Evolution: Education and Outreach

**RESEARCH ARTICLE**

CrossMark

# Do disciplinary contexts impact the learning of evolution? Assessing knowledge and misconceptions in anthropology and biology students

Elizabeth P. Beggrow[1*] and Gena C. Sbeglia[2]

## Abstract

**Background:** Evolution education research has focused on biology populations, while other disciplines organized around evolutionary theory—such as biological anthropology—remain understudied. Cognitive science and education research suggest that learning evolution within the context of human evolution might cause increased understanding of evolutionary theory, as well as reasoning patterns relating to evolutionary change different from those stemming from learning evolution in a more generalized context. Biological anthropology students could offer a test of this hypothesis. This study incorporates this underrepresented population into the evolution education literature in order to generate insights into the effects of disciplinary context on evolutionary knowledge and reasoning.

**Methods:** Undergraduate biology and anthropology students (N = 268) completed two validated and published evolution knowledge instruments: Conceptual Inventory of Natural Selection (CINS) and Assessment of Contextual Reasoning about Natural Selection (ACORNS). We varied the surface features of the ACORNS items [i.e., the trait (familiar, unfamiliar) and taxon (human, non-human)] and evaluated if the populations differed in their instrument scores or sensitivity to item surface features.

**Results:** The populations differed in background and demographic variables. Evolutionary knowledge and reasoning patterns also differed, with biology students having higher CINS scores, more key concepts, fewer naive ideas, and higher frequencies of accurate reasoning models. However, scores were generally poor for both populations. When background and demographic factors were controlled, key concept scores were comparable, but anthropology students continued to display lower measures for the other variables. Additionally, biology students' showed limited sensitivity to the item surface features compared to anthropology students.

**Conclusions:** Anthropology and biology students displayed significantly different demographic and academic backgrounds, making comparisons complex. We did not find evidence that learning evolution within a human/primate context generated greater instrument scores or lower sensitivities to item surface features. Rather, both anthropology and biology students were novice-like in their evolutionary knowledge and reasoning patterns. Anthropology students were more novice-like because they displayed sensitivities to item surface features. Our study raises questions about how best to assess the relative impact of taxon and trait familiarity on the measurement of evolutionary knowledge and reasoning.

**Keywords:** Evolution, Anthropology, Human evolution, Knowledge, Misconceptions, Context, Undergraduates, Surface Features

*Correspondence: beggrow.7@osu.edu
[1] Center for Life Sciences Education, The Ohio State University, 1735 Neil Avenue, Columbus, OH 43210, USA
Full list of author information is available at the end of the article

## Background

Although undergraduate coursework in both biology and biological anthropology uses evolutionary theory as a guiding and organizing principle (AAAS 2011; Fuentes 2011), the contextualization of disciplinary core ideas (e.g., heredity, evolution) differs in important ways. For example, while introductory courses in the biological sciences typically engage with a range of core ideas using an array of taxonomic contexts from across the tree of life (e.g., archaea, bacteria, fungi, plants, animals), anthropology courses cover a somewhat overlapping range of core ideas illustrated using human and other primate examples (e.g., Fuentes 2011). These differences in curricular contexts (e.g., primate focused vs. tree of life focused) provide an untapped research context for evolution education studies. Indeed, many evolution education research studies have utilized biology majors and non-majors to explore the challenges inherent to teaching and learning about evolution (e.g., Nehm and Reilly 2007; Gregory 2009). Interestingly, anthropology undergraduates have not received comparable attention in evolution education research even though evolution also serves as a core feature of that discipline. The overarching goal of our work is to begin to explore evolutionary knowledge and reasoning patterns in anthropology undergraduates, and to compare these findings to samples of biology undergraduates.

## Anthropology, biology, and evolution education

Differences in how evolution is contextualized (e.g., focus on primates) means that evolutionary topics covered in anthropology, such as inheritance, mutation, and phenotypic variation, are often situated within human examples. This focus could provide advantages for students in terms of learning evolution. For one, humans appear to be intrinsically interested in themselves (e.g., Pobiner 2012) and our cognitive tendency to easily differentiate individuals may help to overcome cognitive biases that hamper evolutionary thinking (i.e., essentialism; Sinatra et al. 2008). Anthropology students also learn about variation within a species, which is often a significant barrier to understanding natural selection (Gregory 2009). Finally, the discovery of new fossil taxa creates excitement and interest beyond the sciences, and is often associated with dilemmas and debates (e.g., Does the variation found in a new fossil exemplify intraspecific variation, or should it be named a new species?). The discoveries provide important opportunities for discussing and exploring the nature of science, which in and of itself has been associated with improved understanding of core concepts (e.g., Dagher and BouJaoude 1997; Kampourakis and Zogza 2009).

The potential advantages of teaching evolution using anthropological contexts have not gone unnoticed in science education research. A number of anthropologists and science education researchers have written about the importance of including human examples in evolution education (e.g., Alles and Stevenson 2003; Ashmore 2005; Cunningham and Wescott 2009; DeSilva 2004; Flammer 2006; Hillis 2007; Nickels et al. 1996; Paz-y-Miño and Espinosa 2009; Pobiner 2012, 2016; Price 2012; Wilson 2005) and some have investigated incorporating human examples into biology curricula (e.g., deSilva 2004; Flammer 2006; Price 2012; Pobiner et al. 2018). While there is much evidence to suggest that anthropology curricula may offer a unique and advantageous way of learning evolutionary theory, there has been no formal, comparative research to test this hypothesis. Rather, the current body of work on students' understanding of evolution, their non-normative ideas, and their acceptance of evolutionary theory is primarily based on populations of biology students, teachers, and experts. Studies investigating these traits in biological anthropology students are extremely rare and the results are not readily comparable to other populations, which limits any tests of the role that disciplinary context plays in evolution learning. For instance, Cunningham and Wescott (2009) surveyed students enrolled in an introductory biological anthropology course and found that, despite widespread agreement on the validity of biological evolution, many students held a number of misconceptions regarding evolutionary theory and the nature of science. However, this study was not conducted using published and validated measurement instruments, so it is unclear how these scores compare to populations in other studies or if the inferences generated from these scores are robust. Therefore, the relative evolutionary knowledge of populations of anthropology students, teachers and experts and the impact that human-focused evolution instruction has on that knowledge are in need of additional exploration.

## Learning evolution using human contexts

Folk biology has explored individuals' reasoning about biological kinds and has found that US children utilize essentialism, or an assumption of an underlying causal nature of a kind, in their biological reasoning (Gelman and Wellman 1991; Wellman and Gelman 1992). Similar findings have also been found in other cultures and populations (e.g., Atran 1998; Bishop and Anderson 1990; Gregory 2009; Medin and Atran 2004; Shtulman 2006). These biases extend to the classroom, where learners often do not consider the magnitude of variation within species (Shtulman and Schulz 2008), and they consequently perceive all members of a species as nearly the same (Gregory 2009). Nonetheless, Shtulman and Schulz

(2008) found that an appreciation of individual-level variation by learners is related to a correct understanding of the mechanisms of natural selection, suggesting that learners can overcome this cognitive bias. Because individual variation is crucial to population thinking, essentialistic thinking creates potential obstacles for understanding evolutionary theory, particularly the ideas that species are immutable categories or that variation is best conceptualized as 'noise' (Gelman and Legare 2011). These obstacles impede learners' grasp of within-species variation, and, ultimately, a firm understanding of the processes responsible for evolutionary change.

Typological biases could be the result of evolutionary processes favoring expediency and efficiency. Primates exhibit many social-cognitive abilities in order to facilitate interactions with conspecifics (Axelrod and Hamilton 1981; Barret and Henzi 2005a, b; Dunbar 1993, 1998; Hammerstein 2003; de Waal 1997a, b; Humphrey 1974). Forming coalitions, bonding through grooming, and an overall awareness of who to affiliate with versus who to avoid, are crucial skills for social primates, particularly humans. Indeed, Humphrey (1974) found evidence in rhesus macaques that cognition regarding conspecifics is individual-oriented, while cognition about allospecifics tended to be species-oriented. As of 2018, 55% of people worldwide live in urban areas (Population Division 2018), and for this proportion of the global population, interactions with large numbers of non-human animals are limited. When considering our own evolutionary history, intraspecific interactions certainly outweigh interspecific ones (Medin and Atran 2004) and, cognitively speaking, it seems as though humans operate accordingly.

Although a bias to think in 'kinds' has been documented for people reasoning about non-human animals and plants, there has been research demonstrating that it does not always hold for thinking about other humans, at least biologically (Birnbaum et al. 2010; Rhodes and Gelman 2009). Situating a biological phenomena in a human context appears to change the cognitive principles at play, and thinking about individual variation becomes more 'comfortable' when we are thinking about humans (Nettle 2010). Indeed, support was found among British university students for a stronger tendency towards individual-based reasoning when that reasoning was focused on humans as opposed to non-human animals (Nettle 2010). When reasoning about humans, students were specifically more likely to think that adaptive change could occur within species instead of the species going extinct and/or being replaced by a novel species (as they did with non-human animals), and they were more likely to accept the idea that individuals did not have to change within a lifetime for population-level changes to occur. Furthermore, when reasoning about humans, students

were less likely to think that novel features would automatically become ubiquitous among the entire species and tended not to view competition as a driver of evolutionary change. However, Nettle did find that reasoning about human evolution had no effect on two non-normative ideas: that of the utility of a feature correlating with mutation and heredity (i.e., use/disuse), as well as the notion that change is driven by species needs (i.e., teleology) (2010). Nonetheless, Nettle's (2010) findings support the idea that different domain-specific cognitive biases exist for reasoning about human versus non-human animals (Atran 1998; Atran et al. 2001; Medin and Atran 2004).

