

## Article

# The Dominance Concept Inventory: A Tool for Assessing Undergraduate Student Alternative Conceptions about Dominance in Mendelian and Population Genetics

Joel K. Abraham,\* Kathryn E. Perez,<sup>†</sup> and Rebecca M. Price<sup>‡</sup>

\*Department of Biological Science, California State University–Fullerton, Fullerton, CA 92831; <sup>†</sup>Department of Biology, University of Wisconsin–La Crosse, La Crosse, WI 54601; <sup>‡</sup>School of Interdisciplinary Arts and Sciences, University of Washington–Bothell, Bothell, WA 98011

Submitted August 16, 2013; Revised January 3, 2014; Accepted January 3, 2014  
Monitoring Editor: William B. Wood

Despite the impact of genetics on daily life, biology undergraduates understand some key genetics concepts poorly. One concept requiring attention is *dominance*, which many students understand as a fixed property of an allele or trait and regularly conflate with frequency in a population or selective advantage. We present the Dominance Concept Inventory (DCI), an instrument to gather data on selected alternative conceptions about dominance. During development of the 16-item test, we used expert surveys ( $n = 12$ ), student interviews ( $n = 42$ ), and field tests ( $n = 1763$ ) from introductory and advanced biology undergraduates at public and private, majority- and minority-serving, 2- and 4-yr institutions in the United States. In the final field test across all subject populations ( $n = 709$ ), item difficulty ranged from 0.08 to 0.84 ( $0.51 \pm 0.049$  SEM), while item discrimination ranged from 0.11 to 0.82 ( $0.50 \pm 0.048$  SEM). Internal reliability (Cronbach's alpha) was 0.77, while test–retest reliability values were 0.74 (product moment correlation) and 0.77 (intraclass correlation). The prevalence of alternative conceptions in the field tests shows that introductory and advanced students retain confusion about dominance after instruction. All measures support the DCI as a useful instrument for measuring undergraduate biology student understanding and alternative conceptions about dominance.

## INTRODUCTION

College students struggle to understand genetics, even though genetics concepts are regularly covered in high school and college biology courses (Finley *et al.*, 1982; Mills Shaw *et al.*, 2008; McElhinny *et al.*, 2012) and textbooks (e.g., Campbell *et al.*, 2008; Miller and Levine, 2010; Freeman *et al.*, 2013; Phelan, 2013). The phenomenon of dominance is particularly

challenging for students to understand (Collins and Stewart, 1989; Heim, 1991; Allchin, 2000; Christensen, 2000). Dominance is often ascribed to either an allele or a trait (Allchin, 2000, 2005), but it actually describes the pattern when a phenotype that is associated with only one allele of an allelic pair is expressed in the phenotype of both the homozygote and heterozygote. In reality, this pattern shifts, depending on the makeup of the allelic pair (well demonstrated by the human ABO blood system) and thus is not technically a property of alleles themselves. The shorthand convention, however, which we follow in this paper, uses dominant and recessive as adjectives for alleles or traits. Instructors may not discuss the biochemical processes underlying phenotypic expression when teaching dominance, and thus students may fail to recognize that typically both alleles produce gene products (Allchin, 2000; McElhinny *et al.*, 2012). Additionally, students often mistake trait or allelic frequency in a population, or impact on survival or reproduction, for dominance (e.g., Collins and Stewart, 1989; Heim, 1991; Donovan, 1997; Allchin, 2000). Furthermore, they often inaccurately reason that dominant

DOI: 10.1187/cbe.13-08-0160

Address correspondence to: Joel K. Abraham (jkabraham@fullerton.edu).

© 2014 J. K. Abraham *et al.* CBE—Life Sciences Education © 2014 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

**Table 1.** Field-testing institutions in pilot and main studies<sup>a</sup>

Stage	Institution	Course	Level	Type	Description	<i>n</i>
Pilot	A	1	Adv	4Y, public	Masters, small, Midwest	46
	B	1	Intro	4Y, public	Doctoral, large, Southeast	161
	C	1	Intro	4Y, private	Doctoral, large, Northeast	227
	D	1	Intro	4Y, private	Undergraduate, small, historically black college or university, Southeast	107
	E	1	Adv	4Y, public	Undergraduate, small, liberal arts, Northwest	22
Pilot total						563
Main	F	1	Intro	4Y, public	Masters, large, Hispanic serving, West	78
			2*	Adv		21
	G	1	Intro	4Y, public	Doctoral, large, Midwest	174
	H	1	Intro	4Y, public	Masters, small, Midwest	150
			2	Adv		89
	I	1*	Intro	4Y, public	Undergraduate, small, liberal arts, Northwest	29
		2*	Adv			40
	J	1*	Intro	4Y, public	Doctoral, large, Northwest	318
		2*	Intro			189
		3*	Adv			53
	K	1*	Intro	2Y, public	Associate, large, minority serving, West	9
Main total	L	1*	Intro	2Y, public	Associate, large, West	35
		2*	Intro			15
						1200

<sup>a</sup>2Y = 2-yr institution; 4Y = 4-yr institution; Intro = introductory-level students; Adv = advanced-level students. \* indicates courses used in final testing of DCI.

alleles or traits must increase in frequency in a population over time (e.g., Allchin, 2000; Christensen, 2000).

Alternative conceptions about phenotypic expression of allelic pairs may impact student understanding of other concepts (Allchin, 2002). For example, students may have difficulty understanding how deleterious dominant alleles, such as those responsible for Huntington's chorea, could be retained in a population (Cortopassi, 2002), or mistakenly assume that traits coded for by recessive alleles in human populations, such as blond hair, are "dying out" (BBC News, 2002). With the ongoing permeation of genetics concepts and practices into everyday life (e.g., at-home genetic testing kits), a misunderstanding or misapplication of fundamental ideas in genetics can be detrimental beyond the college classroom (McElhinny *et al.*, 2012).

