Biology and Evolution* **23**:957–963.

R. Haygood, 2004. Sexual conflict and protein polymorphism. *Evolution* **58**:1414–1423.

R. Haygood, A. R. Ives, and D. A. Andow, 2004. Population genetics of transgene containment. *Ecology Letters* **7**:213–220.

R. Haygood, A. R. Ives, and D. A. Andow, 2003. Consequences of recurrent gene flow from crops to wild relatives. *Proceedings of the Royal Society of London Series B, Biological Sciences* **270**:1879–1886.

R. Haygood, 2002. Coexistence in MacArthur-style consumer–resource models. *Theoretical Population Biology* **61**:215–223.

R. Haygood, 1994. Native code compilation in SICStus Prolog. P. Van Hentenryck (editor), *Proceedings of the Eleventh International Conference on Logic Programming*, MIT Press, pp. 190–204.

B. K. Holmer, B. Sano, M. Carlton, P. Van Roy, R. Haygood, W. R. Bush, A. M. Despain, J. M. Pendleton, and T. P. Dobry, 1990. Fast Prolog with an extended general purpose architecture. *Proceedings of the 17th International Symposium on Computer Architecture*, IEEE Computer Society Press, pp. 282–291.

Selected talks

(I've begun compiling materials related to the many talks I've given at conferences, workshops, and other meetings. The following is a sample.)

Interpreting decision-tree models, monthly meeting of Research Triangle Analysts, January 17, 2017.

Data science at ReverbNation, ReverbNation, June 29, 2016.

Selected software

(I’ve begun compiling various software I’ve created that may be useful to other people and that I’m legally entitled to distribute. The following is a sample.)

sklearn-gbmi, which provides a Python module for computing Friedman and Popescu's *H* statistics, in order to look for interactions among variables in scikit-learn gradient-boosting models.

Haygood et al., 2007 HyPhy-ware, which includes the HyPhy Batch Language files used to compute the results in Haygood et al., 2007 and an example of their use.

References

Dr. Devjanee Swain Lenz, Director, Sequencing and Genomic Technologies facility, Duke University, devjanee.swain.lenz@duke.edu.

Dr. Erik Soderblom, Director, Proteomics and Metabolomics facility, Duke University, erik.soderblom@duke.edu.

Dr. Gregory A. Wray, Professor, Biology Department, Duke University, gwray@duke.edu.