Beyond overcoming essentialistic biases, studying evolution using humans may provide other advantages. Some studies suggest that students would actually prefer to learn evolution in the context of humans and that the topic could be a motivational factor (Pobiner et al. 2018; Schrein 2017; Paz-y-Miño and Espinosa 2009; Hillis 2007; Wilson 2005). For example, when asked for feedback on how their experience with the human evolution case studies instructional material compared to previous experiences with evolution content, a majority of students responses were coded as positive and indicated a preference for human examples (Pobiner et al. 2018). A similar preference for learning evolution with human examples was found in both biology majors and non-majors (Paz-y-Miño and Espinosa 2009). These studies suggest that the situations and contexts in which students learn about evolution make a difference.

### Situated cognition and learning

Although learning concepts (e.g., evolution) within a particular taxonomic context (e.g., primates) can have advantages, it can also produce disadvantages (Anderson et al. 1996). In terms of knowledge application, an optimal recipe for learning is a combination of concrete and abstract examples (Anderson et al. 1996). This suggests that learning environments in which evolutionary concepts are taught across a range of contexts should foster improved application skills (e.g., Nehm 2018). It follows that while learning evolutionary concepts situated within anthropology may lead to an ability to apply those concepts within human-related contexts, it may *not* foster the ability to apply those concepts across the tree of life (e.g. both human and non-human contexts).

Within the situated cognition perspective, there is an assumption that knowledge is dependent on the situation(s) in which it is learned and used (Seely Brown et al. 1989). From this perspective, all learning is situated within the context of the social and cultural setting in which it takes place, whether that is in the classroom or out in the community (Sawyer and Greeno 2009).

Although there is debate concerning what it means "to be situated" (Adams and Aizawa 2009; Wilson and Clark 2009), a basic tenet is that cognitive processes are both social and neural, and that knowledge itself is seen as dynamic (in terms of learning, remembering and reinterpreting) and contextualized (Clancey 2009). The contextualization of knowledge can be explored at many different scales, ranging from the social nature of the learning environment to more fine-grained questions relating to assessment tasks.

### Situated cognition, familiarity and reasoning

Novice reasoning is inextricably linked to the context in which it is situated, thus the specific features of that context may contribute to the framing and conceptualization of any problem a novice may face and be a critical part of novice reasoning (Kirsh 2009). The features of a problem that elicit these contextual effects in novice learners are called surface features. The effects of surface features on knowledge acquisition, retrieval and problem solving have been widely investigated within cognitive science (e.g., Caleon and Subramaniam 2010; Chi et al. 1981; DiSessa et al. 2004; Gentner and Toupin 1986; Sawyer and Greeno 2009; Evans et al. 2010; Sabella and Redish 2007). Within biology, the impact of surface features has been explored in a variety of studies, some of which explored context effects in genetics (see Schmiemann et al. 2017 for a review), though most of the research has been focused on understanding of natural selection (e.g., Bishop and Anderson 1990; Clough and Driver 1986; Federer et al. 2015; Kampourakis and Zogza 2009; Nehm et al. 2012; Nehm and Ha 2011; Nehm and Reilly 2007; Nehm and Ridgway 2011; Opfer et al. 2012; Settlage, 1994). Evolutionary biology is perhaps more sensitive to issues of contextuality than other science domains, namely due to the fact that the units of evolution (individuals and species) already vary across space and time (Nehm and Ha 2011), which may make reasoning about these concepts more susceptible to contextual effects. Nehm and colleagues have found evidence for contextual feature effects with assessment items designed to elicit knowledge and non-normative ideas about evolution (Federer et al. 2015; Nehm et al. 2012; Nehm and Ha 2011; Nehm and Reilly 2007; Nehm and Ridgway 2011; Opfer et al. 2012). The reasoning patterns elicited by these items were impacted by the item's surface features such as the taxon in question (e.g., plant/animal/human), the polarity of evolutionary change in traits (e.g., loss or gain of trait) and the familiarity of the taxon and trait in question (e.g., lily vs. labiatae), though such effects diminish as expertise increases (e.g., Nehm and Ridgway 2011; Opfer et al. 2012).

Young children, the quintessential novices, are thought to hold a theory-like structure of naive ideas in biology that includes the necessary knowledge to recognize biological things and phenomena despite little formal education on the topic, but lack the normative ideas about how those phenomena operate (e.g., Inagaki and Hatano 2006; Opfer et al. 2012). For example, children envision plants and animals as separate categories and vary accordingly in how they apply biological ideas to these concepts (e.g., Carey 1986; Inagaki and Hatano 1996; Opfer and Siegler 2004). Furthermore, children will use their understanding of humans as an analog to reason about plants and animals or novel situations (Inagaki and Hatano 2002). This is a potentially useful feature of reasoning that could be leveraged in evolutionary biology instruction by using familiar human examples as a bridge to the less familiar non-human examples (Seoh et al. 2016).

That humans may be thought of as "familiar" is both logical and inferred from research. Beyond the advantages addressed above, the *familiarity* of the construct 'humans' could impact learners when asked to reason about evolutionary change, but there has been little research to determine whether this impact is positive or negative. In their study developing and piloting human evolution case studies, Pobiner and colleagues found gains in measures of understanding post-instruction on an assessment asking students to explain evolutionary change in humans and a non-human taxon (2018). It is important to note, however, that the measures of understanding for this study did not include naive ideas, which, in addition to accurate key concepts, have been found to be higher when students are asked about evolution in familiar taxa compared to unfamiliar taxa (Federer et al. 2015). In contrast to Pobiner and colleagues' findings, Ha and colleagues (2006) looked at student explanations of evolution of human, animal and plant traits and found a negative effect of human taxon category on the responses. Specifically, they found that when asked about human evolution, students' explanations were less likely to explain evolutionary change using natural selection and that both human and animal items were more likely to elicit misconceptions regarding the use/disuse of traits and intentionality (Ha et al. 2006). These studies raise questions about the relationship between context of learning, context of assessment and the reasoning patterns elicited. More specifically, it remains to be seen how these surface features, whose effects on populations of biology learners are better documented for some features (familiarity) over others (taxon category), impact learners whose evolution education is situated completely within the primate/human lineage (i.e., anthropology).

## Research questions

Using a comparative, quantitative research design, this study explores the following research questions:

(RQ 1)  How similar are the students that enroll in anthropology and biology classes?

(RQ 2)  Do evolutionary knowledge and naive ideas differ across anthropology and biology students? If so, how?

(RQ 3)  Is variation in evolutionary knowledge and naive ideas across these populations explained by background and demographic variables?

(RQ 4)  To what extent do surface features impact each population's evolutionary knowledge and naive ideas? Specifically, do evolutionary knowledge and naive ideas differ based on: (RQ 4.1) the taxon (human vs. non-human)? (RQ 4.2) the familiarity of the trait?

## Methods

### Recruitment and instruments

Data were gathered from undergraduate students enrolled in an introductory biological anthropology course (referred to here as anthropology) and an introductory biology course at a large, public, Midwestern university. Courses were sampled once towards the end of the fall semester of 2012. Both courses count towards fulfilling the Natural Science GEC requirement and both require students to enroll in a laboratory component. Learning objectives for both courses included being able to explain the mechanisms of evolution (including genetic drift, natural selection, sexual selection) and how they relate to patterns of speciation and extinction (see Additional file 1: Appendix 1). Approximately seven lecture hours and three laboratory sessions (55 min each) in anthropology were designated for basic evolution content (history of evolutionary thought, cell biology/inheritance/DNA basics, heredity, population genetics, evolutionary mechanisms, macroevolution, modern human variation). Approximately eight lecture hours and three laboratory sessions (2 h each) in biology were designated for basic evolution content (artificial selection and natural selection, microevolutionary mechanisms, macroevolution and systematics, population genetics). Overall, both courses covered the same basic evolutionary concepts for roughly equal amounts of time, while the context in which they were taught differed.

Students were recruited to participate in an online survey accessed through SurveyMonkey®. Points were awarded to students who participated in surveys based on the discretion of the instructors. Though amounts varied between sections, all amounts were nominal relative to the total grades. Surveys were comprised of a consent agreement, a section for demographic information, and three instruments. Demographic information (e.g., gender, year, and ethnicity) was gathered in accordance with IRB approval, as well as information concerning whether English was a first language, previous college level biology courses taken, and previous college level anthropology courses taken. Although participants were asked to identify cultural anthropology courses previously taken in the survey, these courses were not included in the analysis of the data. Year in school was coded as freshman, sophomore, junior or senior. Ethnicity was collapsed into two categories and coded as either white only or non-white. In addition to the aforementioned survey components, the survey was comprised of (1) the multiple-choice Conceptual Inventory of Natural Selection (CINS) instrument (Anderson et al. 2002), (2) the open-response Assessment of Contextual Reasoning about Natural Selection (ACORNS) instrument (Nehm et al. 2012), and (3) a familiarity rating scale for 28 biological terms (see Additional file 2: Appendix 2).

### CINS

The multiple-choice CINS instrument consists of 20 items with one correct response option. Each item's alternate answer choices were designed to address typical non-normative ideas regarding natural selection (Anderson et al. 2002). The items are scored as correct/incorrect, providing a total score ranging from 0 to 20. Although the CINS has been reported to display some psychometric limitations (Battisti et al. 2010), it is a widely used instrument for natural selection knowledge and is generally recognized as an instrument capable of generating valid inferences about general levels of participants' evolutionary knowledge (Smith 2010). The original CINS paper purports it to be a test of natural selection knowledge, but its questions about speciation mean that the concept of macroevolution is addressed (Futuyma 2009), making it a test of both micro- and macroevolutionary concepts.