To date, the lack of diagnostic tools has made it difficult to gauge the prevalence of these alternative conceptions. While there are several concept inventories (CIs) that assess student understanding of genetics, such as the Genetics Concept Assessment (GCA; Smith *et al.*, 2008), the Genetics Literacy Assessment Instrument (Bowling *et al.*, 2008), and the Genetics Concept Inventory (Elrod, 2007), they are broad in scope. To focus specifically on dominance, we have developed the Dominance Concept Inventory (DCI) for students of college biology. The DCI focuses on how students understand the term *dominance* as it relates to phenotypic expression of allelic pairs, but does not explore students' understanding of the molecular basis for the multiple mechanisms that lead to the pattern of dominance. The narrow focus of the DCI reflects the most common alternative conceptions encountered by us in our own teaching and discussed in the literature. We carried out testing and validation with diverse student populations, including community colleges and minority- and majority-serving universities. The DCI contains 16 items, including true-false, multiple-choice, and two-tiered multiple-choice questions.

## METHODS

We conducted this study in two stages: the pilot study (Fall 2009–Summer 2010) and the main study (Winter 2011–Fall 2013). We were granted approval for this research by IRBs at MIT (expedited: 0705002250), the University of Wisconsin–La Crosse (exempt: no number assigned), the University of Washington (expedited: 42505), California State University–Fullerton (expedited: HSR-12-0432), and all other participating institutions during testing (additional approval numbers are not reported to preserve the instructors' and subjects' anonymity). Details about the participating institutions are listed in Table 1. Participating institutions included community colleges and minority- and majority-serving universities. When we conducted interviews, we made every effort to balance the participant pool for gender, to include racial and ethnic diversity, and to include ESL learners.

### Pilot Study

In the first stage of this study, we used an iterative process to identify and verify student alternative conceptions in biology topics; this involved rounds of literature searches, interviews with college students, and discussion with colleagues (as in Smith *et al.*, 2008; Abraham *et al.*, 2012; Hiatt *et al.*, 2013; Table 2). Throughout this paper, we use *alternative conceptions* in place of common phrases such as *misconceptions* or *naïve conceptions*. While it may be that *misconceptions* is the more popular term (Crowther and Price, 2014), some have suggested that *alternative conceptions* better respects research demonstrating that meaningful learning may start from scientifically inaccurate conceptions (Maskiewicz and Lineback, 2013).

We developed an initial set of open-response and multiple-choice questions about Mendelian and population genetics to use in interviews. We then recruited subjects ( $n = 26$ ) from

**Table 2.** Overview of the DCI development process<sup>a</sup>

Pilot study	1	Identified and verified student alternative conceptions in population and Mendelian genetics through literature searches, interviews ( $n = 26$ ), and discussions with instructors.
	3	Drafted multiple-choice pilot test on population and Mendelian genetics.
	4	Experts ( $n = 5$ ) reviewed test; pilot test revised based on feedback.
	5	Administered 31-item test to students ( $n = 563$ ) across five campuses.
Main study	6	EvoCI group revised or created new test items and developed target concepts relating to dominance.
	7	Draft DCI sent to experts ( $n = 7$ ) for review; DCI revised based on feedback.
	8	Administered revised DCI to students and conducted follow-up interviews ( $n = 16$ ); used feedback to make additional changes to wording and formatting.
	9	Administered DCI to students ( $n = 491$ ) across three campuses (institutions F, G, and H). Revised wording of two DCI items.
	10	Administered final version of DCI as paper- ( $n = 133$ ) and computer-based test ( $n = 576$ ) to students across five campuses (institutions F, I, J, K, and L). Calculated internal reliability, difficulty, discrimination for each test administration and subject population. Readministered test to one course at institution J ( $n = 53$ ) for test-retest reliability.

<sup>a</sup>Institution codes from Table 1.

Boston-area private and public 2- and 4-yr colleges and universities (Table 2). We held interviews with subjects, alone or in pairs, and asked them to answer our oral and written questions, explain their answers, and define terminology. After each interview, we modified questions or added additional multiple-choice questions. Interviews lasted no longer than 2 h, and no subject took part in more than one interview. Subjects were paid for their participation. An example interview protocol is included in the Supplemental Material.

We identified a number of alternative conceptions relating to population genetics, Mendelian genetics, and specifically phenotypic expression of allelic pairs, in the literature and interviews. We ultimately focused on four alternative conceptions (Table 3), with the recognition that additional alternative conceptions around dominance undoubtedly exist and affect student learning.

Two conceptions are related, in that students assume a relationship between the dominance and the frequency of alleles or traits in the population. *DomFreq* is the idea that the *current* frequency of an allele in a population is linked to dominance (Collins and Stewart, 1989; Heim, 1991; Donovan, 1997; Allchin, 2000); students conflate dominance with pre-dominant. *DomInc* is the idea that *increases* in the frequency of an allele over time are linked to dominance (Allchin, 2000;

Christensen, 2000). Both of these alternative conceptions appear in the literature (Table 3) and were among the most commonly encountered during interviews. For instance, when asked in interview questions about a two-allele system in which an allele ( $g_1$ ) is at a higher frequency in the population, Student A responded,

Student A:  $g_1$  is dominant because it has a greater frequency.

That same student, when asked about a different two-allele system, in which black hair is at a higher frequency than white hair, responded,

Student A: Black hair because it has a greater frequency.

Other students offered additional details in their responses:

Student B: From what I saw, black is more dominant ... if the trait is dominant, it will be expressed more frequently. I guess it depends on your sample. If a large sample size, and black hair is more frequent than white hair, can assume black is probably dominant.

Student C: If the allele is dominant over other alleles, it will obviously rise in frequency.

