Turning randomness into meaning at the molecular level using Muller's morphs

Kathleen Henson[1,*], Melanie M. Cooper[2] and Michael W. Klymkowsky[3,‡]

[1]School of Education, University of Colorado, Boulder, CO 80309, USA
[2]Department of Chemistry, Clemson University, Clemson, SC 29634, USA
[3]Molecular, Cellular and Developmental Biology and CU Teach, University of Colorado, Boulder, Colorado, 80309-0347, USA

*Present address: Miss Porter's School, 60 Main Street, Farmington, CT 06032, USA
‡Author for correspondence (michael.klymkowsky@colorado.edu)

Biology Open 1, 405–410
doi: 10.1242/bio.2012031

Summary

While evolutionary theory follows from observable facts and logical inferences (Mayr, 1985), historically, the origin of novel inheritable variations was a major obstacle to acceptance of natural selection (Bowler, 1992; Bowler, 2005). While molecular mechanisms address this issue (Jablonka and Lamb, 2005), analysis of responses to the Biological Concept Inventory (BCI) (Klymkowsky et al., 2010), revealed that molecular biology majors rarely use molecular level ideas in their discourse, implying that they do not have an accessible framework within which to place evolutionary variation. We developed a "Socratic tutorial" focused on Muller's categorization of mutations' phenotypic effects (Muller, 1932). Using a novel vector-based method to analyzed students' essay responses, we found that a single interaction with this tutorial led to significant changes in thinking toward a clearer articulation of the effects of mutational change. We suggest that Muller's morphs provides an effective framework for facilitating student learning about mutational effects and evolutionary mechanisms.

Key words: Mutation, Phenotype, Evolution, Randomness, Effective learning

Introduction

The inability to explain the mechanism by which inheritable variations are generated was a major contributor to the "eclipse of Darwinism" at the turn of the 20th century. As Bowler notes, "*It was precisely the unpredictability of the sequence of events in Darwin's universe which made the worldview which we now call "Darwinism" so unacceptable to the late nineteenth-century mind.*" (Bowler, 1992). This conceptual chasm closed, at least for scientists, with the synthesis of evolutionary thinking, population genetics, and molecular biology. As stated by Muller "*there remains no reason to doubt the application of the dictum 'all life from pre-existing life' and 'every cell from a pre-existing cell,' to the gene: 'every gene from a pre-existing gene'*" (Muller, 1936).

That said, Mendel's insights were based on highly artificial conditions; a Mendelian dream-world in which in-bred strains (common genetic backgrounds) and strongly expressive and highly penetrant alleles produced distinctive phenotypes. In the real world, the connection between genotype and phenotype is generally more complex. Genetic background effects and the stochastic and interconnected nature of biological systems, responsible for their robustness (Wagner, 2005), lead to situations in which incomplete expressivity and variable penetrance are the rule rather than the exception. Moreover, evolutionary innovation generally involves gene duplication rather than a static genome (Bergthorsson et al., 2007), a fact rarely emphasized in most genetics or molecular biology textbooks. This makes for conceptual obstacles when connecting abstract genetic rules to biological realities. All too often students are taught, and remember, iconic systems such as sickle cell anemia, cystic fibrosis, lactose tolerance, or the *lac* operon, that fail to provide a coherent framework within which to analyze new and significantly more complex systems as well as the origins of phenotypic novelty.

At the same time, students' difficulties are not due simply to "crimes of omission". Many, if not most, foundational concepts of science are counter-intuitive, e.g. the ideas of genes, molecules, these random motions, and events (Garvin-Doxas and Klymkowsky, 2008; Klymkowsky, 2011) and are often "delivered" to students through passive instruction such as lecture (Havighurst, 1929; Powell, 2003). It is not surprising, therefore, that at its root, much student confusion is didaskalogenic, that is instruction/instructor-induced (Nersessian, 1989; Taber, 2001). Students commonly have misconceptions about how mutations affect phenotype and questions that address the origins of variation are challenging for them to answer correctly (Gregory, 2009). At the same time genetics concept-based assessments do not explicitly examine the molecular origins of genetic variation and new phenotypes (Anderson et al., 2002; Bowling et al., 2008; Smith et al., 2008). Not surprisingly students, like the scientists who turned to "*non-Darwinian mechanisms such as Lamarckism, orthogenesis and saltationism*" (Bowler, 2005), tend to embrace active, i.e. purposeful drivers. While natural selection is not itself a random process, its effects are often modulated by random events (such as accidental death); that said, in addition to mutation, evolutionary trajectories are influenced by what are