### ACORNS

The ACORNS is an open-response instrument that asks participants to reason about evolutionary change. The items prompt participants to explain mechanisms that account for between-species change, thereby testing both micro- and macroevolutionary knowledge. Previous work has shown the test to generate valid and reliable inferences among populations of university level biology students (Beggrow et al. 2014; Beggrow and Nehm 2012; Nehm et al. 2012; Nehm and Ha 2011). We developed eight isomorphic items in which we varied the taxon and

**Table 1  ACORNS Items**

| Taxon | Familiar | Unfamiliar |
|---|---|---|
| Non-human | *Item 1:* How would scientists explain how a descendant species of *dolphin* with a large brain evolved from an ancestral species of *dolphin* that lacked a large brain?[a] | *Item 3:* How would scientists explain how a descendant species of *horse* with a navicular evolved from an ancestral species of *horse* that lacked a navicular?[a] |
|  | *Item 2:* How would scientists explain how a descendant species of *camel* with long eyelashes evolved from an ancestral species of *camel* that lacked long eyelashes?[a] | *Item 4:* How would scientists explain how a descendant species of *koala* with dermatoglyphics evolved from an ancestral species of *koala* that lacked dermatoglyphics?[a] |
| Human | *Item 5:* How would scientists explain how a descendant species of *human* with a large brain evolved from an ancestral species of *human* that lacked a large brain?[b] | *Item 7:* How would scientists explain how a descendant species of *human* with a navicular evolved from an ancestral species of *human* that lacked a navicular?[b] |
|  | *Item 6:* How would scientists explain how a descendant species of *human* with long eyelashes evolved from an ancestral species of *human* that lacked long eyelashes?[b] | *Item 8:* How would scientists explain how a descendant species of *human* with dermatoglyphics evolved from an ancestral species of *human* that lacked dermatoglyphics?[b] |

[a]  Items present on version 1 of the survey. Version 1 consisted of only non-human items and was assigned to half of the biology and anthropology student populations

[b]  Items present on version 2 of the survey. Version 2 consisted of only human items and was assigned to the other half of the biology and anthropology student populations

the trait. Specifically, half of the items used non-human taxa (i.e., dolphin, camel, horse, koala) and the other half used humans (Table 1). Likewise, half of the items used familiar traits (i.e., brain, eyelashes) and the other half used unfamiliar traits (i.e., navicular, dermatoglyphics) (Table 1). Familiarity of the taxa and traits was hypothesized apriori using Google™ PageRank (cf., Nehm et al. 2012; see Additional file 2: Appendix 2) and confirmed a posteriori. We intended for half of the traits and all of the taxa to be familiar to the survey respondents. All items focused on trait gain. We organized these eight isomorphic items into two versions of the survey: a version focusing on evolution of the four traits in non-human animals (items 1–4) and a version focusing on evolution of the same four traits but in humans (items 5–8) (Table 1). Half of the biology students and half of the anthropology students were assigned to each version of the survey and each student took only one version.

The order of the ACORNS items in the survey was randomly generated for each participant to help control for order effects on responses (e.g., Federer et al. 2015). The ACORNS responses were scored using automated scoring models (EvoGrader; Moharreri et al. 2014) developed to assess the accuracy of nine evolutionary concepts: six key concepts (KCs; variation, heritability, competition, limited resources, differential survival/reproduction, and non-adaptive reasoning) and three naive ideas (NIs; adapt, need, use/disuse) (Nehm et al. 2010; Beggrow et al. 2014). KC scores for each item ranged from 0 to 6 (referred to as per-item total KC), and NI scores for each item ranged from 0 to 3 (per-item total NI). A sum of all KCs used across all four items generated a Total KC score and a sum of all NIs used across all four items generated a Total NI score. Model type (MT) was also scored as

either no model (no direct answer to the question), naive model (non-normative ideas only), mixed model (non-normative and normative ideas) or pure scientific model (normative accurate ideas only; Moharreri et al. 2014).

### Students' familiarity with item words

After students completed the open response items, we asked them to rate their familiarity with each trait and taxon along the following scale: (1) "I have never seen/heard the word before" (i.e., unfamiliar), (2) "I have seen/heard the word before but do not know what it means" (i.e., somewhat unfamiliar), (3) "I have seen/heard the word before and may know what it means" (i.e., familiar), (4) "I have seen/heard the word before and are certain of its meaning" (i.e., very familiar). All terms were listed individually and devoid of contextual cues. We asked students to provide self-reported familiarity ratings for terms, including those used in the ACORNS items, to confirm a priori hypotheses of familiarity levels. Terms were chosen based on Google™ PageRank scores to represent a selection of biological and anthropological terms that would range from unfamiliar to familiar for both anthropology and biology students (see Additional file 2: Appendix 2). The ratings also helped to generate more accurate measurements of familiarity that varied for each student; this variation was then included in our models.

### Sample demographics

A total of 654 students took the survey, with three students declining to consent to the study (99.5% consent rate). Out of those surveys, 67 were incomplete and removed from the dataset. If students had taken five or more anthropology or biology courses (7) or had completed or were currently enrolled in both anthropology

and biology courses (109), they were removed from the dataset. Of the 468 remaining students, 19 students had missing demographic or background data and were removed from the relevant analyses.

We classified students as anthropology or biology students depending on their prior and current anthropology and biology courses. For the purposes of this study, anthropology students were classified as those who had completed or were currently enrolled in biological anthropology courses but had not taken any, and were not currently enrolled in any, biology courses (N = 208). Biology students were classified as those who had completed or were currently enrolled in biology courses and had not taken any, and were not currently enrolled in any, biological anthropology courses (N = 260).
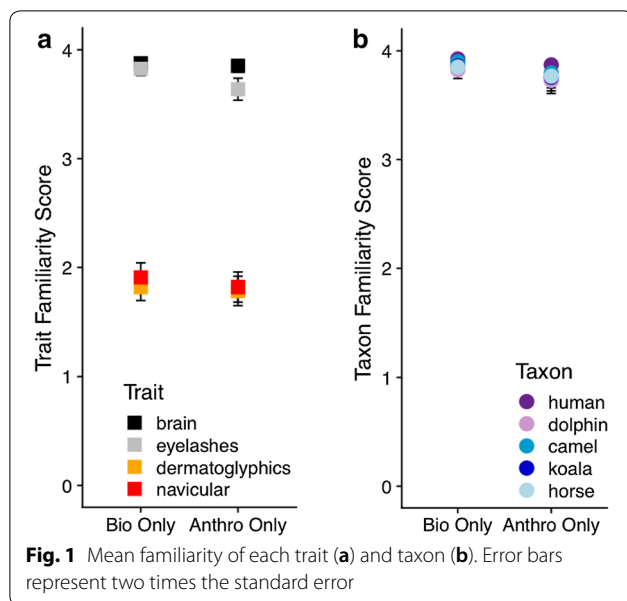
## Analyses
### Validity evidence
New items for the ACORNS instrument were introduced with this study and to establish convergent validity, Kendall's Tau B correlation coefficients were calculated between CINS scores and ACORNS Total KC scores using SPSS v.20. Conversions were made according to Gilpin (1993) in order to make them comparable to published results. The CINS test was used here to establish validity evidence for the ACORNS items because it is considered a proxy for natural selection knowledge (Nehm and Schonfeld 2010).

To address RQ1 (*How similar are the students enrolled in anthropology and biology classes?*), we compared demographic variables (i.e., gender, ethnicity), other student background variables (i.e., year, number of prior or current anthropology or biology courses, word count, English as a first language), and evolution knowledge and reasoning variables (i.e., CINS, ACORNS per-item total KCs, ACORNS per-item total NIs, ACORNS MT) between biology and anthropology students. We compared the distributions of demographic and background variables between biology and anthropology students using a Chi-Squared test. We compared the knowledge and reasoning variables between biology and anthropology students using a suite of regressions aligned with data type. CINS scores are numeric, and were analyzed using a linear regression with student classification as the single independent variable. ACORNS per-item total KCs and per-item total NIs are ordinal and were analyzed using separate cumulative link mixed-effects models with a logit link function via the R package ordinal (v. 2018.8-25; Christensen 2018). ACORNS MT data were converted into binary categorical variables (i.e., pure scientific MT vs. all other MTs) and were analyzed using a generalized linear mixed-effects model via the R package lme4 (Bates et al. 2018). As each student completed four

ACORNS items and thus had four data points for each ACORNS outcome variable, student id was coded as a random effect in these models. These regression models will be built upon in subsequent research questions, and so these models will be referred to as model set 1.

To address RQ2 (*Do evolutionary knowledge and naive ideas differ across anthropology and biology students? If so, how?*), we ran the same class of regression models as described above for CINS scores, ACORNS per-item total KCs, ACORNS per-item total NIs and ACORNS MT but, in addition to including student classification as a predictor variable (as in model set 1), we also included background (i.e., year, number of prior or current anthropology or biology courses, English as a first language) and demographic variables (i.e., gender, ethnicity). These regression models will be referred to as model set 2. With this model, we can then ask RQ3 (*Is variation in evolutionary knowledge and naive ideas across these populations explained by background and demographic variables?*). We report on the impact of student classification for explaining the variation in each of the four knowledge and reasoning outcome variables (CINS scores, ACORNS per-item total KCs, per-item total NIs and MT) while controlling for all background and demographic variables. We report unstandardized regression coefficients (*b*). We examined the effect size of each significant variable using generalized eta squared ($\eta^2 G$) via the R package Analysis of Factorial Experiments (afex, v. 0.21-2) (Singmann et al. 2018). $\eta^2 G$ measures the additional variance explained by a variable as compared to a model in which it was excluded. $\eta^2 G$ can be compared across regression analyses and studies, and is appropriate for use in mixed model designs (Bakeman 2005; Lakens 2013; Olejnik and Algina 2003). The following cutoff values for interpretation can be used: small effect = 0.01, medium effect = 0.06, and a large effect = 0.14 (Olejnik and Algina 2003). We use a critical p-value of 0.01 for all analyses.