**Table 3.** Descriptions of the concepts and alternative conceptions covered in the DCI<sup>a</sup>

	Description	DCI item
Target concept		
TC1	Evolutionary processes determine the frequency of an allele in a population.	1, 3, 4, 7, 10, 11, 12, 13, 16
TC2	The selective advantage of a phenotype in a population is determined by its impact on survival and reproduction.	2, 5, 8, 9, 13, 14, 15
Alternative conception		
DomFreq	The frequency of an allele in a population is related to dominance (e.g., Collins and Stewart, 1989; Heim, 1991; Donovan, 1997; Allchin, 2000).	1, 10
DomInc	Dominant alleles increase in frequency in a population (e.g., Allchin, 2000; Christensen, 2000).	3, 4, 7, 11, 12, 13, 16
DomSelect	Dominance is related to the selective advantage/disadvantage of an allele or allelic pair (e.g., Heim, 1991; Allchin, 2000).	2, 5, 13, 14
HeteroSelect	Heterozygotes have a selective advantage over other genotypes.	8, 9, 15

<sup>a</sup>The DCI and the mapping of DCI items to DCI question numbers and answers can be found in the Supplemental Material.

Student D: I would think that the dominant one would grow to be more dominant, and the recessive would get less and less. After longer period of time, would expect dominant to win out.

More than half of the interviews included frequency-related alternative conceptions about dominance.

Two other alternative conceptions occur when students conflate selective advantage with dominance or genotype. *DomSelect* is the alternative conception that dominant alleles code for traits that are selectively advantageous (Heim, 1991; Allchin, 2000). When asked about dominance and the selective advantage of traits, some students replied,

Student C: Since g1 is dominant, and if it has been for some time, I would say it is selectively advantageous.

Student D: If an allele is dominant, and stays dominant, it is selectively advantageous.

Student E: Usually populations like to have the dominant take over because it is more helpful or beneficial to the population.

*HeteroSelect* is the idea that individuals who are heterozygous at a particular locus have a selective advantage over homozygous individuals, often referred to as *overdominance*. While single-locus overdominance is a real phenomenon, such as the commonly presented sickle-cell anemia example, it does not apply in every case. Previous instruction on hybrid vigor, the pattern of higher fitness in hybrid offspring relative to their parent strains, may also influence student perspectives on this alternative conception. To our knowledge, *HeteroSelect* has not been documented in the literature, yet several of us have encountered it in our teaching, as well as in interviews:

Student F: In general, heterozygous genotypes are more advantageous to a population.

Student G: Yeah, because heterozygous is basically when two different types of traits are being matched. Homozygous . . . [there is] no room for improvement in terms of adaptation to the environment. I think for heterozygous it will be more advantageous when the offspring are a combination of traits that can survive in the environment.

Another student, when asked about a two-allele system showing incomplete dominance at a locus for horn length, responded,

Student H: The medium-horned lizard is probably the most selectively advantageous, because it is heterozygous, which usually indicates a higher fitness.

We used student responses to modify our multiple-choice questions and match distracters to alternative conceptions. Once we completed a draft CI, we asked five college-level population genetics instructors to review the items and make suggestions for improvements. We corrected scientific inaccuracies and improved the clarity of the questions. The resulting 31-item pilot test focused on student conceptions and knowledge of Mendelian and population genetics.

We administered the pilot paper-based test to students in introductory or advanced courses ( $n = 563$ ) at five different institutions, which were a mix of public or private and

majority or minority serving (Table 1). Testing took place after students received instruction in population or Mendelian genetics. Students who left multiple questions blank were removed from the study.

After this point, we decided to focus on dominance relationships. A nine-item portion of the original test focused on dominance and phenotypic expression of alleles and served as the foundation for the DCI. One additional item from the original test was also useful for developing the DCI, so we report student responses to this item in the *Results* as well:

Which of the following descriptions best fits the relationship between the terms **gene**, **allele**, **locus**, and **chromosome**?

- A. Alleles are different forms of a gene, which can be found at a locus on a chromosome.
- B. Genes are made up of chromosomes, which can be found at a locus on an allele.
- C. Chromosomes are different forms of a gene, which can be found at an allele on a locus.
- D. A locus is made up of alleles, which can be found at a gene on a chromosome.

We generated the options in this item from student responses to an open-response question about the terms.

### Main Study

We then began revising items and identifying target concepts for the DCI in consultation with members of the EvoCI working group (a working group funded by the National Evolutionary Synthesis Center to develop CIs about evolution). From these discussions, we created a preliminary list of target concepts and alternative conceptions and revised items for the DCI. After each revision, other EvoCI working group members reviewed the items and offered suggestions on wording. This constituted an initial informal expert review ( $n = 5$ ). The final list of target concepts and alternative conceptions is in Table 3.

We sent a revised draft of the DCI to seven more experts for content review. Experts were identified through convenience sampling (Marshall, 1996); six experts are classical genetics, population genetics, or evolution instructors at 4-yr institutions, while the seventh expert is a population genetics researcher. We first asked experts to review and comment on the target concepts and alternative conceptions in the DCI. We then asked them to read the DCI and respond to the following questions:

1. This item addresses the target concept. (1–strongly agree to 5–strongly disagree)
2. This item is scientifically accurate. (1–strongly agree to 5–strongly disagree)
3. Please make any suggestions for improving the clarity and accuracy of this item. If the answer(s) that we have chosen are incorrect, please explain why. (Open response)

We addressed items for which any expert selected “disagree” or “strongly disagree” for target concept or scientific accuracy, and only retained items for which the average score for either measure was greater than two. Average scores for target concept and scientific accuracy were 1.64/5 and 1.46/5,

respectively. We removed two draft items and used the feedback to modify the language, structure, and content of seven items.

We then administered paper- and computer-based versions of the DCI to, and conducted follow-up interviews with, 16 undergraduates from institution F. Students ranged from their first to fifth year of college and were from introductory biology to upper-division evolution courses. Subjects were given gift cards for their participation. We interviewed subjects individually, and they described their thought processes for their answers to each item. We subsequently modified the items to increase clarity.

We then implemented large-scale testing in courses at institutions F (paper based) and G and H (computer based); see Table 1 ( $n = 489$ ). Modifications to the DCI after this stage were minor and addressed items that had confusing wording, typos, or very low or very high difficulty and discrimination scores (see *Data Analysis*). The resulting final version of the DCI and the answer key can be found in the Supplemental Material.

