Investigating Causal Genetic Variation in the yellow Gene of Drosophila melanogaster as a Means of Teaching Foundational Molecular Genetic Concepts & Techniques

Robert B. Page, Matt Crook

Abstract
How genetic variation influences phenotypic variation is of importance to many biological disciplines, including evolutionary biology, biomedicine, and agriculture. Nevertheless, students frequently struggle to make connections across levels of biological organization, which can make it challenging to facilitate understanding of how nucleotide variation gives rise to organismal variation. At the same time, biology students are now expected to gain early experience with cornerstone techniques from molecular biology, so that these skills can be reinforced and expanded upon. Here we describe a five-to-seven-week sequencing project that examines genetic and phenotypic variation in wild-type and yellow-bodied fruit flies and, in the process, exposes students to several foundational techniques in molecular biology. In addition, students analyze partial yellow gene sequences from PCR products using the freely available bioinformatics suite UGENE and in doing so are introduced to core bioinformatics skills. The entire project is framed around the axiom that if the yellow gene controls phenotypic differences in body color between wild-type and yellow-bodied flies, it should be possible to identify causal variation in yellow sequences from wild-type versus yellow-bodied flies. This project relies on guided inquiry and can be used in 1000- or 2000-level molecular biology courses and advanced high school laboratories.

Key Words: bioinformatics; causal sequence variant; DNA sequencing; Drosophila melanogaster; guided inquiry; yellow body.

Introduction
The causal relationship between nucleotide variation in DNA and phenotypic variation at the organismal level is among the most fundamental phenomena in all biology. It is essential to biological evolution (Futuyma, 1998), plant and animal breeding programs (Walsh and Lynch, 1998), genetic engineering (Copeland et al., 2001; Doudna & Charpentier, 2014), the risk and occurrence of disease (Botstein & Risch, 2003; Santoro et al., 2016), and many other features of biological systems. However, teaching this foundational concept to lower-division biology students is often challenging, and there is evidence that biological variation is a threshold concept (sensu Meyer & Land, 2003) that students must grasp in order to deepen their understandings of other biological processes, such as natural selection (Ross et al., 2010; Walck-Shannon et al., 2019). In addition, a firm understanding of causal genetic variation requires integration of knowledge across topics that span several levels of biological organization, which can also be challenging for students to understand and instructors to convey (National Research Council, 2009; American Association for the Advancement of Science, 2011). While lower-division students are struggling with foundational biological concepts, they also face pressure to gain exposure to knowledge and skills that will give them a competitive advantage with professional/graduate programs and the job market (Page et al., 2018). Indeed, it is now standard practice to expose lower-division college students to cornerstone techniques from molecular biology, such as DNA isolation, electrophoresis, and Sanger sequencing, as early exposure to these methods serves as a foundation that can be built upon in upper-division courses. Moreover, given the role that high-throughput techniques play in modern life sciences research, it is also important to expose students to biological databases and computerized analysis of biological data early in their careers (e.g., Hoatling et al., 2018).

The common fruit fly, Drosophila melanogaster, is one of the most time-honored model organisms in genetics and has played a role in many seminal discoveries including gene linkage (Morgan, 1911), sex-linkage (Morgan & Bridges, 1916), and the roles of homeobox genes in animal development (Kaufman et al., 1980; Nüsslein-Volhard & Wieschaus, 1980). As such, its genome is

“Our objectives with this series of modules are to teach the link between phenotype and molecular genetics and to introduce students to the skills they need to succeed in upper division classes and the workplace.”
The yellow gene is pleiotropic, and loss of function leads to decreased pigmentation, viability, and male copulation success (Wilson et al., 1976). Yellow encodes a dopamine conversion enzyme (Heinze et al., 2017; Wittkopp et al., 2002; Xu et al., 2011) and has a readily available loss of function allele (y1) which is an A-to-C transversion that ablates the ATG start codon required for mRNA translation (Geyer et al., 1990). Thus there is a direct causal relationship between the y1 allele and the yellow-body phenotype via lack of yellow protein function.