To address RQ4 (*How do surface features impact each population's evolutionary knowledge and naive ideas?*), we built upon model set 2 by adding two additional predictor variables that addressed the following surface features: the specific taxon (i.e., human vs. non-human), and the familiarity of the trait (familiar or unfamiliar). These models were run for anthropology students separately from biology students so that we could compare the nature of the impact of surface features for each population. This set of models will be referred to as model set 3 in this paper. We used these models to test if the per-item total number of KCs, NIs, and MTs differ based on the taxon category (RQ 4.1) or the trait familiarity (RQ 4.2). For each of the significant surface feature variables, we report the unstandardized coefficients and $\eta^2 G$. Because

**Fig. 1** Mean familiarity of each trait (**a**) and taxon (**b**). Error bars represent two times the standard error

all surface feature variables were included in the model simultaneously, when observing the impact of a particular surface feature variable, the analysis controls for the impact of all the others. We used a critical p-value of 0.01 for all analyses.

## Results

### Validity evidence

Kendall's Tau B correlation analyses revealed that the CINS scores and ACORNS total KC scores are significantly correlated for the non-human taxa items ($\tau = 0.375$, $p < 0.01$; $r = 0.562$). Both CINS scores ($\tau = -0.252$, $p < 0.01$; $r = -0.383$) and ACORNS total KC scores ($\tau = -0.310$, $p < 0.01$; $r = -0.468$) are negatively associated with the ACORNS total NI scores. For the human items, the CINS scores had a very strong and significant association with the ACORNS total KC scores ($\tau = 0.411$, $p < 0.01$; $r = 0.600$) and both the ACORNS total KC scores ($\tau = -0.258$, $p < 0.01$; $r = -0.397$) and CINS scores ($\tau = -0.160$, $p < 0.01$; $r = -0.249$) were significantly negatively associated with the ACORNS total NI scores.

### Trait and taxon classifications

Plots of mean familiarity scores for traits revealed clear distribution differences (Fig. 1). We therefore categorized each trait as either familiar or unfamiliar. In contrast, the taxa were viewed as similarly familiar. Specifically, brain and eyelashes were given a score of 3 or 4 by nearly all of the biology and anthropology students (Fig. 1a). Conversely, dermatoglyphics and navicular were given a score of 1 or 2 by most

biology and anthropology students (Fig. 1a). All taxa were given a score of 4 by nearly all students (Fig. 1b). Therefore, for this population of students, the traits brain and eyelashes were classified as familiar and the traits dermatoglyphic and navicular were classified as unfamiliar. All taxa were classified as familiar but labeled as human or non-human in the models. Therefore, trait familiarity (familiar vs. unfamiliar) and taxon category (human vs. non-human) were the surface features examined in this study.

### RQ1 (How similar are the students enrolled in anthropology and biology classes?)

Anthropology and biology students display significantly different patterns for all demographic variables and most background variables. The anthropology population had fewer females ($\chi^2 = 12.69$, $df = 1$, $p < 0.001$), fewer white students ($\chi^2 = 23.78$, $df = 1$, $p < 0.001$), fewer students for whom English was a first language ($\chi^2 = 153.15$, $df = 1$, $p < 0.001$), more students early in their college career ($\chi^2 = 181.9$, $df = 3$, $p < 0.001$), and fewer previous and current courses ($\chi^2 = 1746.8$, $df = 3$, $p < 0.001$). Word count for the open response ACORNS items was not significantly different between biology and anthropology students. See Table 2 for means and standard errors.

### RQ2 (Do evolutionary knowledge and naive ideas differ across anthropology and biology students? If so, how?)

Anthropology and biology students also showed significant differences in the evolution knowledge and reasoning variables. The anthropology population had lower CINS scores ($b = -2.92$, $t = -15.92$, $p < 0.001$, $\eta^2 G = 0.12$), fewer KCs ($b = -0.91$, $Z = -4.5$, $p < 0.001$, $\eta^2 G = 0.04$), more NIs ($b = 1.22$, $Z = 5.58$, $p < 0.001$, $\eta^2 G = 0.06$), and a lower probability of a pure scientific model ($b = -1.62$, $Z = -6.73$, $p < 0.001$, $\eta^2 G = 0.09$) (Fig. 2a–d). See Table 2 for means and standard errors. As indicated by $\eta^2 G$, the size of the effect of student classification on evolutionary knowledge was small for KCs, and medium for NIs, MT, and CINS.

### RQ3 (Is variation in evolutionary knowledge and naive ideas across these populations explained by background and demographic variables?)

The difference in the per-item total KCs between anthropology and biology students was explained by background and demographic variables. Specifically, when controlling for background and demographic variables, the per-item total KCs was no longer significantly different between the two populations. Rather, the number of prior and current biology and anthropology courses was the only significant predictor for the per-item total KCs ($b = 0.80$, $z = 2.68$, $p < 0.01$).
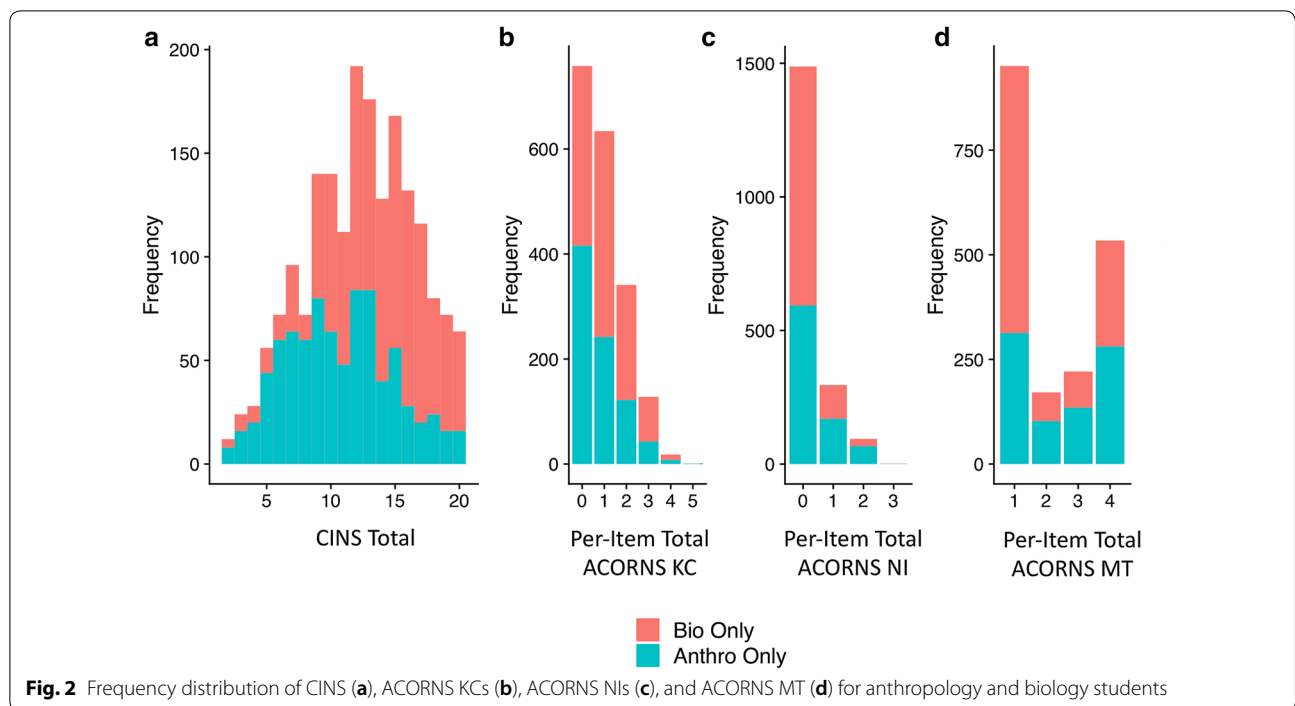
**Table 2 Demographic, background and evolutionary knowledge measures for biology and anthropology students**

|  | Biology | Anthropology |
|---|---|---|
| # of responses | 1040 | 832 |
| # of participants | 260 | 208 |
| Total # of prior or current courses | Mean = 2.18<br>SE = 0.015 | Mean = 1.04<br>SE = 0.007 |
| Year | Freshman: 31 (12%)<br>Sophomore: 116 (45%)<br>Junior: 63 (24%)<br>Senior: 49 (19%) | Freshman: 72 (35%)<br>Sophomore: 80 (39%)<br>Junior: 43 (21%)<br>Senior: 11 (5%) |
| Sex | Male: 119 (46%)<br>Female: 140 (54%) | Male: 112 (54%)<br>Female: 94 (46%) |
| Ethnicity | White only: 182 (72%)<br>Non-white: 71 (28%) | White only: 122 (61%)<br>Non-white: 78 (39%) |
| English 1st language | Yes: 235 (90%)<br>No: 25 (10%) | Yes: 140 (67%)<br>No: 68 (33%) |
| Word count | Mean = 29.88<br>SE = 0.76 | Mean = 28.56<br>SE = 1.01 |
| CINS | Mean = 13.6<br>SE = 0.12 | Mean = 10.68<br>SE = 0.14 |
| Per-item total ACORNS KC | Mean = 1.07<br>SE = 0.03 | Mean = 0.78<br>SE = 0.03 |
| Per-item total ACORNS NI | Mean = 0.18<br>SE = 0.01 | Mean = 0.37<br>SE = 0.02 |
| Per-item total ACORNS MT | Pure scientific: 631 responses<br>Other: 409 responses | Pure scientific: 313 responses<br>Other: 519 responses |

In contrast, the difference in the per-item total NIs, the probability of a pure scientific model, and CINS scores between anthropology and biology students were partially, but not fully explained by demographic and background variables as demonstrated by the sustained significance, but decreased effect size, of the student classification variable (Table 3). Specifically, when controlling for background and demographic variables, the anthropology population maintained significantly lower CINS scores ($b = -1.34$, $t = -4.16$, $p < 0.001$, $\eta^2 G = 0.009$), more NIs ($b = 1.73$, $z = 4.24$, $p < 0.001$, $\eta^2 G = 0.04$), and a lower probability of a pure scientific model ($b = -1.16$, $z = -2.71$, $p < 0.01$, $\eta^2 G = 0.01$) (Table 3). The classification of the student was the only significant predictor variable for the per-item total NIs and the probability of a pure scientific model. However, in addition to student classification being a significant predictor variable for CINS scores, course history ($b = 0.72$, $t = 3.00$, $p < 0.001$, $\eta^2 G = 0.005$) and English as a first language ($b = -2.40$, $t = -10.35$, $p < 0.001$, $\eta^2 G = 0.02$) also had significant unique effects on CINS scores. Of all the predictor variables, English as a first language had the largest effect size on CINS scores.