essentially random processes such as founder effects and population bottlenecks (Garvin-Doxas and Klymkowsky, 2008). The embarrassingly low level of acceptance of the validity of biological evolution by the American public (Berkman and Plutzer, 2011) raises an obvious question: to what degree are students presented with a coherent framework (Donovan and Bransford, 2005) within which to develop the conceptual sophistication required to deal with what are, in fact, quite complex processes and ideas? If students were to consider the nature of mutations and alleles in the same way in which the pioneers of genetics did, would their understanding increase?

Results and Discussion

To explore this question in greater depth, we drew on data obtained during the development and administration of the Biological Concepts Instrument (BCI) (Klymkowsky et al., 2010). At the heart of the problem of understanding the molecular basis of evolutionary novelty is the question "How might a mutation be creative?" since mutations, through their effects on the time, place, and level of gene expression or the activities of the gene product, whether polypeptide, regulatory, or structural RNA, provide the raw material from which new phenotypes arise. The only logical choice on offer (BCI question 14) is "If the mutation altered the gene product's activity" since creative implies new and new implies an alteration in the gene product. Nevertheless, typically only ~40% of students (and 62.7% of surveyed teachers) selected this response (Fig. 1). Students' responses did not change significantly with their progression through a molecular biology curriculum (Klymkowsky et al., 2010) and similar results were obtained with students at a major European University (see http://www.educ.ethz.ch/modern_biology/program/ppt_Klymkowsky.pdf). Since this choice is not based on sophisticated biological reasoning, but rather the acceptance of the idea that "creative" implies "new", this observation suggests that there are significant conceptual barriers captured in the distractors "If the mutation inactivated a gene that was harmful" and "If the mutation had no

BCI Question	ave. ± σ	teachers
Q14: How might a mutation be creative? a. It could not be; all naturally occurring mutations are destructive. b. If the mutation inactivated a gene that was harmful. **c. If the mutation altered the gene product's activity.** d. If the mutation had no effect on the activity of the gene product.	41.8 ± 7.4	62.7
Q21: A mutation leads to a dominant trait; what can you conclude about the mutation's effect? a. It results in an overactive gene product. b. It results in a normal gene product that accumulates to higher levels than normal. c. It results in a gene product with a new function. **d. It depends upon the nature of the gene product and the mutation.**	33.0 + 11.0	61.4
Q24. A mutation leads to a recessive trait; what can you conclude about the mutation's effect? a. It results in a non-functional gene product. b. It results in a normal gene product that accumulates to lower levels than normal. c. It results in a gene product with a new function. **d. It depends upon the nature of the gene product and the mutation.**	29.5 + 4.8	64.9
Q27. Consider a diploid organism that is homozygous for a particular gene. How might the deletion of this gene from one of the two chromosomes produce a phenotype? a. If the gene encodes a multifunctional protein. **b. If one copy of the gene did not produce enough gene product.** c. If the deleted allele were dominant. d. If the gene encoded a transcription factor.	23.3 +4.8	58.1

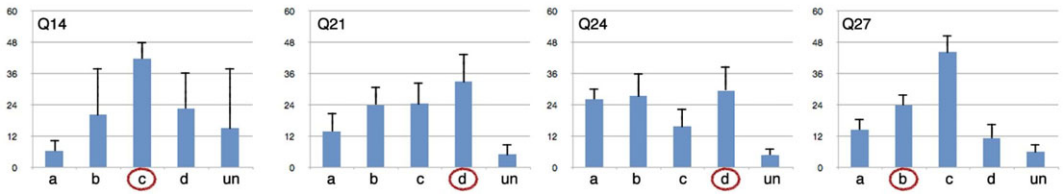

Fig. 1. BCI based insights into student "genetic" thinking. The table (top) illustrates the overall correct responses (in bold) to four questions from the BCI. The graphs (below) and the right two columns of the table reflect responses to these same questions by a group of fourteen introductory college level biology classes (a total of 2197 students) at seven different colleges and universities, administered over the period from 2006 to 2011, together with the responses from a group of 85 middle and high school teachers who answered the BCI online in response to an email to the National Science Teachers Association biology listserv. The graphs display the responses to the questions (a,b,c,d and unanswered "un") together with the standard deviation for each response. The correct choices are indicated by red circles.