Final testing occurred at institution F, one course each at institutions J and K, and two courses at institution L (paper based,  $n = 133$ ), as well as one course at institution I and two courses at institution J (computer based,  $n = 576$ ); see Table 1. Students were given credit for completing the DCI but not for their performance on the test. The analyses in the *Main Study* section of the *Results* are based on only this final round of testing ( $n = 709$ ).

We readministered the DCI in one advanced course at institution J 2 wk after subjects completed the first test (paper-based,  $n = 53$ ). No deliberate instruction on DCI content took place during that time. We used those results to calculate test-retest reliability (Crocker and Algina, 1986), but excluded the second DCI administration from the rest of the study.

While we did not systematically collect completion time data, all students at institution F completed the paper-based DCI within 15 min, and at institution J, the average completion time for the computer-based DCI was  $\sim 12$  min.

### Data Analysis

We scored the DCI by dividing the number of correct responses by the total possible number of correct responses. Similarly, we calculated the prevalence of alternative conceptions by dividing the number of responses indicating a given alternative conception by the total number of items that target that alternative conception. Because the pilot and main study test items differed, we calculated performance and prevalence of alternative conceptions separately for each test.

We estimated both internal reliability and test-retest reliability, a measure of stability, for the DCI using R (R Development Core Team, 2012). To estimate internal reliability, we used Cronbach's alpha (Cronbach, 1951). We estimated test-retest reliability using two different calculations: Pearson's product moment correlation (PPC; also described as Pearson's  $r$ ) and intraclass correlation (ICC). Researchers favor different approaches (e.g., Rousson *et al.*, 2002; Weir, 2005), so we chose to run both calculations. For ICC, we used Model 2,1, which treats test implementations as equivalent and uses individual student responses, as opposed to averaged responses, in its calculation (Weir, 2005).

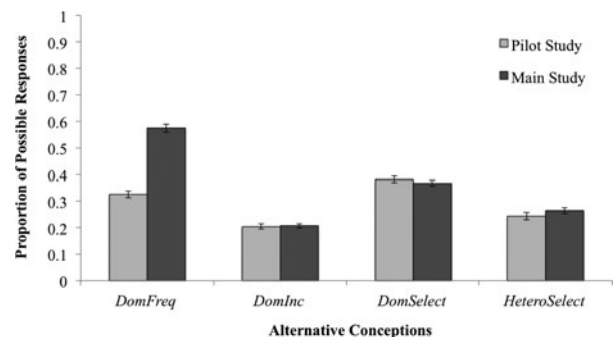
We calculated item difficulty ( $P$ ) and item discrimination ( $D$ ) indices for each version of the DCI in the main study, but include only the values for the final round of testing. Item difficulty ( $P$ ) is calculated by dividing the number of correct responses by the total number of responses for each item; the lower the value, the more difficult the item is (Crocker and Algina, 1986). Item discrimination ( $D$ ) measures how well an item distinguishes between students who perform well on the test and those who perform poorly; lower values of  $D$  indicate poorer discrimination (Crocker and Algina, 1986). We defined high-performing students as those scoring in the top third of students, while low-performing students were those who scored in the bottom third of students; and we calculated  $D$  by subtracting the average item difficulty for low performers from that of high performers for each item (Crocker and Algina, 1986).

We calculated combined and subgroup internal reliability, item statistics, and performance. We had defined two types of subgroups: administration method and subject population (Table 4). The administration method was computer- or paper-based administration of the DCI. The subject populations were introductory students at 2-yr institutions (2Y-Intro), students at 4-yr institutions (4Y-Intro), and advanced students at 4-yr institutions (4Y-Adv). We compared subgroup difficulty, discrimination, and performance with a series of independent two-tailed  $t$  tests (administration method) or single-factor analysis of variance (subject population).

## RESULTS

### Subject Performance and Prevalence of Alternative Conceptions

Subjects in the pilot study scored an average of 0.52 ( $\pm 0.012$  SEM) on the pilot DCI. Subjects in the main study scored an average of 0.52 ( $\pm 0.010$  SEM) on the final DCI, and performance did not vary significantly across subgroups (administration method:  $t = -0.24$ ,  $p = 0.811$ ; student population:  $F = 1.15$ ,  $df = 2$ ,  $p = 0.317$ ; Table 4). DomSelect and DomFreq were the most prevalent alternative conceptions in the pilot study, appearing in 0.38 ( $\pm 0.012$  SEM) and 0.33 ( $\pm 0.013$  SEM) of the responses, respectively (Figure 1).



**Figure 1.** Frequency of target alternative conceptions found in student responses in the pilot ( $n = 563$ ) and main study (final version of the DCI,  $n = 709$ ). We calculated alternative conception frequencies by dividing the number of responses indicating a given alternative conception by the total number of items that target that alternative conception in the pilot or main study test. Error bars are 1 SEM.

**Table 4.** Summary of mean subject performance (proportion correct), mean alternative conception prevalence (proportion of possible responses), mean item statistical values, and reliability (Cronbach's alpha) of DCI for full data set (combined), as well as each testing subgroup (administration method, subject populations) from the main study<sup>a</sup>