**RQ4.1–4.2 (To what extent do surface features impact each population's evolutionary knowledge and naive ideas? Specifically, do evolutionary knowledge and naive ideas differ based on: (RQ 4.1) the taxon (human vs. non-human)?, (RQ 4.2) the familiarity of the trait?)**

For biology students, trait familiarity and taxon category did not explain per-item total KCs, per-item total NIs, or the probability of a pure scientific MT (Fig. 3a–d). Therefore, biology students' open response answers were not impacted by these surface features. In contrast, for anthropology students, the trait familiarity ($b = 0.39$, $z = 2.59$, $p < 0.01$, $\eta^2 G = 0.009$) and taxon category ($b = -0.83$, $z = -2.82$, $p < 0.01$, $\eta^2 G = 0.03$) explained per-item total KCs, with the highest scores occuring for familiar traits and in a non-human context (Fig. 3a, c). Trait familiarity and taxon category did not impact per-item total NIs for these same anthropology students (Fig. 3b, d). Therefore, anthropology students' open response answers were impacted by these surface features for KCs but not NIs. Furthermore, taxon category ($b = -0.95$, $z = -2.839$, $p < 0.01$, $\eta^2 G = 0.03$), but not trait familiarity, significantly explained the probability of a pure scientific MT for anthropology students, with the highest probability occurring for the non-human items.

**Fig. 2** Frequency distribution of CINS (**a**), ACORNS KCs (**b**), ACORNS NIs (**c**), and ACORNS MT (**d**) for anthropology and biology students
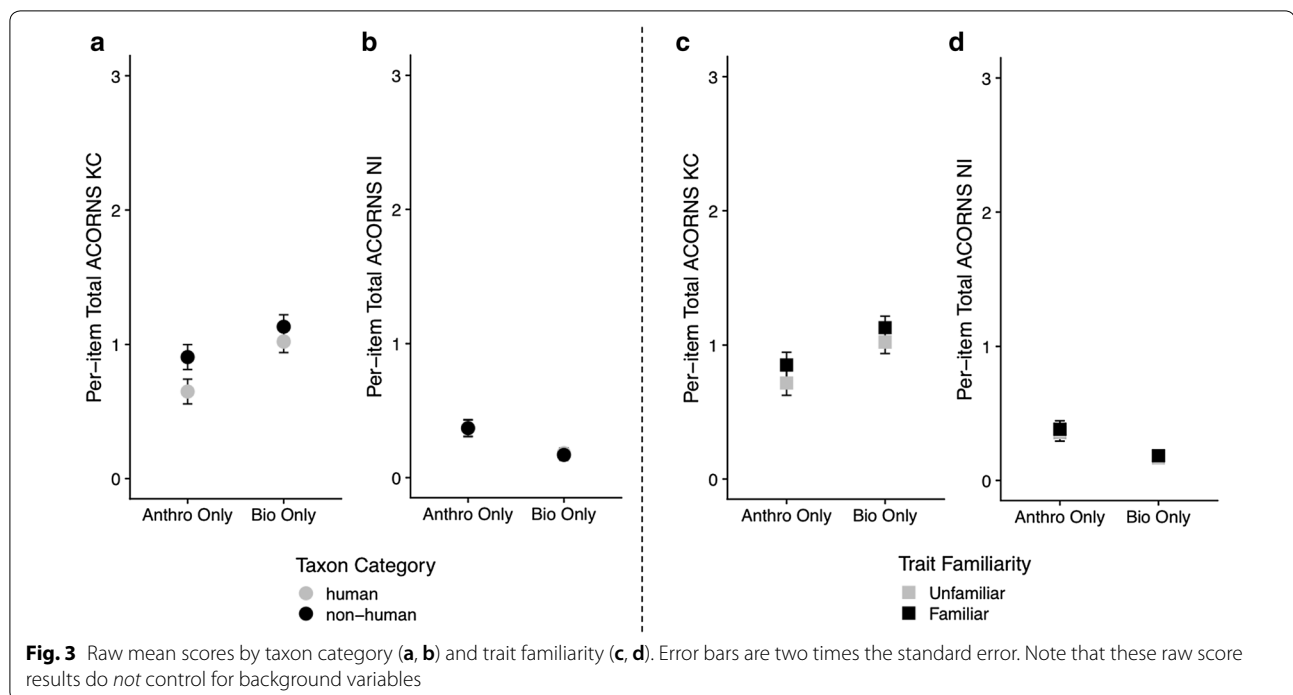
## Discussion

Undergraduate science education reform has focused attention on the teaching, learning, and assessment of core concepts, such as the disciplinary core idea of evolution (e.g., NRC 2001a, b, 2012a, b; AAAS 2011; NGSS Lead States 2013; Sinatra et al. 2008). A large body of research in evolution education has resulted from these initiatives. Much of this work has been directed at student understanding of evolution and non-normative ideas about evolution, sometimes with the intention of developing pedagogies to initiate conceptual change (e.g., Bishop and Anderson 1990; Demastes et al. 1995a, b, 1996; Nehm and Schonfeld 2007; Scharmann 1994; Nehm and Reilly 2007). These studies form a substantial literature regarding the magnitudes of evolutionary knowledge, non-normative ideas, and acceptance of biology students and teachers. Yet remarkably little is known about evolutionary knowledge and reasoning in another undergraduate population taught in a very different context: biological anthropology (e.g., Cunningham and Wescott 2009). Indeed, while biological anthropology and biology share a common 'language' of evolution (Wilson 2005), they offer distinct experiences when learning evolutionary theory. Anthropology offers a unique learning environment focusing on a single lineage and associated case studies of evolution occurring in that lineage. Do these different educational experiences produce significant differences in knowledge, misconceptions, and reasoning patterns?

**Table 3 Comparison of measures between populations with and without controlling for demographic and background variables**

|  | No control variables | With control variables |
|---|---|---|
| CINS | Anthro < Bio ($\eta G^2 = 0.12$) | Anthro < Bio ($\eta G^2 = 0.009$) |
| Per-item total KC | Anthro < Bio ($\eta G^2 = 0.04$) | Anthro = Bio |
| Per-item total NI | Anthro > Bio ($\eta G^2 = 0.06$) | Anthro > Bio ($\eta G^2 = 0.04$) |
| Per-item total MT | Anthro < Bio ($\eta G^2 = 0.09$) | Anthro < Bio ($\eta G^2 = 0.01$) |

The overarching goal of our work was to begin to explore evolutionary knowledge and reasoning patterns in this population and compare them to undergraduate biology students.

The courses from which our populations of students were sampled appeared to be comparable on paper. Both courses represent one of the two (biology) or three (anthropology) introductory level offerings for each program, the order of which are unimportant. Both require a laboratory component in addition to the lecture component. Despite these similarities and the fact that both anthropology courses and biology courses use evolutionary theory as their foundation, our findings show that the students who come from these backgrounds displayed demographic *and* knowledge differences. In fact, there

**Fig. 3** Raw mean scores by taxon category (**a**, **b**) and trait familiarity (**c**, **d**). Error bars are two times the standard error. Note that these raw score results do *not* control for background variables

were significant differences for all demographic and background variables tested. For example, the anthropology students in our sample were actually less experienced in terms of how many evolution-related courses they had already taken and therefore, had not progressed as far in their overall college coursework. Given this information, it is perhaps not surprising that the two populations displayed differences in their understanding of evolutionary concepts. Across all measures of knowledge and reasoning, anthropology students had worse scores than the biology students, despite their open-response answers being comparable in terms of verbosity (c.f., Federer et al. 2015). These differences in knowledge and misconceptions were largely (i.e., ACORNS KC) or partially explained (i.e., CINS, ACORNS NI, ACORNS MT) by controlling for demographic and background variables, but significant differences, with small effect sizes, remained. Specifically, when controlling for background and knowledge variables, anthropology and biology students no longer differed in the number of accurate ideas that they used in their evolutionary explanations. Nevertheless, as compared to biology students, anthropology students displayed lower CINS scores, were more likely to bring non-normative ideas into their evolutionary explanations, and remained further from expert-like reasoning.

Many different variables can be used to place a learner along a novice-expert continuum (e.g., Beggrow and Nehm 2012). In this study, we focused on three variables:

amount of knowledge, amount of misconceptions, and sensitivity to surface features in evolutionary reasoning. Experts are expected to have high knowledge, few misconceptions, and low sensitivity to surface features (Nehm and Ridgway 2011). It is possible for respondents to demonstrate novice-like behavior for some of these variables and expert-like behavior for others. Biology and anthropology students demonstrated novice-like levels of evolutionary knowledge. Specifically, both populations performed poorly on the CINS, a non-majors test of evolutionary knowledge (Anderson et al. 2002), with mean scores of 13.6 and 10.68, respectively. Furthermore, while both biology and anthropology students demonstrated few misconceptions in their explanations of evolutionary change (i.e., few NIs, 0.18 and 0.37, respectively), they also demonstrated low levels of knowledge (i.e., few KCs, 1.07 and 0.78, respectively) and inconsistent evolutionary models (i.e., low rates of purely scientific models, 61% and 38%, respectively).