Herein, we describe a guided-inquiry-based lab series that recapitulates the molecular characterization of a mutant allele that would occur once a genetic locus had been identified for a mutation generated in a forward genetic screen. This series entails generating, analyzing, and interpreting sequences from the 5’ end of the yellow gene derived from wild-type vs. yellow-bodied D. melanogaster. The entire project is framed around the axiom that if the yellow gene controls body color, then it should be possible to identify causal sequence variation between wild-type and yellow-bodied fruit flies in the yellow gene. This lab series can be conducted over five to seven weeks and exposes students to several foundational molecular biology techniques including DNA isolation, UV spectrophotometry, agarose gel electrophoresis, PCR, PCR purification, and Sanger sequencing. Furthermore, during the analysis and interpretation stage of the project, students are exposed to several core skills in bioinformatics and sequence analysis, such as working with biological databases, editing chromatograms, mapping reads to a reference sequence, and multiple sequence alignment. Lastly, in addition to considering causal sequence variation, students also consider eukaryotic gene structure and the consequences of point mutations in coding regions on gene expression.

**Concept-Based Learning Objectives**

- Define/explain what is meant by the phrase causal sequence variant.

**Techniques & Methods Introduced**

- DNA isolation
- Use of UV spectrophotometry to quantify and quality DNA isolates
- Use of gel electrophoresis to size separate DNA fragments
- Polymerase chain reaction (PCR)
- Use of spin column technology for purification of PCR products
- Sanger (dideoxy) sequencing
- Use of databases to obtain biological sequences and information about biological sequences
- Use of bioinformatic tools to process and analyze DNA sequences
  - Editing chromatograms
  - Mapping reads to a reference sequence
  - Using multiple sequence alignment to inspect nucleotide variation

**Lab Series Design**

**Overview**

This inquiry-based exercise is designed to be completed in five to seven weeks and is presented in six modules, beginning with DNA isolation and ending with sequence analysis, that mimic workflows routinely carried out in research labs (Table 1). It does not presume anything about students’ molecular biology skills beyond reasonable competency with micropipettes. We have found that the exercise presented by Boker (2012) is useful in helping students develop accuracy with and confidence in their pipetting.

| Module               | Length                          | Supplemental Materials                           |
|----------------------|---------------------------------|--------------------------------------------------|
| Phenotyping          | One 1 hr. session               | • Protocol 1                                     |
| DNA isolation        | One 2–3 hr. session             | • Protocol 2                                     |
| Isolate quality control | One 2–3 hr. session             | • Protocol 3                                     |
| PCR                 | One 1–3 hr. session             | • Protocol 4                                     |
| PCR purification     | One 2–3 hr. session             | • PCR presentation slides                        |
| Sequence analysis    | One to three 2–3 hr. sessions (depending on endpoint) | • Protocol 6                                     |
|                      |                                 | • Sanger sequencing presentation slides          |
|                      |                                 | • yellow sequences zip archive                   |
Module 1: Exploring the yellow Phenotype

Linking phenotype and genotype is the key aim of this series of modules, so the first task is to explore the phenotypic differences between wild-type and yellow-bodied animals. The two objectives of this module are to familiarize students with the use of a stereo dissecting microscope and to use this microscope to compare the body color differences between wild-type and yellow-bodied animals.

Module 2: DNA Isolation

The primary goals of this module are twofold. First, as can be seen in Supplemental Protocol 2, students are made aware of the fact that there are several ways to perform DNA isolations but that all of these approaches consist of a series of physiochemical steps that facilitate specific occurrences (e.g., release of DNA from cells and removal of protein). Second, students are provided with a protocol that will enable them to isolate DNA from whole fruit flies. The protocol that we have provided is for use with Qiagen's DNeasy Blood and Tissue Kit (product 69504), as we have found that this kit's streamlined approach maximizes the likelihood of students generating isolates of sufficient quality for downstream analyses. However, minimally toxic protocols based on ammonium acetate precipitation (e.g., Fetzner, 1999, rescaled for cell-rich tissues from small arthropods) are less expensive and potentially just as effective.

Module 3: Isolate Qualification and Quantification

As can be seen in Supplemental Protocol 3, the third module in this lab series introduces students to UV spectrophotometry as a means of quantifying the yield obtained via DNA isolation and the purity of the isolates. As such, this lab introduces students to the basic relationship between absorbance and concentration via the Beer-Lambert Law. The procedure that we provide in Protocol 3 is for use with a microvolume spectrophotometer, such as a NanoDrop (Thermo Scientific); however, in principle, any instrument that measures absorbance at wavelengths between 200 and 300 nm would be sufficient. In addition, this lab teaches students how to use agarose gel electrophoresis to assess the degree to which their DNA isolates have been degraded. Lastly, students make dilutions suitable for loading template DNA into PCRs using their spectrophotometer results and the equation \( V_1 C_1 = V_2 C_2 \).