effect on the activity of the gene product", that lure students away from the correct response. Three BCI questions are relevant to the relationship between evolutionary mechanisms and genetic and molecular level processes: questions 21 and 24, address the relationship between dominant and recessive alleles and molecular level mechanisms while question 27, examines the relationship between gene number and molecular effects (Fig. 1). In each case, fewer than half of the students responding (and less than 65% of teachers) selected the correct response. Since these questions deal directly with the nature of dominant (Q21) and recessive (Q24) mutations and their phenotypes, as well as the molecular level effects of a deletion (Q27), at the very least, these results suggest a level of uncertainty as to how to interpret mutational effects.

In the light of these observations, and our own experiences with the inability of upper division students to connect gene activity, genetic variation, and phenotype, we considered the possibility that students simply have not been supplied, or have not incorporated into their thinking, a coherent, robust and generalized framework within which to think about mutations and their effects. In that light, it seemed that Muller's work on mutations and phenotypic traits (Muller, 1932), might provide just such a framework; it places all mutations with discernible phenotypes into five distinct categories linked to gene activity: amorphic (no function), hypomorphic (reduced function), hypermorphic (increased function), antimorphic (antagonistic function), or neomorphic (new function). If we include the class of mutation with no effect on phenotype, we have captured all possible mutational effects. Together with an explicit recognition that existing genes do not appear *de novo*, but are derived from pre-existing genes through duplication and various types of recombination events (see above), Muller's morphs could provide students with a relatively simple and coherent framework within which to consider the effects of mutational change. A review of six popular introductory genetics textbooks reveals that gene duplication is either not mentioned, or is restricted to a page or two toward the end of the books, while Muller's classification of the phenotypic effects of mutations, does not occur at all. While there are mentions of "loss of function" or "gain of function", these are not presented within a coherent molecular, cellular or physiological framework. How students interpret these phrases is not clear, and deserves some consideration - for example, are we talking about loss of one or multiple functions? Moreover, such statements imply that gene products have a single function, which is rarely the case.

Considering these issues, and influenced by a reading of "How Students Learn" (Donovan and Bransford, 2005), we developed a formative "Socratic tutorial" targeted on mutations and their effects. Its design emphasizes scenarios that encourage metacognition, student self regulation and cooperative learning - a Socratic style of learning (video of students working through the activity is available at the http://besocratic.colorado.edu website). As a generic control and to serve as a comparison for working with such tutorials, we developed a second tutorial on the topic of "Graphical Thinking" (see below; the "Graphical Thinking" tutorial is available at http://besocratic.colorado.edu/Graph-thinking/graphical-home.htm). The "Mendel's Factors and Muller's Mutations" tutorial is available at http://besocratic.colorado.edu/mutations/mutations-home.htm (supplementary material Table S1 provides examples of students responses to the tutorial). The responses to these tutorials were captured in an on-line database. Because of privacy concerns, we cannot allow free access to student responses. Instructors interested in using these materials need only give their students a course code, this will enable us to provide instructors with a detailed summary of their (anonymous) students' responses upon request. While the activities, particularly the Graphical Thinking activity, is currently best performed on paper, we are in the process of transforming both activities into ones that will be accessible through our interactive, graphical input based beSocratic platform (http://beSocratic.clemson.edu).