Although both populations demonstrated novice-like knowledge and reasoning patterns, biology students performed significantly better for all of these variables than anthropology students. The difference was the most striking for evolutionary reasoning, where biology students had nearly twice the rate of normative evolutionary models as anthropology students. Therefore, for the purposes of this paper, we will classify biology students as novices and anthropology students as extreme novices. For anthropology students then, doing worse on these

three measures (CINS, ACORNS NI, and ACORNS MT) compared to the biology students, could be reflective of their relatively early stage of learning about evolution. As extreme novices are learning, non-normative ideas can often persist while new and normative scientific ideas are integrated into their knowledge frameworks (e.g., Vosniadou et al. 2008; Kelemen and Rosset 2009; Nehm 2010), resulting in a synthetic model of both normative and non-normative ideas (e.g.,. Beggrow and Nehm 2012; Nehm and Ha 2011; Vosniadou et al. 2008). Accordingly, when a task cues that synthetic model, all the knowledge (normative and non-normative) will be elicited together. This could explain why anthropology students had KCs similar to the biology students but, because they are still in the early stages of building their evolution knowledge frameworks, their misconceptions were elicited as well, thereby resulting in a majority of explanations that exhibited non-scientific reasoning models. Similarly, on the CINS multiple-choice test, it is likely that for anthropology students, enough misconceptions are being cued, such that the incorrect choices (designed to highlight typical non-normative ideas; Anderson et al. 2002) appear as viable options. Meanwhile, biology students, while they performed as novices overall, did have a slight majority of explanations scored as pure scientific models. On the novice-expert continuum, some of these explanations fit the "emerging expert" category (adaptive reasoning using key concepts only), which is is not completely unexpected given prior research findings with similar populations (Beggrow and Nehm 2012; Nehm and Ha 2011; Nehm and Schonfeld 2008).

Sensitivity to item surface features can also be used to place learners along a novice-expert continuum. The fact that item surface features affect student learning and problem solving has been well-documented (e.g., Caleon and Subramaniam 2010; Chi et al. 1981; diSessa et al. 2004; Evans et al. 2010; Gentner and Toupin 1986; Nehm and Ha 2011; Sabella and Redish 2007; Sawyer and Greeno 2009; Schmiemann et al. 2017). In evolutionary biology, changing various types of item surface features (e.g., animal vs. plant taxon; loss vs. gain of trait; familiar vs. unfamiliar taxon/trait) has been found to influence reasoning patterns of novices (Federer et al. 2015; Ha et al. 2006; Nehm et al. 2012; Nehm and Ha 2011; Nehm and Reilly 2007; Nehm and Ridgway 2011; Opfer et al. 2012), yet experts tend to see beneath these surface feature effects (e.g., Chi et al. 1981; Nehm and Ridgway 2011; Opfer et al. 2012). We used two types of surface features in this study—trait familiarity and taxon—and will discuss the results for each in turn.

## Surface feature 1
### Trait familiarity

Our study used items in which all taxa were standardized as familiar, but traits were presented that were both familiar or unfamiliar. Levels of familiarity were hypothesized a priori using Google™ PageRank (see Additional file 2: Appendix 2), but confirmed a posteriori using student familiarly ratings. To our knowledge, this is the first study to explore the effects of surface feature familiarity on evolutionary reasoning while keeping constant the familiarity of the taxon. This approach is essential to tease apart the role of familiarity with "who" evolves vs. "what" evolves. Therefore, this study is the only one we know of that allows the robust investigation of trait familiarity in evolutionary knowledge and reasoning patterns. We found that when we varied the familiarity of traits (i.e., what is evolving) in our items, but kept the taxon (i.e., who is evolving) familiar, biology and anthropology students demonstrated different reasoning patterns. Specifically, biology students' explanations were not sensitive to trait familiarity for all knowledge and reasoning outcome variables. The anthropology student explanations were similarly resistant to this surface feature in terms of their misconceptions and evolutionary reasoning, but did not exhibit comparable resistance in terms of the number of KCs used. Previous research investigating the impact of the familiarity of item surface features on student evolutionary reasoning using the ACORNS instrument has shown more pronounced effects. However, these studies differ from ours in that familiarity was standardized across *both* the taxon (i.e., who is evolving) and the trait (i.e., what is evolving) (e.g., Nehm and Ha 2011; Opfer et al. 2012). Therefore, it is possible that the specific surface feature (e.g., trait vs. taxon) and the number of surface features (e.g., trait/taxon only vs. taxon and trait) designated as unfamiliar may impact research findings. For example, Federer et al. (2015) found that students used *more* KCs and NIs in their explanations for items of familiar taxa/familiar traits compared to items of unfamiliar taxa/unfamiliar traits. We did not find this to be the case with either biology or anthropology students, instead we saw anthropology students using more KCs but no difference in their NIs. Another study also found students to use more KCs in their explanations for items of familiar taxa/familiar traits compared to items of unfamiliar taxa/unfamiliar traits, but no difference for cognitive biases (e.g., teleological misconceptions; Opfer et al. 2012). These results demonstrate a similar pattern to ours, but use slightly different measures of non-normative ideas. Again, it is important to note that both of these studies differed from ours in that the authors designed their items such that both traits *and* the taxa were familiar or unfamiliar. Therefore, even though we

did find some effects of familiarity on student knowledge and reasoning patterns, our results did not completely align with those from previous ACORNS research. It raises the question as to whether keeping one item feature familiar is sufficient to mitigate some potential effects unfamiliarity has on student reasoning. Indeed, outside of evolution, in an investigation of familiarity effects on genetics understanding, Schmiemann et al. (2017) compared measures across items that featured familiar or unfamiliar plants and animals with familiar traits and found no effects of their surface features on students' genetic reasoning. Similar to our study, only the familiarity of one surface feature was altered while the other remained familiar across items. However, while our study varied trait familiarity, their study varied taxon familiarity. Taking their findings into consideration with ours, the question of why it might matter *who* evolves, or *what* evolves, remains open. Additionally, while many studies have shown surface features are not expected to impact experts (e.g., Chi et al. 1981; Chi 2006; Nehm and Ha 2011; Nehm and Ridgeway 2011; Opfer et al. 2012), it is not known how the *familiarity* of surface features would affect experts. Because other surface features do not significantly impact experts, it is likely that experts would not be affected by the familiarity of the surface features we used here. Therefore, referring back to a novice-expert continuum, biology students demonstrate more expert-like reasoning (relative to anthropology students) in their low sensitivity to the varying familiarity of our item surface features used here, although to confirm this characterization, studies with experts are needed.

## Surface feature 2

### *Taxon*

While research into the effects of surface feature familiarity is minimal, there is even less work regarding whether the construct of *human* impacts students' evolutionary reasoning patterns. Using human examples in evolution education has been suggested to help to: motivate interest in the topic, form a bridge to less familiar contexts (i.e., non-human), and help students overcome misconceptions (e.g., Hillis 2007; Medin and Atran 2004; Nettle 2010; Paz-y-Miño and Espinosa 2009; Pobiner et al. 2018; Seoh et al. 2016; Wilson 2005). However, anthropology students learn evolutionary theory within a single context (primate lineages) and their knowledge might be more tightly bound to this context compared to that of biology students (diverse array of taxa) (Bjork and Richardson-Klavehn 1989). Thus, any differences we would expect to see in anthropology students' reasoning would be between human and non-human item measures; specifically, we would have expected the human context to elicit more key concepts (even if more naive ideas were

also provided). Indeed, our study did find feature effects of taxon category on anthropology students' knowledge measures and reasoning patterns, but not for the biology students. However, contrary to what was expected for anthropology students, non-human items had higher key concept scores and were significantly more likely to elicit a pure scientific MT, though the effect size was small. These results raise the question of why their knowledge patterns were not as they were predicted. The only other study, to our knowledge, that has looked at differences in evolutionary reasoning across human and non-human items did find similar results (Ha et al. 2006). Ha and colleagues used items asking about evolution in humans and non-humans to examine students' explanations across various ages for accurate scientific ideas and misconceptions. They found that when asked about human evolution, students' were less likely to use an accurate scientific explanation of evolution by natural selection. Furthermore, both human and animal items were more likely to elicit naive ideas regarding the use/disuse of traits as well as intentionality (Ha et al. 2006). While Ha et al. looked at these patterns in elementary through high school level students (who are not learning evolutionary theory situated within a human context), the similarity in their results align with our placement of anthropology students (who have received very little evolutionary instruction overall) on the extreme novice end of the continuum for evolutionary reasoning in regards to their sensitivity to taxon category. Our results generated little evidence in support of the claim that learning evolution within a human evolution context (i.e., primate lineage) is advantageous. Incorporating human examples may still be beneficial, but only when interspersed with examples of other taxonomic contexts. Our results raise numerous questions about what might be effective ways of integrating human examples into evolutionary instruction.

A number of studies suggest that the inclusion of human evolution into evolution instruction has the potential to improve learning; only two studies to our knowledge have directly investigated these effects. Evidence for positive impacts resulting from the inclusion of human evolution has been found for both human evolution instruction followed and human evolution assessment items (e.g., Nettle 2010; Pobiner et al. 2018). In a study with college-level psychology students, Nettle found that participants who were taught evolution in the context of humans performed better on questionnaires that invoked human evolution rather than evolution in non-human taxa, particularly regarding misunderstandings stemming from the lack of attention to intraspecific variation (other non-normative ideas also persisted). Weaknesses of Nettle's (2010) study worth noting include a limited focus on assessing students on human

vs. non-human evolution (as opposed to investigating impacts of human context on learning evolution) and he neglected to establish evidence for validity and reliability for the instrument. In contrast, Pobiner et al. (2018) developed human evolution curriculum mini-units for high school biology students and measured evolutionary knowledge both pre- and post-instruction using instruments for which validity and reliability evidence has been gathered (e.g., ACORNS). They found that students displayed a gain in knowledge measures post-instruction, though their analysis was limited to three key concepts (Pobiner et al. 2018). Even though this finding aligns with our results (anthropology students did not differ from biology students in their ACORNS key concept scores), their analyses did not include naive ideas nor did it compare their human evolution curriculum with non-human evolution curriculum (Pobiner et al. 2018). Thus, their findings are limited and, beyond student interest or motivation, do not provide strong evidence for an advantage of human evolution instruction (Pobiner et al. 2018). Given the paucity of empirical research on human evolution instruction, it is entirely possible that the human context itself provides no such advantages described above for learning and applying evolutionary concepts and the advantages seen are rather from increasing the diversity of contexts of evolutionary content, in general.