Module 4: Polymerase Chain Reaction

In this module, students are introduced to PCR and its many applications. In addition to providing students with the background material in Supplemental Protocol 4, we give a 20–30 minute presentation on the mechanics of PCR, including how real-time PCR can be used for viral and mRNA quantification. Students then work together to produce a master mix that is aliquoted into 0.2 ml tubes along with 2 µl of the 10 ng/µl dilution of template DNA that they will have made at the end of the previous session. The PCR recipe and thermal cycler conditions that we provide in Protocol 4 have been thoroughly tested using the Promega GoTaq Flexi Polymerase kit (product M8295) and dNTP mix (product U1515), as well as the T100 thermal cycler that is available from BioRad. However, use of other reagents and equipment will likely only require minor, if any, adjustments. We designed the primers described in Protocol 4 to target a region of yellow that includes a portion of the 5’ UTR, all of the first exon, and a portion of the first intron, which is appropriate given that the mutation we are assessing is located in the gene’s start codon. We typically use 10 micromolar as our working primer concentration, and the primer volumes in Protocol 4 assume that the primers are at this concentration.

Module 5: PCR Verification & Cleanup

The primary objectives of this module are to (1) determine which PCRs from the previous session were successful, (2) remove impurities from the PCR products to make them suitable for Sanger sequencing, and (3) determine the concentration of the purified PCR product via UV spectrophotometry. In the interest of time, we typically have students set up their gels first and move on to the PCR purification protocol while their gels are running. The procedure we have given in Supplemental Protocol 5 is for use with the Wizard SV Gel and PCR Clean-Up System available from Promega (product A9281); however, in principle, any PCR cleanup procedure could be used in this step.

Module 6: Sanger Sequencing & Sequence Analysis

Between this module and the previous module, we typically submit students’ purified PCR products to a genomics facility for Sanger sequencing. However, after doing this over the course of several semesters we have amassed several high-quality wild-type and mutant sequences that are suitable for students to work with, and these are included in the supplemental materials. As such, submission of student PCRs for sequencing is optional. Irrespective of whether student PCRs are sequenced or analyses are based on the ABI files we provide, it is advisable that students be introduced to the mechanics of Sanger sequencing, which we typically do via a 20–30 minute presentation. However, exercises such as the one offered by Conley et al. (2016) may be used to further emphasize key features of dideoxy sequencing.

Once the students have an appreciation for cycle sequencing reactions, capillary electrophoresis, and electropherograms, they use FlyBase, the NCBI website, and the freely available bioinformatics suite UGENE (Okonechnikov et al., 2012), to explore and analyze their sequences. Guidance for students on how to use these tools is given in Supplemental Protocol 6. Ultimately, the work students perform when completing this project can have one of two endpoints. The first possible endpoint is answering the question set provided in Protocol 6. The second possible endpoint is production of a scientific manuscript (i.e., introduction, methods, results, and discussion) based on this lab series.

Discussion

Our objectives with this series of modules are to teach the link between phenotype and molecular genetics and to introduce students to the skills they need to succeed in upper division classes and the workplace. We chose D. melanogaster as our model because it is easy to obtain and maintain, has a large number of well-characterized visible mutant phenotypes, has a wealth of genetic information available, and is a model organism used by research laboratories to study a wide range of biological questions. We chose the yellow gene in large part because of how well the genetic basis for this phenotype, an A to C transversion in the start codon (Geyer et al., 1990), ties into the larger concept of gene expression, which underpins a large part of modern genetics. The modules form a series of guided inquiries utilizing the same approach that geneticists use to determine the molecular basis of mutant alleles recovered from a forward genetic screen. Each module is part of a whole but also stands alone as a functioning protocol.
Because misconceptions about genetics are common among college students and the general public (Stern & Kampourakis, 2017), it is crucial for instructors to be sensitive to the possibility that students will draw conclusions or make comparisons that were not intended. Given that this lab series examines a body color phenotype that results from a loss of function mutation in a gene that acts pleiotropically to affect behavior, these exercises should be conducted within the context of a course that actively seeks to disabuse students of any essentialist notions that they may have. Given that biological essentialism can be reinforced by examples that strongly link race to the alleles present at a single locus (e.g., sickle cell anemia in African Americans) and is associated with misunderstandings of intraspecific diversity (Donovan, 2015), it is important to convey to students that most traits in natural populations exhibit complex modes of inheritance and are not under single gene control. In addition to introducing students to foundational concepts from quantitative genetics (e.g., polygenes, additivity, and genotype-by-environment interaction), it should be emphasized that while molecular mechanisms are frequently evolutionarily conserved, the genetic and physiological circuitries of humans and other animals differ (Bolker, 2019; Greek & Rice, 2012).