The first tutorial, "Mendel's Factors and Muller's Mutations", was intended as the experimental treatment. It begins (page 1) by asking students to reflect on what Mendel knew about his factors in physical or molecular terms, and then introduces (page 2) Muller's discovery that X-rays could induce mutations, asking students to reflect on what that observation implied about the nature of genes and gene products. Page 3 introduces the idea of "gene product" and how mutations affect gene product activity; asking students to explain some of their ideas about that relationship. Page 4 introduces the logic of Muller's experimental approach, based on the ability to generate deletions and duplications in *Drosophila*; students are asked to think about exceptions to the assumption that a deletion or duplication always produces either a 1/2 or 2-fold change, respectively, in the level of gene product. Page 5 introduces the amorphic and hypomorphic classes of mutation, and asks how a mutation might "reduce but not eliminate or change the activity of a gene product". Page 6 introduces the hypermorphic class of mutation/allele and asks how a mutation could increase activity. Pages 7 and 8 introduce the antimorphic and neomorphic classes and asks how in molecular terms these could come about. Page 9 serves as a review, and asks students to reflect on "How do mutations generate novel structures and behaviors?" and "Why might the random nature of mutations lead some people to reject biological evolution?"

The second tutorial, "Graphical Thinking", was designed to serve as a control for the effects of working on-line on a tiered activity, similar in structure to "Mendel/Muller". It focusses on the meaning of graphical data, specifically when the points on a graph should be connected, or used to determine a "best fit" line or curve. The tutorial walked students through a series of questions about data and graphs and ended with an opportunity for students to summarize and review their ideas. In the context of our studies, this tutorial served as a control for content, rather than for the effects of engaging students in a metacognitive, collaborative activity. That said, Graphical Thinking does not control for the simple act of introducing the topics of mutations, phenotype, and molecular level phenomena to students and we are not claiming that the effects observed following students working with Mendel/Muller are necessarily due to the specific structure of the activity. We therefore refer to the students who were assigned to the Graphical Thinking group as the comparison, rather than the control group.

Both tutorials were designed to be used in a small group discussion context. Students worked through the tutorials independent of instructor feedback. In the case of Mendel/Muller, we piloted the tutorial (Fall, 2009), videotaped student groups working through the tutorial, analyzed their responses, and made minor revisions where necessary to clarify the tutorial for students. To test the efficacy of Mendel/Muller as a tool for increasing student understanding of the basics of molecular/phenotypic novelty, we introduced the tutorial (Spring, 2010) in

an undergraduate genetics laboratory course at a large midwestern public university and collected preliminary data from student responses (273 students). All students were given a survey (supplementary material Table S2) before working with either tutorial. The survey consisted of five questions, two derived from the BCI address the topics of natural selection and mutational effects, two newly generated questions on the interpretation of graphs; and a fifth that asked students to produce an open-ended response to the question "How might a mutation be creative?". The 13 laboratory sections were assigned at random to work through either the Mendel/Muller (treated group) or Graphical Thinking (comparison group) tasks; both tasks took approximately 30 to 40 minutes to complete (video examples of a group of students working through the Mendel/Muller activity is available at the http://besocratic.colorado.edu website). The random assignment was carried out on the section level rather at than at the level of individual students and therefore does not account for real differences that may have existed between sections. Four weeks later, each section was given the same survey.

Students' written responses to the "How might a mutation be creative?" question were analyzed based on what we believe to be a novel and easy to use coding strategy designed to capture the full complexity of student thinking (Fig. 2). Responses were analyzed by two readers. The readers disagreed on fewer than ten percent of all responses and resolved conflict through discussion. Each written response was coded as correct (and the extent of correctness) as increasing values along the X-axis. A response was given X=+1 if it contained evidence for "new function" and X=+2 if it indicated an understanding of "new phenotype". For example, the response "It changes a base in a DNA sequence which in turn alters the structures and behaviors or phenotype of the cell" would have earned a score of (2,0). An X value of −1 was given if the student explicitly denied that mutations could produce new functions (a very rare response). Responses that consisted of, or contained extraneous language, which we term "cant", that is, "to use pretentious language, barbarous jargon, or technical terms; to talk with an affectation of learning" (Webster's Dictionary), which equates with "conceptual noise",