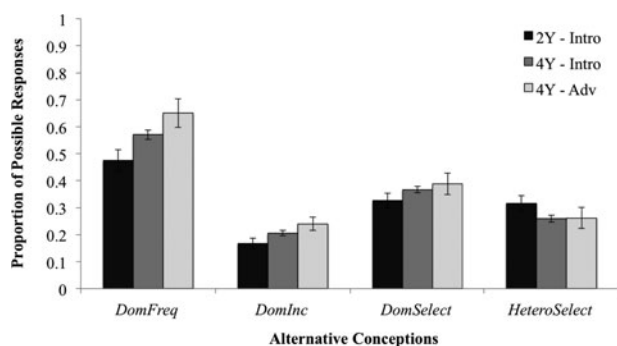
Testing group	<i>n</i>	Performance	DomFreq	DomInc	DomSelect	HeteroSelect	Difficulty ( <i>P</i> )	Discrimination ( <i>D</i> )	Reliability
Administration method									
Paper	133	0.51 (0.02)	0.57 (0.02)	0.20 (0.02)	0.35 (0.03)	0.29 (0.03)	0.51 (0.05)	0.48 (0.05)	0.79
Computer	576	0.50 (0.01)	0.57 (0.04)	0.21 (0.01)	0.37 (0.01)	0.26 (0.01)	0.51 (0.05)	0.49 (0.05)	0.76
Subject population									
2Y-Intro	59	0.49 (0.03)	0.47 (0.05)	0.17 (0.02)	0.33 (0.04)	0.32 (0.04)	0.49 (0.05)	0.49 (0.06)	0.77
4Y-Intro	547	0.51 (0.01)	0.57 (0.02)	0.21 (0.01)	0.37 (0.01)	0.26 (0.01)	0.52 (0.05)	0.50 (0.05)	0.76
4Y-Adv	103	0.48 (0.02)	0.65 (0.04)	0.24 (0.02)	0.39 (0.03)	0.26 (0.03)	0.48 (0.05)	0.51 (0.05)	0.80
Combined	709	0.51 (0.01)	0.57 (0.02)	0.21 (0.01)	0.37 (0.01)	0.26 (0.01)	0.51 (0.05)	0.50 (0.05)	0.77

<sup>a</sup>When mean values are given,  $\pm 1$  SEM is included in parentheses. There were no significant differences in performance, difficulty, or discrimination among any of the subgroups. 2Y = 2-yr institution; 4Y = 4-yr institution; Intro = introductory-level students; Adv = advanced-level students.

DomInc and HeteroSelect averaged 0.2 ( $\pm 0.01$  SEM) and 0.24 ( $\pm 0.014$  SEM), respectively (Figure 1).

As in the pilot study, DomFreq and DomSelect were the most frequent alternative conceptions in the main study (Figure 1), with 0.57 ( $\pm 0.016$  SEM) of possible subject responses indicating the DomFreq alternative conception, while 0.37 ( $\pm 0.010$  SEM) of responses indicated the DomSelect alternative conception. Both DomInc and HeteroSelect appeared less frequently in student responses, averaging 0.21 ( $\pm 0.01$  SEM) and 0.26 ( $\pm 0.010$  SEM), respectively (Figure 1). Administration method did not affect the rank order or prevalence of alternative conceptions (Table 4). The prevalence of alternative conceptions did vary across subject populations, and the rank order was similar across subgroups (Figure 2). No subject population consistently outperformed the others (Figure 2; Table 4).

On the pilot study item about the relationship between gene, allele, chromosome, and locus, subjects selected the correct option (A) 65% of the time. However, performance on this item varied quite a bit among institutions. In one advanced course, ~80% of the students chose the correct option.



**Figure 2.** Frequency of target alternative conceptions found in each subject population of the main study: introductory students at 2-yr institutions (2Y-Intro,  $n = 59$ ), introductory students at 4-yr institutions (4Y-Intro,  $n = 547$ ), and advanced students at 4-yr institutions (4Y-Adv,  $n = 103$ ). We calculated alternative conception frequencies by dividing the number of responses indicating a given alternative conception by the total number of items that target that alternative conception. Error bars are 1 SEM.

In the lowest-scoring course, another advanced course, only 50% of the students answered this question correctly.

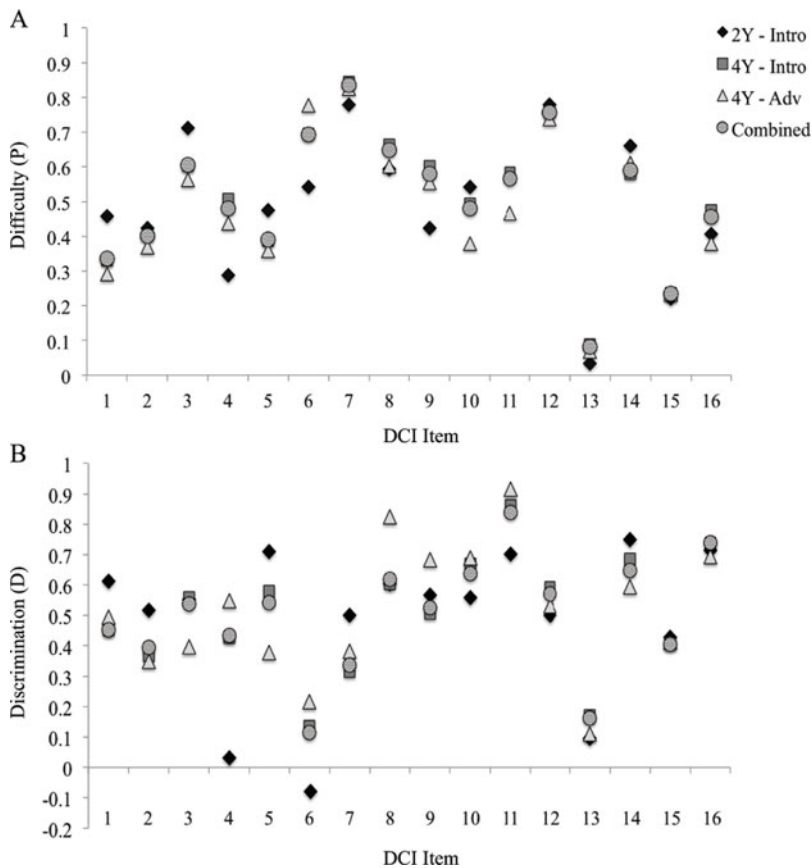
### Test Characteristics

Mean difficulty (*P*) and discrimination (*D*) were statistically similar for paper-based versus computer-based implementations (difficulty:  $t = -0.031$ ,  $p = 0.98$ ; discrimination:  $t = 0.13$ ,  $p = 0.898$ ) as well as for introductory 2-yr, 4-yr, and advanced 4-yr students (difficulty:  $F = 0.142$ ,  $df = 2$ ,  $p = 0.868$ ; discrimination:  $F = 0.068$ ,  $df = 2$ ,  $p = 0.935$ ; Table 4). The combined item difficulty and discrimination across all subject populations were generally within the range of desirability (Crocker and Algina, 1986; Haladyna, 2004; Figure 3): difficulty (*P*) ranged from 0.08 to 0.84 (lower values are more difficult items), with a mean *P* of 0.51 ( $\pm 0.049$  SEM); discrimination (*D*) ranged from 0.11 to 0.82, with a mean *D* of 0.50 ( $\pm 0.048$  SEM). However, *P* and *D* varied for some items, particularly items 4 and 6, across different subject populations (Figure 3).