The lab series we describe has worked well and is presented here in its final form after three years of iterative development and testing. Student feedback about the lab series tends to reflect their engagement with a longer duration project that deals with a “big picture idea” rather than a series of unlinked lab exercises or specific modules such as gel electrophoresis or Sanger sequencing. We feel that this lab series would be an excellent model organism module to run in series with others, such as *Caenorhabditis elegans* RNAi (Sengupta, 2013) or *Arabidopsis thaliana* breeding modules (Price et al., 2018). An alternative would be to pair this lab series with other *D. melanogaster* experiments, such as studying the X-linked inheritance of the yellow and white genes (Lobo & Shaw, 2008) or using a reverse genetics approach to conditionally knock down yellow in a wild-type background (Dietzl et al., 2007). In addition, a simple assay in which yellow-bodied vs. wild-type males compete for mates would give students the opportunity to study the influence of yellow on a complex trait—namely male courtship / copulatory success (Massey et al., 2019).

### Suppemental Materials

The protocols for Modules 1–6—containing background and procedures, PowerPoint slides describing PCR and Sanger sequencing, and yellow.ABI sequencing files—are available at https://jaguar-my.sharepoint.com/:f:/g/personal/rpage_tamusa_edu/EvlZrsiSoXIEu rcWDeziQYBEfxbZ95J6slCHwkwX7xdpH?e=uASYSL

### Acknowledgments

We thank T. Marie Tipps for assistance during the early phases of development of this lab series.

### References

American Association for the Advancement of Science. (2011). *Vision and Change in Undergraduate Biology Education*. Washington DC: AAAS.

Bolker, J. (2012). *You sank my … bacteriophage?* *American Biology Teacher*, 74, 422–23.

Bolker, J. A. (2019). *Selection of models: evolution and choice of species for translational research*. *Brain, Behavior and Evolution*, 93, 82–91.

Botstein, D. & Risch, N. (2003). *Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease*. *Nature Genetics Supplement*, 33, 228–37.

Conley, J.E., Meisel, A.J. & Smith, J.J. (2016). Using M&Ms to model Sanger’s dideoxy DNA sequencing method. *American Biology Teacher*, 78, 516–22.

Copeland, N.G., Jenkins, N.A. & Court, D.L. (2001). Recombining: a powerful new tool for mouse functional genomics. *Nature Reviews Genetics*, 2, 769–79.

Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Melliner, G., Basser, K., Kinsey, K., Oppel, S., Scheiblauer, S., Couto, A., Marra, V., Kelemen, K., Dickson, B.J. (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature*, 448, 151–56.

Donovan, B.M. (2015). Reclaiming race as a topic of the U.S. biology textbook curriculum. *Science Education*, 99, 1092–117.

Doudna, J.A. & Charpentier, E. (2014). *CRISPR-Cas9: The new frontier of genome editing*. Nature, 527, 43–48.

Fetzner, J.W. (1999). Extracting high-quality DNA from shed reptile skins: a simplified method. *Biotechniques*, 26, 1052–54.

Futuyma, D.J. (1998). *Evolutionary Biology, 3rd Edition*. Sunderland, MA: Sinauer Associates.

Geyer, P.K., Green, M.M & Corces, V.G. (1990). Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in *Drosophila*. *EMBO Journal*, 9, 2247–56.

Greek, R. & Rice, M.J. (2012). Animal models and conserved processes. *Theoretical Biology and Medical Modelling*, 9, 40.

Heinze, S.D., Kohlbrenner, T., Ippolito, D., Meccariello, A., Burger, A., Mosimann, C., Saccoone, G. and Bopp, D. (2017). CRISPR-Cas9 targeted disruption of the yellow ortholog in the housefly identifies the brown body locus. *Scientific Reports*, 7, 4582.

Hoatling, S., Slabach, B.L. & Weisrock, D.W. (2018). Next-generation teaching: a template for bringing genomic and bioinformatic tools into the classroom. *Journal of Biological Education*, 52, 301–13.