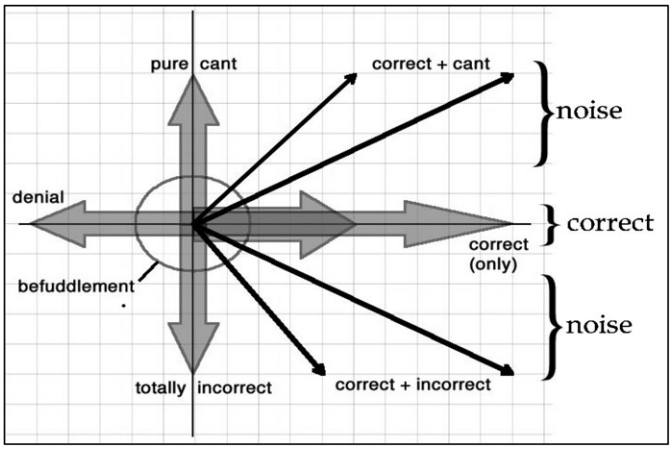

Fig. 2. Vector analysis of student responses. This illustrates our scheme for visualizing the clarity, correctness, and confusion present in student responses. The widths of arrows (and the diameter of the circle centered around 0,0) reflect the number/percentage of students in that group. Arrows that fall off the X axis contain aspects of correctness and either cant or mistakes.

were plotted as positive +1 on the Y axis. Misconceptions were assigned a Y=−1. For example the following response, "When it snows and you aren't able to survive in the cold you need to mutate to survive" represents a misconception that mutations occur based on need. This response would have received a score of (0, −1) No effort was made to categorized the "size" of misconceptions or the irrelevance of cant responses: this could, in theory, be done, leading to Y values greater than 1 or −1. Responses that contained both correct and either cant or misconceptions were placed in (1,1), (2,1), (1, −1) or (2, −1) bins. It is possible for responses to contain correct ideas together with both cant and misconceptions; a response that contained both cant and misconceptions was assigned to the misconception bin (Y=−1). Clearly a more accurate alternative would be to plot misconceptions and cant on the Y and Z axes, respectively, so as to generate a three- rather than a two-dimensional characterization of student responses. We chose not to do this here for sake of visual simplicity. Responses that could not be readily characterized were assigned scores of 0,0 and their frequency is indicated by the diameter of the circle. The frequency of other classes of responses is indicated by vector thickness. All responses were decoded "blind", before we had determined which of the treatments, Mendel/Muller (experimental) or Graphical Thinking (comparison) the students had been assigned. In this coding scheme, success, which includes answers that are correct, complete, free of cant and misconceptions, would produce an X axis vector of length +2 (Fig. 2).

The major differences between the experimental Mendel/Muller (N=134) and the comparison group, Graphical Thinking (N=139) (Fig. 3) was a significant decrease in the percentage of students answering the question with cant (Y=1) and a significant increase in the percentage of students indicating a novel phenotypical (X=2) effect in the experimental Mendel/Muller group. The students in both groups experienced significant drops in the use of cant answers (control; $p<0.003$, treatment; $p<0.002$, respectively). The students in the treatment group, however, experienced a significant increase in their use of answers indicating a novel phenotype while students in the control group did not demonstrate such an increase ($p<0.001$, $p<0.291$ respectively). It should be noted that both groups were enrolled in a genetics laboratory course, most in the associated lecture course, and both showed similar increases in the idea of new effects/functions (X=+1). Since both tutorials involve collaborative exercises, general effects of metacognitive practice could influence the results; we have not examined "untreated" students. That said, a single, short introduction to Muller's characterization of mutant traits/phenotypes appears to have had a measurable effect, and one would be tempted to speculate that a more concerted and continuous emphasis on providing students with a coherent conceptual framework within which to think about mutational effects could have even greater (desirable) effects.

A common problem with many educational interventions is that they are designed to provoke a desired response; many students are capable of recognizing what is expected of them. This is, in fact, one lesson to be derived from the Force Concept Inventory (Heller and Huffman, 1995; Hestenes and Halloun, 1995; Hestenes et al., 1992), our studies on student understanding of stochastic processes (Garvin-Doxas and Klymkowsky, 2008; Klymkowsky and Garvin-Doxas, 2008), chemical concepts