The NRC (2001a, b) emphasizes that an integrative mental framework utilized across a range of contexts is essential for achieving competency in science. If biology students are better at applying the evolutionary ideas that they have learned across situational features (i.e., non-human and human evolutionary change), it raises the question as to what it is about biology, which anthropology lacks, that fosters this more flexible conceptual framework. Theory suggests that this lack of flexibility could be a by-product of the focused nature of evolutionary theory learners experience in anthropology (e.g., Jacobson and Spiro 1995; Spiro et al. 1989). By only representing evolutionary theory using a single theme (e.g., evolution in the primate lineage), the construct of evolution becomes oversimplified, the likelihood of embedded misconceptions increases, and the likelihood of achieving flexible, transferable knowledge frameworks decreases (Jacobson and Spiro 1995). Incorporating a variety of examples across a diversity of contexts has been suggested as a more optimal method for teaching (Anderson et al. 1996; Jacobson and Spiro 1995; Nehm, 2018; Opfer et al. 2012; Spiro et al. 1989). Accordingly, the biology students demonstrate some ability to consistently apply their evolutionary knowledge across such a range - a skill the anthropology students do not seem to have mastered yet.

Ultimately, biology students' explanatory frameworks appear to be relatively more developed and coherent than those of the anthropology students as they exhibit consistency in application across taxon categories and across trait familiarity (Kampourakis and Zogza 2009; Nehm 2018). Considering that experts are better at seeing beneath surface features (e.g., Chi 2006), and that transfer is a factor of representation and degree of practice (Anderson et al. 1996), it seems an advantage for learning evolutionary concepts and fostering more advanced conceptual frameworks lies in teaching a construct, like evolution, across a diversity of contexts.

While we did control for many demographic and background variables, an alternative explanation could be that some other differences in biology and anthropology students that we did not control for accounted for the sensitivity to taxon that the anthropology students displayed. Their sensitivity to the human taxon could be a result of their limited exposure to anthropology (the majority of the students' only anthropology course was the one they were currently enrolled in). Future studies including anthropology students with more experience in terms of coursework could help resolve this issue.

### Implications for instruction
The finding that naïve ideas were more common in anthropology students compared to biology students (when demographic and background features were held constant) suggests that targeting naive ideas should be an instructional goal for anthropology education. Additionally, considering the positive effects associated with incorporating human examples into *biology* instruction found by other authors (e.g., deSilva 2004; Flammer 2006; Nettle 2010; Price 2012; Pobiner et al. 2018; Seoh et al. 2016), another potential instructional goal could be incorporating non-human comparative examples into *anthropology* instruction. Providing a greater diversity of contexts for anthropology students could help build a greater flexibility into their conceptual frameworks and foster more expert-like reasoning. Clearly, more studies including anthropology students, instructors, and experts are called for, as they will continue to help clarify how contextual factors impact the learning of evolution.

### Limitations
One major limitation is that biology and anthropology students may be different populations as evidenced by their significantly different patterns of demographic and background variables. One of the most striking differences is that the vast majority of anthropology students have taken only one anthropology class (i.e., the one they were in while completing the survey). In contrast, most biology students had already taken biology classes in

addition to the one they were in during the survey. Therefore, although both populations were sampled at a similar time in their academic careers, these findings demonstrate that care must be taken to ensure that comparisons between anthropology and biology students are appropriate. However, even when controlling for the number of prior courses, significant differences between the two populations were still found using regression analyses. A potentially more appropriate method for comparing these two populations could be propensity score matching using a larger data set. Additionally, recruiting students from higher level courses could potentially help mitigate these concerns.

As described above, anthropology and biology students may differ in evolutionary knowledge and reasoning patterns due to their respective training. However, it is also possible that the populations enrolling in each of these courses are different in the first place, and thus, the outcomes may not be indicative of the impact of their respective types of evolutionary training. We controlled for many of the differences among students in the analyses, but we were not able to control for every student variable. For example, is possible that motivation and interest may differ among the biology and anthropology students in the sampled population. Specifically, the introductory biology course in which this study took place was designed for biology majors and most of the students in the class were biology majors. There are alternative introductory level biology courses at the university for non-major students. In contrast, the introductory anthropology class used in this study is taken by both majors and non-majors, and there are no other introductory course offerings for non-majors. The different introductory course structures for these two disciplines may have contributed to the discrepancy in previous coursework observed between our two populations, and may differentially impact student motivation and/or interest. In terms of the former limitation, sampling from upper level courses for comparison or, alternatively, sampling introductory anthropology along with a *non-major* introductory biology course could lead to more comparable populations. In addition, gathering pre-test data on the populations could also help with this limitation. In terms of the latter limitation, the interaction between context and motivation/interest was beyond the scope of this study, but raises important questions that could be addressed in future work.

Although we were able to determine that there are differences between populations of biology and anthropology students, we are unable to tease apart the program these students are situated within and the instructional variation the students are experiencing. In other words,

is it the nature of the content (evolution via biology vs. evolution via anthropology) or characteristics of the instructors in these programs? Accordingly, an alternative explanation for the differences in measures of knowledge and reasoning seen between the populations is the anthropology students' lack of familiarity with the assessment format. The biology program involved in this research is strongly rooted in biology education research, conducts its own research studies and incorporates evidence-based teaching practices. Thus, the ACORNS item format used in this study, while novel to the anthropology students, is not novel to the biology students. While it is possible that this discrepancy in assessment format familiarity could have impacted the anthropology students performance (Norman et al. 1996; Opfer et al. 2012; Schmiemann et al. 2017), it seems unlikely considering there was no difference in KC measures between populations. However, the instruction itself could be impacting the results if research on novices' non-normative ideas is being addressed through targeted instruction. These ambiguities could be addressed with future research including larger samples of students across programs with diverse involvement in biology education research.

## Additional files

**Additional file 1: Appendix 1.** Table S1. Biological terms rated for familiarity and their expected familiarity based on Google™ PageRank.

**Additional file 2: Appendix 2.** Table S2. Learning objectives for courses sampled as presented in syllabi.

**Author details**
[1] Center for Life Sciences Education, The Ohio State University, 1735 Neil Avenue, Columbus, OH 43210, USA. [2] Department of Ecology & Evolution, Stony Brook University, 650 Life Sciences Building, Stony Brook, NY 11794, USA.

**References**
Adams F, Aizawa K. Why the mind is still in the head. In: Robbins P, Aydede ME,
    editors. The Cambridge handbook of situated cognition. Cambridge:
    Cambridge University Press; 2009. p. 78–95.
Alles DL, Stevenson JC. Teaching human evolution. Am Biol Teach.
    2003;65:333–9.
American Association for the Advancement of Science. Vision and change in
    undergraduate biology education. Washington, DC. 2011. http://visio
    nandchange.org. Accessed 1 Nov 2018.
Anderson JR, Reder LM, Simon HA. Situated learning and education. Educ
    Researcher. 1996;25:5–11.
Anderson DL, Fisher KM, Norman GJ. Development and evaluation of the
    conceptual inventory of natural selection. JRST. 2002;39:952–78.
Ashmore PC. Role of physical anthropology in intermediate and secondary
    education. Am J Phys Anthropol. 2005;48:154–62.
Atran S. Folk biology and the anthropology of science: cognitive universals
    and cultural particulars. Behav Brain Sci. 1998;21:547–609.
Atran S, Medin D, Lynch E, Vapnarsky V, Ucan Ek'E, Sousa P. Folkbiology doesn't
    come from folkpsychology: evidence from Yukatek Maya in cross-
    cultural perspective. J Cogn Cult. 2001;1:3–42.
Axelrod R, Hamilton WD. The evolution of cooperation. Science.
    1981;211:1390–6.
Bakeman R. Recommended effect size statistics for repeated measures
    designs. Behav Res Methods 2005;37:379–84. https://doi.org/10.3758/
    BF03192707
Barret L, Henzi SP. Monkeys, markets and minds: biological markets and
    primate sociality. In: Kappelar P, van Schaik CP, editors. Cooperation in
    primates and humans: mechanisms and evolution. Berlin: Springer;
    2005a. p. 209–32.
Barret L, Henzi SP. The social nature of primate cognition. P R Soc B.
    2005b;272:1865–75.
Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, Dai B,
    Scheipl F, Grothendieck G, Green P. Linear mixed-effects models using
    'Eigen' and S4. R package. 2018; v. 1.1-18-1.
Battisti BT, Hanegan N, Sudweeks R, Cates R. Using item response theory to
    conduct a distractor analysis on Conceptual Inventory of Natural Selec-
    tion. Int J Sci Math Educ. 2010;8:845–68.
Beggrow EP, Nehm RH. Students' mental models of evolutionary causation:
    natural selection and genetic drift. EEO. 2012;5:429–44.
Beggrow EP, Ha M, Nehm RH, Pearl DK, Boone WJ. Assessing scientific practices
    using machine-learning methods: how closely do they match clinical
    interview performance? J Sci Educ Technol. 2014;23:160–82.
Birnbaum D, Deeb I, Segall G, Ben-Eliyahu A, Diesendruck G. The development
    of social essentialism: the case of Isreali children's inferences about Jews
    and Arabs. Child Dev. 2010;81:757–77.
Bishop BA, Anderson CW. Student conceptions of natural selection. JRST.
    1990;27:415–27.
Bjork RA, Richardson-Klavehn A. On the puzzling relationship between
    environment context and human memory. In: Izawa C, editor. Current
    issues in cognitive processes: The Tulane Flowerree. Symposium on
    Cognition. Hillsdale: Erlbaum; 1989. p. 313–44.
Caleon IS, Subramaniam R. Do students know what they know and what they
    don't know? Using a four-tier diagnostic test to assess the nature of
    students' alternative conceptions. Res Sci Educ. 2010;40:313–37.
Carey S. Cognitive science and science education. Am Psychol.
    1986;41:1123–30.
Chi MT. Two approaches to the study of experts' characteristics. In: Charness N,
    Feltovich P, Hoffman RE, editors. The Cambridge handbook of expertise