Kaufman, T.C., Lewis, R. & Wakimoto, B. (1980). Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: the homeotic gene complex in chromosome interval 8A-A-B. Genetics, 94, 115–33.

Lobo, I. & Shaw, K. (2008). Thomas Hunt Morgan, genetic recombination, and gene mapping. *Nature Education*, 1(1), 205.

Massey, J.H., Chung, D., Siwanowicz, I., Stern, D.L. & Wittkopp, P.J. (2019). The yellow gene influences *Drosophila* male mating success through sex comb melanization. *eLife*, 8, e49388.

Meyer, J.H.F. & Land, R. (2003). Threshold concepts and troublesome knowledge: linkages to ways of thinking and practicing within the disciplines. In International Improving Student Learning Symposium, Rust, C. & Oxford Centre for Staff Development. *Improving student learning theory and practice - 10 years on: Proceedings of the 2002 10th International Symposium Improving Student Learning*. Oxford: Oxford Centre for Staff & Learning Development.

Morgan, T.H. (1911). Random segregation versus coupling in Mendelian inheritance. *Science*, 39, 384.

Morgan, T.H. & Bridges, C.B. (1916). *Sex-linked Inheritance in Drosophila*. Washington DC: Carnegie Institution.

National Research Council. (2009). *A New Biology for the 21st Century*. Washington DC: National Academies Press.

Nüsslein-Volhard, C. & Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 287, 795–801.

Okonechinskikov, K., Goloysa, O., Fursov, M. & the UGENE team. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, 28, 1166–67.
Page, R.B., Espinosa, J., Mares, C.A., Del Pilar, J. & Shelton, G.R. (2018). The curvy road to student success in underserved populations. *Journal of College Science Teaching, 47*(5), 6–7.

Price, C.G., Knee, E. M., Miller, J.A., Shin, D., Mann, J., Crist, D.K., Grotewold, E. & Brikjacic, J. (2018). Following phenotypes: an exploration of Mendelian genetics using *Arabidopsis* plants. *American Biology Teacher, 80*(4), 291–300.

Ross, P.M., Taylor, C.E., Hughes, C., Whitaker N., Lutze-Mann, L., Kofod, M. & Tzioumis, V. (2010). Threshold concepts in learning biology and evolution. *Biology International, 47*, 47–52.

Sanctoro, M.L., Moretti, P.N., Pellegrino R., Gadelha, A., Abilio V.C., Hayashi, M.A.F., Belangero, S.I. & Hakonarson, H. (2016). A current snapshot of common genomic variants contribution in psychiatric disorders. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 171*, 997–1005.

Sengupta, S. (2013). Bringing RNA Interference (RNAi) into the high school classroom. *American Biology Teacher, 75*(9), 698–703.

Stern, F. & Kampourakis, K. (2017). Teaching for genetics literacy in the post-genomic era. *Studies in Science Education, 53*, 193–225.

Walck-Shannon, E., Batzli, J., Pultorak, J. & Boehmer, H. (2019). Biological variation as a threshold concept: can we measure threshold crossing? *CBE—Life Sciences Education, 18*, ar36.

Walsh, B. & Lynch, M. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates.

Wilson, R., Burnet, B., Eastwood, L. & Connolly, K. (1976). Behavioural pleiotropy of the yellow gene in *Drosophila melanogaster*. *Genetics Research, 28*, 75–88.

Wittkopp, P.J., True, J.R. & Carroll, S.B. (2002). Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development, 129*(8), 1849–58.

Xu, X., Oliveira, F., Chang, B.W., Collin, N., Gomes, R., Teixeira, C., Reynoso, D., My Pham, V., Elnaiem, D.E., Kamhawi, S., Ribeiro, J.M., Valenzuela, J.G. & Andersen, J.F. (2011). Structure and function of a “yellow” protein from saliva of the sand fly *Lutzomyia longipalpis* that confers protective immunity against *Leishmania major* infection. *Journal of Biological Chemistry, 286*(37), 32383–93.

**Robert B. Page, PhD, is an associate professor of biology, and Matt Crook, PhD, is an assistant professor of biology, both in the Department of Life Sciences at Texas A&M University—San Antonio, San Antonio, TX; e-mails: Robert.Page@tamusa.edu and Matthew.Crook@tamusa.edu.**