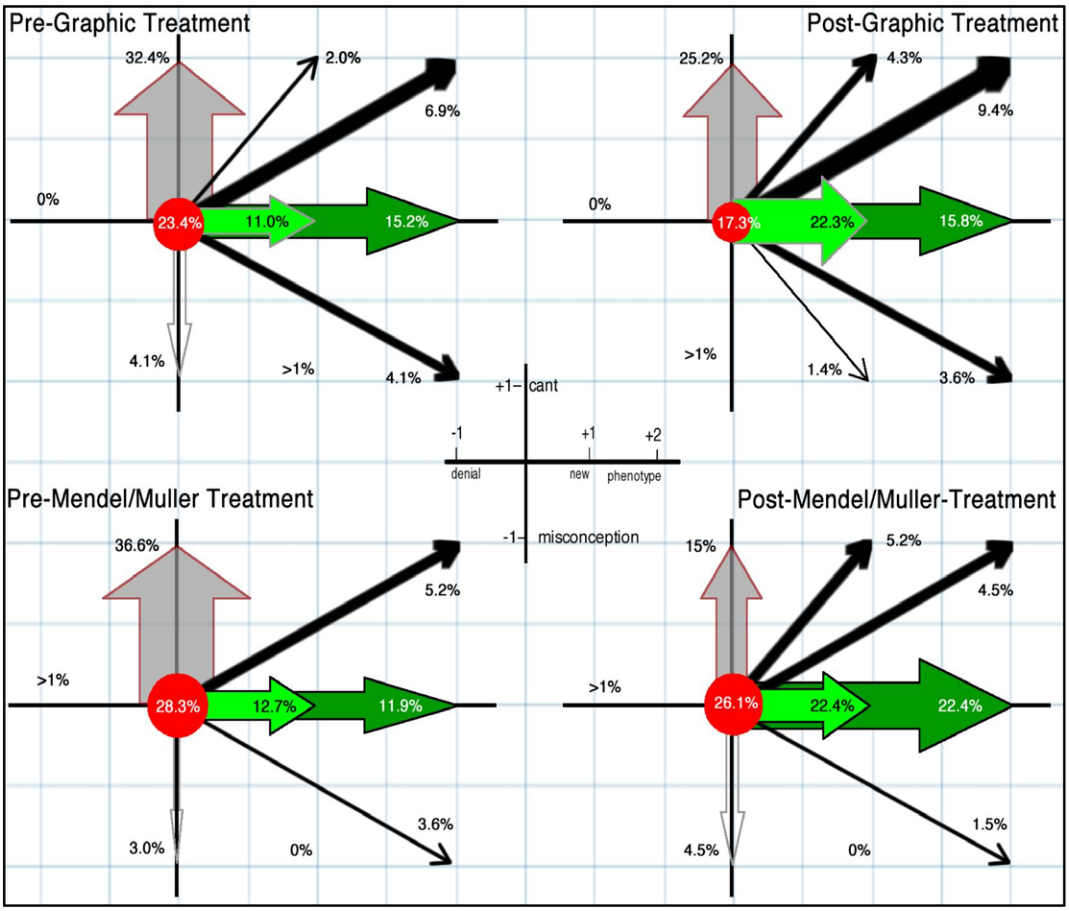

Fig. 3. Changes in student thinking. Students were asked to work through either the Graphical Thinking (top panels) or the Mendel/Muller (bottom panels) activities in groups. Student responses to the "How might a mutation be creative?" question pre- (left panels) and post- (right panels) treatment were analyzed.

(unpublished observations), and a range of other studies (Smith and Tanner, 2010), namely that students can pass typical tests without a rigorous and transferable understanding of the underlying and important ideas. We attempted to avoid this in both the Mendel/Muller and Graphical Thinking tutorials. For example, Mendel/Muller does not talk explicitly about the nature of gene products (that is, RNAs and polypeptides) nor does it mention exactly what a mutation can do, in molecular terms, to such gene products. It simply introduces the various types of mutational effects in the context in which they were conceived of by one of the pioneers of genetics. The tutorial does not talk about mutations as being creative, although the discussion of neomorphic mutations does, per force, address questions of new and novel functions. We were, in fact, amazed that four weeks after the activity, we detected a significant signal, namely a decrease in cant and an increase in recognition that mutations can give rise to new phenotypes. The results, while admittedly restricted in scope, support the hypothesis that student understanding increases significantly if mutations/alleles are introduced in the same way as they were considered by the pioneers of genetics, namely as abstract entities with various "functions" (Sturtevant, 1965).