Internal reliability (Cronbach's alpha) of the final version of the 16-item DCI was high for both the paper-based ( $n = 133$ ,  $\alpha = 0.79$ ) and computer-based ( $n = 576$ ,  $\alpha = 0.76$ ) implementations; the combined internal reliability was 0.77. Internal reliability also did not differ appreciably between different subject testing subgroups (Table 4). Moreover, these values are comparable with those found in other published CIs, such as the Concept Inventory of Natural Selection ( $\alpha = 0.58$ – $0.64$  [Anderson *et al.*, 2002]) or the EvoDevoCI ( $\alpha = 0.31$ – $0.73$  [Perez *et al.*, 2013]). Smith *et al.* (2008) estimated test-retest reliability for the GCA using data from testing in two semesters of a course (coefficient of stability = 0.93). We took the more conventional approach of comparing two successive test administrations in the same population to estimate test-retest reliability (Crocker and Algina, 1986), using two methods of calculation: Pearson's PCC and the ICC (Model 2,1). PCC and ICC test-retest values were 0.74 and 0.77, respectively.

### DISCUSSION

After several rounds of development, the DCI satisfies a number of criteria for a useful CI to assess student understanding



**Figure 3.** DCI test item statistics for each main study subject subgroup: introductory students at 2-yr institutions (2Y-Intro,  $n = 59$ ), introductory students at 4-yr institutions (4Y-Intro,  $n = 547$ ), and advanced students at 4-yr institutions (4Y-Adv,  $n = 103$ ). (A) Difficulty ( $P$ ); (B) discrimination ( $D$ ). Higher values of  $P$  and  $D$  indicate easier items and better discrimination between high- and low-performing students, respectively.

and the prevalence of alternative conceptions held by undergraduates concerning dominance and phenotypic expression of alleles. We recruited subjects from institutions in many regions of the United States (West, Northwest, Midwest, Northeast, and South) during test development, and from a range of institution types, including community colleges, minority-serving institutions, small liberal arts colleges, and large research institutions. We included introductory and advanced undergraduates across a wide range of demographics throughout the initial interviews, test revisions, and final testing, and made use of input from experts to improve the items.

The DCI has a high and comparable level of internal reliability in both computer- and paper-based implementations, which allows for flexibility in implementation. Both the internal and test-retest reliability of the DCI are within recommended ranges and compare well with other published CIs. Additionally, test item characteristics are comparable with those for other published CIs. Item difficulty and item discrimination values are both generally in the desired ranges. However, items 6 and 13 had very low discrimination; these items also had components that were tangentially related to the DCI core concepts and alternative conceptions (see Supplemental Material: Answer Key and Alternative Conception Alignment for DCI). One agree/disagree option within item 13 stated, "Alleles that are harmful are quickly removed from populations"; this option counted for much of the difficulty and lack of discrimination. Although this alternative conception regularly appeared in the pilot

interviews, it is not directly related to dominance. We chose to remove that option (reflected in the Supplemental Material: DCI), which raised item 13 discrimination from 0.07 to 0.11. Similarly, item 6 (see Supplemental Material: DCI) is not a central concept or alternative conception. Both low- and high-performing subjects found item 6 relatively easy, so it had relatively low discrimination ( $D = 0.11$ ). However, we retained this item in the final DCI, because it includes answer choices that differ from those asked in previous questions.

Difficulty and discrimination were fairly consistent across the various testing subgroups, with the exceptions of items 4 and 6 (Figure 3). In each case, one of the two 2Y-Intro courses was primarily responsible for the low discrimination values; all of the students in this course scored similarly on the items. Across the DCI, the directionality and magnitude of differences in performance among subject populations were not consistent. For instance, on some items, introductory students from 2-yr institutions (2Y-Intro) outperformed advanced students from 4-yr institutions (4Y-Adv), while the pattern was reversed on other items (Figure 3A).

While other CIs address genetics, such as the GCA (Smith *et al.*, 2008) and the Genetics Literacy Assessment Instrument (Bowling *et al.*, 2008), they are broad in scope. For instance, the 25-item GCA has a single item that addresses alternative conceptions related to dominance. While these broader CIs are appropriate for assessing an entire genetics course, the DCI is a more narrowly focused tool that can be used to help quantify the prevalence of alternative conceptions in students or assess the effectiveness of a

specific teaching tool or approach. For instance, some suggested approaches to teaching dominance are to separate instruction on heritability from instruction on gene expression (Allchin, 2000; McElhinny *et al.*, 2012; Redfield, 2012). Allchin (2000) recommends teaching complete dominance as a rare phenomenon, rather than the default condition. Others suggest teaching meiotic division before Mendelian genetics (Moll and Allen, 1987). These approaches may help students overcome some of the common dominance-related alternative conceptions, but the lack of an instrument limits an instructor's ability to assess the effectiveness of his or her revised teaching units or activities. It is our hope that the DCI will aid instructors in gauging the impact of such curricular changes on student conceptions about dominance.

Additionally, although we focused on four main alternative conceptions, we retained some distracters from the pilot study that related to other documented alternative conceptions. For instance, in Q2 we present an option that matches the probability of dominant and recessive phenotypes in a monohybrid cross of two heterozygotes. Students often default to ratios they have memorized from previous problems when considering new genetic problems (Browning and Lehman, 1988). Instructors may wish to use the DCI as a starting point for class discussion based on student responses to these items.

The DCI, as a narrowly focused CI, has limitations. The short nature of the DCI and selected-response format preclude deeper explorations of alternative conceptions about dominance and phenotypic expression of alleles. Additionally, the DCI focuses on a subset of potential concepts and alternative conceptions related to Mendelian and population genetics, and therefore should not be used as the sole summative assessment instrument in courses. Because the DCI focuses primarily on patterns students associate with dominance, little can be said about student understanding of the many molecular mechanisms through which the phenomenon of dominance manifests.

### Prevalence of Alternative Conceptions

The DCI targets four alternative conceptions described in the literature or identified through student interviews (Table 3). We found that each of these alternative conceptions is fairly prevalent across our student populations, including students in upper-division courses, which suggests that the DCI would not suffer from a ceiling effect if used throughout a biology curriculum. Moreover, the rank order of the alternative conceptions in terms of prevalence was stable across the three subject populations (2Y-Intro, 2Y-Intro, and 4Y-Adv; Figure 2; Table 4). More than one-third of student responses were linked to DomFreq and DomSelect alternative conceptions in both the pilot and the main study, and between 20 and 30% of possible responses were linked to DomInc and HeteroSelect (Figure 1). While it is possible that the prevalence of DomFreq is overestimated in the main study (after expert review, only two items remain that are linked to this alternative conception), the frequency with which we sampled these alternative conceptions in both populations supports anecdotal evidence from instructors and the published literature.

Some of the confusion apparent among undergraduate students in genetics could be due to their prior learning and pre-

conceptions on the subject. However, some confusion may be due to a lack of comprehension of basic genetic terms, such as gene and allele. Cho *et al.* (1985) found that a variety of high school textbooks use the terms gene and allele interchangeably, without clarification. Lewis *et al.* (2000) found in a survey of British high school students that only one-third of their sampled students recognized the term allele, and only 3% of the total survey produced a scientifically accurate definition. In that same survey, 19% stated that a gene is bigger than a nucleus in a size-sorting activity. Lewis *et al.* (2000) also found that many students in their survey focused on the societal implication of genetics, such as the use of DNA to identify individuals, implying that most of their understanding of genes was from media sources, rather than the classroom or textbooks. Marbach-Ad (2001) found that ninth- and 12th-grade students in Israel considered the terms gene and trait interchangeable.

This problem is not limited to high school, and in fact may be exacerbated by college-level instruction. While our data are from a single item in the pilot study, it is clear that, even at advanced levels, students may struggle with basic terminology. Similarly, Hiatt *et al.* (2013) found that 8.9% of the biology majors they surveyed misused vocabulary related to genetics, using the terms gene, allele, and genome interchangeably in open-response questions, and failed to distinguish among the terms in follow-up interviews that focused on use of those words. Hiatt *et al.* (2013) also found students conflated gene expression (i.e., transcription) with the phenotypic expression of an allele, which Allchin (2000), Donovan (1997), and Lewis and Kattmann (2004) suggest could play a role in student confusion about dominance. A study by Pashley (1994) found that college students in the United Kingdom tend to confuse gene and allele, but encouragingly, when this conceptual difficulty is remedied, overall genetics understanding increases.

The multiple definitions for biological terms may also play a role in student alternative conceptions. Confusions about the meaning of dominance may be driven by an equivocation of its colloquial, behavioral, and genetic definitions (Allchin, 2002; Donovan, 1997). For instance, dominance within groups of social animals is often based on direct competition with other group members; it is unsurprising that students who learn about social behavior might misattribute the same dynamic to gene expression. In addition, students appear to confuse the pattern of dominance in genetics with *predominant*, meaning "most common." The HeteroSelect alternative conception may be an inappropriate extension of overdominance or heterosis, both of which are commonly discussed in introductory courses and textbooks (e.g., Campbell *et al.*, 2008; Miller and Levine, 2010; Freeman *et al.*, 2013; Phelan, 2013). It is not difficult to imagine that students might extend the classic sickle-cell example of overdominance to other examples of heterozygotes.

These references and our work demonstrate that students have considerable difficulty with terminology, which may hinder their understanding of genetics. In this paper, we demonstrate this with data on student understanding of the terms gene and allele, and speculate on other contributing factors to difficulties with the terms recessive and dominant. It may be, as Allchin (2002) argues, that replacing these culturally significant terms with more neutral terms may better serve students and science.



## Future Research

Although we have documented several alternative conceptions, we acknowledge that the DCI covers only a subset of student conceptions about dominance. Work remains to explore influences on student conceptions of gene regulation and gene expression and student understanding of potential mechanisms by which traits coded for by recessive alleles could be masked in heterozygotes. More immediately, we plan to make use of the DCI to assess the effectiveness of the above recommended changes in our approach to teaching Mendelian genetics. The utility of the DCI would be extended with additional validity evidence for its use in high school biology courses and with the general public. Dominance is typically first covered in grades 9–12, and gene expression is a core component of many state standards, including the Next Generation Science Standards being adopted by 26 states (Achieve, 2013). Lanie *et al.* (2004) and the BBC World News, report discussed above (BBC News, 2002) highlight a number of common alternative conceptions the public holds about genetics. As advances in genetics become a larger part of daily life, it is important that we continue to develop tools to assess changes in understanding and the impact of informal science interventions.

## ACKNOWLEDGMENTS

A portion of this study was supported by grant DUE-0717495 from the National Science Foundation (NSF). Additional support was provided by the National Evolutionary Synthesis Center (NSF grant number EF-0905606). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. We thank the students who participated in this study and the instructors who gave up class time in support of this project; expert reviewers; student research assistants Gloria Blanquel, Calvin Lung, and Matthew Bischoff; other members of the EvoCI Toolkit Working Group, especially Kathleen Fisher, Terri McElhinny, Mike U. Smith, Ryan Walker, and NESCent. We received invaluable support and input from Jon Herron and from SimBio employees, particularly Eli Meir and Susan Maruca. Finally, we thank two anonymous reviewers for their helpful comments on an earlier draft of the manuscript.