Our choice of Muller's 1932 characterization of phenotypes and their association with specific mutations/alleles was based on its abstract character. It can be applied to essentially any system and the terms used (morphs) can be readily translated into terms of gene product function: for example, amorph equals no function while neomorph equals new function (see above). At the same time, mutations without overt effects on phenotype (and which can be "captured" through genetic drift) generate a reservoir of genetic variation that can be used later through exadaption (Blount et al., 2008; Wagner, 2005).

The advantages of this framework are two fold. First, it applies to all mutational events, from point mutations to gene duplications and deletions. It therefore provides a universal language which can be used to talk about the functions of direct (DNA and RNA) and indirect (polypeptide and protein) gene products. Through conversations in this language students can begin to consider the fact that many gene products have multiple functions, that gene duplication is essential for functional specialization, and that genetic drift is a key evolutionary mechanism (Aharoni et al., 2005; Copley, 2003; Lynch, 2007). It addition, it enables students to connect what may seem, superficially, to be unrelated functions. Consider the role of molecular chaperones, proteins that act to facilitate the folding of polypeptides and the assembly of proteins and macromolecular complexes. In the context of Muller's morphs, one can appreciate their role in evolutionary processes, where they have been proposed to act as a "buffer". Effects of a mutation that would destabilize a gene product, and lead to a hypomorphic or amorphic effect can be ameliorated by a chaperone, converting for example, the amorphic to the hypomorphic or the

hypomorphic to "normomorphic." This type of effect has been proposed to make a wide range of evolutionary change possible (Lindquist, 2009; Tokuriki and Tawfik, 2009). Only after providing our students with a learnable language by which to talk generally and coherently about mutational effects and evolutionary mechanisms, such as gene duplication and chaperone function, can we reasonably expect them to have cogent things to say.

Acknowledgements

We thank Bob Boswell, Mark Winey, Clayton Lewis, Ross Nehm, and Erin Furtak for comments on previous versions of the manuscript, Erik Hedl for help with database creation and searching, Will Leary for his work interviewing students; Erin Furtak and the students of Teaching and Learning Biology, Fall 2008 and the participants in NSTA's various listservs for sharing their questions, comments, and observations. This project was supported in part by NSF grants DUE 0405007, DUE 0816692, and DUE-1043707, NSF-Noyce funding for Will Leary, and an ASSETT grant from the University of Colorado, Boulder.

Competing Interests

The authors declare that there are no competing interests.

Note added in proof

The importance of stochastic events and molecular chaperones in the relationship between alleles and phenotypes has been called to attention in two recent high profile papers (Burga et al., 2011; Casanueva et al., 2012).

References

Aharoni, A., Gaidukov, L., Khersonsky, O., McQ Gould, S., Roodveldt, C. and Tawfik, D. S. (2005). The 'evolvability' of promiscuous protein functions. *Nat. Genet.* **37**, 73-76.

Anderson, D. L., Fisher, K. M. and Norman, G. J. (2002). Development and evaluation of the conceptual inventory natural selection. *J. Res. Sci. Teach.* **39**, 952-978.

Bergthorsson, U., Andersson, D. I. and Roth, J. R. (2007). Ohno's dilemma: evolution of new genes under continuous selection. *Proc. Natl. Acad. Sci. USA* **104**, 17004-17009.

Berkman, M. B. and Plutzer, E. (2011). Science education. Defeating creationism in the courtroom, but not in the classroom. *Science* **331**, 404-405.

Blount, Z. D., Borland, C. Z. and Lenski, R. E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proc. Natl. Acad. Sci. USA* **105**, 7899-7906.

Bowler, P. J. (1992). *The Eclipse Of Darwinism: Anti-Darwinian Evolution Theories In The Decades Around 1900.* Baltimore: Johns Hopkins University Press.

Bowler, P. J. (2005). Revisiting the Eclipse of Darwinism. *J. Hist. Biol.* **38**, 19-32.