## REFERENCES

- Abraham JK, Perez KE, Downey N, Herron JC, Meir E (2012). Short lesson plan associated with increased acceptance of evolutionary theory and potential change in three alternate conceptions of macroevolution in undergraduate students. *CBE Life Sci Educ* 11, 152–164.
- Achieve (2013). Next Generation Science Standards. [www.nextgenscience.org](http://www.nextgenscience.org) (accessed 7 April 2014).
- Allchin D (2000). Mending Mendelism. *Am Biol Teach* 62, 633–639.
- KluwerAllchin D (2002). Dissolving dominance. In: *Mutating Concepts, Evolving Disciplines: Genetics, Medicine, and Society*, ed. L Parker and R Ankeny, Dordrecht, Netherlands: Kluwer.
- Allchin D (2005). The dilemma of dominance. *Biol Philos* 20, 427–451.
- Anderson DL, Fisher KM, Norman GJ (2002). Development and evaluation of the conceptual inventory of natural selection. *J Res Sci Teach* 39, 952–978.
- BBC News (2002). Blondes “to die out in 200 years.” <http://news.bbc.co.uk/2/hi/health/2284783.stm> (accessed 28 June 2013).
- Bowling BV, Acra EE, Wang L, Myers MF, Dean GE, Markle GC, Moskalik CL, Huether CA (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genetics* 178, 15–22.
- Browning ME, Lehman JD (1988). Identification of student misconceptions in genetics problem solving via computer program. *J Res Sci Teach* 25, 747–761.
- Campbell NA, Reece JB, Urry LA, Cain ML, Wasserman SA, Minorsky PV, Jackson RB (2008). *Biology*, 8th ed., San Francisco, CA: Benjamin Cummings.
- Cho H, Kahle JB, Nordland FH (1985). An investigation of high school biology textbooks as sources of misconceptions and difficulties in genetics and some suggestions for teaching genetics. *Sci Educ* 69, 707–719.
- Christensen AC (2000). Cats as an aid to teaching genetics. *Genetics* 155, 999–1004.
- Collins A, Stewart JH (1989). The knowledge structure of Mendelian genetics. *Am Biol Teach* 51, 143–149.
- Cortopassi GA (2002). Fixation of deleterious alleles, evolution and human aging. *Mech Ageing Dev* 123, 851–855.
- Crocker L, Algina J (1986). *Introduction to Classical and Modern Test Theory*, Orlando, FL: Holt, Rinehart and Winston.
- Cronbach LJ (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334.
- Crowther GJ, Price RM (2014). Re: misconceptions are “so yesterday.” *CBE Life Sci Educ* 13, 3–5.
- Donovan MP (1997). The vocabulary of biology and the problem of semantics. *J Coll Sci Teach* 26, 381–382.
- Elrod S (2007). Genetics Concept Inventory. <http://bioliteracy.colorado.edu/Readings/papersSubmittedPDF/Elrod.pdf> (accessed 20 July 2013).
- Finley FN, Stewart J, Yaroch WL (1982). Teachers’ perceptions of important and difficult science content. *Sci Educ* 66, 531–538.
- Freeman S, Quillin K, Allison L (2013). *Biological Science*, 5th ed., San Francisco, CA: Benjamin Cummings.
- Haladyna T (2004). *Developing and Validating Multiple-choice Test Items*, Mahwah, NJ: Erlbaum.
- Heim WG (1991). What is a recessive allele? *Am Biol Teach* 53, 94–97.
- Hiatt A, Davis GK, Trujillo C, Terry M, French DP, Price RM, Perez KE (2013). Getting to evo-devo: concepts and challenges for students learning evolutionary developmental biology. *CBE Life Sci Educ* 12, 494–508.
- Lanie AD, Jayaratne TE, Sheldon JP, Kardias SLR, Anderson ES, Feldbaum M, Petty EM (2004). Exploring the public understanding of basic genetic concepts. *J Genet Couns* 13, 305–320.
- Lewis J, Kattmann U (2004). Traits, genes, particles and information: re-visiting students’ understandings of genetics. *Int J Sci Educ* 26, 195–206.
- Lewis J, Leach J, Wood-Robinson C (2000). All in the gene? Young people’s understanding of the nature of genes. *J Biol Educ* 34, 74–79.
- Marbach-Ad G (2001). Attempting to break the code in student comprehension of genetic concepts. *J Biol Educ* 35, 183–189.
- Marshall MN (1996). Sampling for qualitative research. *Fam Pract* 13, 522–525.
- Maskiewicz AC, Lineback JE (2013). Misconceptions are “so yesterday!” *CBE Life Sci Educ* 12, 352–356.
- McElhinny TL, Dougherty MJ, Bowling BV, Libarkin JC (2012). The status of genetics curriculum in higher education in the United States: goals and assessment. *Sci Educ* 23, 445–464.
- Miller KR, Levine JS (2010). *Biology*, Upper Saddle River, NJ: Prentice Hall.

- Mills Shaw KR, Van Horne K, Zhang H, Boughman J (2008). Essay contest reveals misconceptions of high school students in genetics content. *Genetics* 178, 1157–1168.
- Moll MB, Allen RD (1987). Student difficulties with Mendelian genetics problems. *Am Biol Teach* 49, 229–233.
- Pashley M (1994). A-level students: their problems with gene and allele. *J Biol Educ* 28, 120.
- Perez KE, Hiatt A, Davis GK, Trujillo C, French DP, Terry M, Price RM (2013). The EvoDevoCI: a concept inventory for gauging students' understanding of evolutionary developmental biology. *CBE Life Sci Educ* 12, 665–675.
- Phelan J (2013). *What Is Life? A Guide to Biology with Physiology*, 2nd ed., New York: Freeman.
- R Development Core Team (2012). R: A Language and Environment for Statistical Computing. [www.R-project.org](http://www.R-project.org) (accessed 20 July 2013).
- Redfield RJ (2012). "Why do we have to learn this stuff?" A new genetics for 21st century students. *PLoS Biol* 10, e1001356.
- Rousson V, Gasser T, Seifert B (2002). Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Stat Med* 21, 3431–3446.
- Smith MK, Wood WB, Knight JK (2008). The Genetics Concept Assessment: a new concept inventory for gauging student understanding of genetics. *CBE Life Sci Educ* 7, 422–430.
- Weir JP (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 19, 231–240.