Bowling, B. V., Acra, E. E., Wang, L., Myers, M. F., Dean, G. E., Markle, G. C., Moskalik, C. L. and Huether, C. A. (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genetics* **178**, 15-22.

Burga, A., Casanueva, M. O. and Lehner, B. (2011). Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature* **480**, 250-253.

Casanueva, M. O., Burga, A. and Lehner, B. (2012). Fitness trade-offs and environmentally induced mutation buffering in isogenic C. elegans. *Science* **335**, 82-85.

Copley, S. D. (2003). Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr. Opin. Chem. Biol.* **7**, 265-272.

Donovan, M. S. and Bransford, J. D. (2005). *How Students Learn: History, Mathematics, And Science In The Classroom*. Washington, DC: National Academies Press.

Garvin-Doxas, K. and Klymkowsky, M. W. (2008). Understanding randomness and its impact on student learning: lessons learned from building the Biology Concept Inventory (BCI). *CBE Life Sci. Educ.* **7**, 227-233.

Gregory, T. R. (2009). Understanding natural selection: essential concepts and common misconceptions. *Evo. Edu. Outreach* **2**, 156-175.

Havighurst, R. J. (1929). Reform in the chemistry curriculum. *J. Chem. Educ.* **6**, 1126-1129.

Heller, P. and Huffman, D. (1995). Interpreting the force concept inventory: A reply to Hestenes and Halloun. *Phys. Teach.* **33**, 503, 507-511.

Hestenes, D. and Halloun, I. (1995). Interpreting the force concept inventory: A response to March 1995 critique by Huffman and Heller. *Phys. Teach.* **33**, 502-506.

Hestenes, D., Wells, M. and Swackhamer, G. (1992). Force concept inventory. *Phys. Teach.* **30**, 141-166.

Jablonka, E. and Lamb, M. J. (2005). *Evolution In Four Dimensions: Genetic, Epigenetic, Behavioral, And Symbolic Variation In The History Of Life.* Cambridge, MA: MIT Press.

Klymkowsky, M. W. (2011). Why is evolution so hard to understand? *ASBMB Today* **March**, 14-15.

Klymkowsky, M. W. and Garvin-Doxas, K. (2008). Recognizing student misconceptions through Ed's Tools and the Biology Concept Inventory. *PLoS Biol.* **6**, e3.

Klymkowsky, M. W., Underwood, S. M. and Garvin-Doxas, K. (2010). Biological Concepts Instrument (BCI): A diagnostic tool for revealing student thinking. *arXiv*:1012.4501.

Lindquist, S. (2009). Protein folding sculpting evolutionary change. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 103-108.

Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. USA* **104** Suppl 1:, 8597-8604.

Mayr, E. (1985). *The Growth Of Biological Thought: Diversity, Evolution, And Inheritance.* Cambridge, MA: Belknap Press.

Muller, H. J. (1932). Further studies on the nature and causes of gene mutations. *Sixth Int. Cong. Genet.* **1**, 213-255.

Muller, H. J. (1936). Bar Duplication. *Science* **83**, 528-530.

Nersessian, N. J. (1989). Conceptual change in science and in science education. *Synthese* **80**, 163-183.

Powell, K. (2003). Science education: spare me the lecture. *Nature* **425**, 234-236.

Smith, J. I. and Tanner, K. (2010). The problem of revealing how students think: Concept inventories and beyond. *CBE Life Sci. Educ.* **9**, 1-5.

Smith, M. K., Wood, W. B. and Knight, J. K. (2008). The Genetics Concept Assessment: a new concept inventory for gauging student understanding of genetics. *CBE Life Sci. Educ.* **7**, 422-430.

Sturtevant, A. H. (1965). *A History Of Genetics*. New York: Cold Spring Harbor Laboratory Press.

Taber, K. S. (2001). Constructing chemical concepts in the classroom?: Using research to inform practice. *Chem. Educ. Res. Pract. Eur.* **2**, 43-51.

Tokuriki, N. and Tawfik, D. S. (2009). Protein dynamism and evolvability. *Science* **324**, 203-207.

Wagner, A. (2005). Robustness, evolvability, and neutrality. *FEBS Lett.* **579**, 1772-1778.