and expert performance. Cambridge: Cambridge University Press; 2006.
    p. 21–30.
Chi MT, Feltovich PJ, Glaser R. Categorization and representation of physics
    problems by experts and novices. Cogn Sci. 1981;5:121–52.
Christensen RHB. Ordinal—regression models for ordinal data. R package.
    2018;v. 2018.8-25.
Clancey WJ. Scientific antecedents of situated cognition. In: Robbins P, Aydede
    ME, editors. The Cambridge Handbook of situated cognition. Cam-
    bridge: Cambridge University Press; 2009. p. 11–34.
Clough EE, Driver R. A study of consistency in the use of students' conceptual
    frameworks across different task contexts. Sci Educ. 1986;70:473–96.
Cunningham DL, Wescott DJ. Still more "fancy" and "myth" than "fact" in stu-
    dents' conceptions of evolution. Evol Educ Outreach. 2009;2:505. https
    ://doi.org/10.1007/s12052-009-0123-6.
Dagher ZR, BouJaoude S. Scientific views and religious beliefs of college
    students: the case of biological evolution. JRST. 1997;34:429–45.
de Waal FB. Good natured: origins of right and wrong in humans and other
    animals. Cambridge: Harvard University Press; 1997a.
de Waal FB. The chimpanzees' service economy: food for grooming. Evol Hum
    Behav. 1997b;18:375–86.
Demastes SS, Good RG, Peebles P. Students' conceptual ecologies and the
    process of conceptual change in evolution. Sci Educ. 1995a;79:637–66.
Demastes SS, Settlage J, Good RG. Students' conceptions of natural selection
    and its role in evolution: cases of replication and comparison. JRST.
    1995b;32:535–50.
Demastes SS, Good RG, Peebles P. Patterns of conceptual change in evolution.
    JRST. 1996;33:407–31.
DeSilva J. Interpreting evidence: an approach to teaching human evolution in
    the classroom. Am Biol Teach. 2004;66:257–67.
DiSessa A, Gillespie N, Esterly J. Coherence versus fragmentation in the devel-
    opment of the concept of force. Cogn Sci. 2004;28:843–900.
Dunbar RI. Coevolution of neocortical size, group-size and language in
    humans. Behav Brain Sci. 1993;16:681–94.
Dunbar RI. The social brain hypothesis. Evol Anthropol. 1998;6:178–90.
Evans EM, Spiegel AN, Gram W, Frazier BN, Tare M, Thompson S, Diamond S. A
    conceptual guide to natural history museum visitors' understanding of
    evolution. JRST. 2010;47:326–53.
Federer MR, Nehm RH, Opfer JE, Pearl D. Using a constructed-response instru-
    ment to explore the effects of item position and item features on the
    assessment of students' written scientific explanations. Res Sci Educ.
    2015;45:527–53.
Flammer L. The evolution solution: teaching evolution without conflict. Am
    Biol Teach. 2006;68:e1–7.
Fuentes A. Biological anthropology: concepts and connections. 2nd ed. New
    York: McGraw-Hill Education; 2011.
Futuyma D. Evolution. 2nd ed. Sunderland: Sinauer Associates; 2009.
Gelman SA, Legare CH. Concepts and folk theories. Annu Rev Anthropol.
    2011;4:379–98.
Gelman SA, Wellman HM. Insides and essences: early understandings of the
    non-obvious. Cognition. 1991;38:213–44.
Gentner D, Toupin C. Systematicity and surface similarity in the development
    of analogy. Cogn Sci. 1986;10:277–300.
Gilpin AR. Table for conversion of Kendall's tau to spearman's rho within the
    context of measures of magnitude of effect for meta-analysis. Educ
    Psychol Meas. 1993;53:87–92.
Gregory T. Understanding natural selection: essential concepts and common
    misconceptions. EEO. 2009;2:156–75.
Ha M, Lee JK, Cha HY. A cross-sectional study of students' conceptions on
    evolution and characteristics of conception formation about it in terms
    of the subjects: human, animals and plants. JKAIE. 2006;26:813–25.
Hammerstein P. Genetic and cultural evolution of cooperation. Cambridge:
    MIT Press; 2003.
Hillis DM. Making evolution relevant and exciting to biology students. Evolu-
    tion. 2007;61:1261–4.
Humphrey NK. Species and individuals in the perceptual world of monkeys.
    Perception. 1974;3:105–14.
Inagaki K, Hatano G. Young children's recognition of commonalities between
    animals and plants. Child Dev. 1996;67:2823–40.
Inagaki K, Hatano G. Young children's naive thinking about the biological
    world. New York: Psychology Press; 2002.

Inagaki K, Hatano G. Young children's conception of the biological world. Curr Dir Psychol Sci. 2006;15:177–81.

Jacobson MJ, Spiro RJ. Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: an empirical investigation. J Educ Comp Res. 1995;12:301–33.

Kampourakis K, Zogza V. Preliminary evolutionary explanations: a basic framework for conceptual change and explanatory coherence in evolution. Sci Educ. 2009;18:1313–40.

Kelemen D, Rosset E. The human function compunction teleological explanation in adults. Cognition. 2009;111:138–43.

Kirsh D. Problem solving and situated cognition. In: Robbins P, Aydede ME, editors. The Cambridge handbook of situated cognition. Cambridge: Cambridge University Press; 2009. p. 264–306.

Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol. 2013;2013(4):1–12.

Medin DL, Atran S. The native mind: biological categorization reasoning in development and across cultures. Psychol Rev. 2004;111:960–83.

Moharreri K, Ha M, Nehm RH. EvoGrader: an online formative assessment tool for automatically evaluating written evolutionary explanations. EEO. 2014;7:1–15.

National Research Council. Investigating the influence of standards: a framework for research in mathematics, science, and technology education. Washington, DC: National Academy Press; 2001a.

National Research Council. Knowing what students know. Washington, DC: National Academy Press; 2001b.

National Research Council. Thinking evolutionarily: evolution education across the life sciences. Washington, DC: National Academy Press; 2012a.

National Research Council. A framework for K-12 science education: practices, crosscutting concepts, and core ideas. Washington, DC: The National Academies Press; 2012b.

Nehm RH. Chapter 14: evolution. In: Reiss M, Kampourakis K, editors. Teaching biology in schools. London: Routledge; 2018.

Nehm RH, Ha M. Item feature effects in evolution assessment. JRST. 2011;48:237–56.

Nehm RH, Reilly L. Biology majors' knowledge and misconceptions of natural selection. Bioscience. 2007;57:263–72.

Nehm RH, Ridgway J. What do experts and novices "see" in evolutionary problems? EEO. 2011;4:666–79.

Nehm RH, Schonfeld IS. Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching of evolution in schools? J Sci Teach Educ. 2007;18:699–723.

Nehm RH, Schonfeld IS. The future of natural selection knowledge measurement: a reply to Anderson et al. (2010). JRST. 2010;47:358–62.

Nehm RH, Ha M, Rector M, Opfer J, Perrin L, Ridgway J, Mollohan K. Scoring guide for the open response instrument (ORI) and evolutionary gain and loss test (EGALT). Technical Report of National Science Foundation REESE Project 0909999; 2010.

Nehm RH, Beggrow EP, Opfer JE, Ha M. Reasoning about natural selection: diagnosing contextual competency using the ACORNS instrument. Am Biol Teach. 2012;74:92–8.

Nehm RH, Schonfeld IS. Measuring knowledge of natural selection: a comparison of the CINS, an open-response instrument, and an oral interview. J Res Sci Teach. 2008;45(10):1131–60.

Nettle D. Understanding of evolution may be improved by thinking about people. Evol Psychol. 2010;8:205–28.

NGSS Lead States. Next generation science standards: for states, by states. Washington, DC: The National Academies Press; 2013.

Nickels MK, Nelson CE, Beard J. Better biology teaching by emphasizing evolution and the nature of science. Am Biol Teach. 1996;58:332–6.

Norman GR, Swanson DB, Case SM. Conceptual and methodological issues in studies comparing assessment formats. Teach Learn Med. 1996;8:208–16.

Olejnik S, Algina J. Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychol Methods. 2003;8:434–47.

Opfer JE, Siegler RS. Revisiting preschoolers' living things concept: a microgenetic analysis of conceptual change in basic biology. Cogn Psychol. 2004;49:301–32.

Opfer JE, Nehm RH, Ha M. Cognitive foundations for science assessment design: knowing what students know about evolution. JRST. 2012;49:744–77.

Paz-y-Miño GC, Espinosa A. Assessment of biology majors' versus nonmajors' views on evolution, creationism, and intelligent design. EEO. 2009;2:75–83.

Pobiner BL. Use human examples to teach evolution. Am Biol Teach. 2012;74:71–2.

Pobiner BL. Accepting, understanding, teaching and learning (human) evolution: obstacles and opportunities. Am J Phys Anthropol. 2016;159:S232–74.

Pobiner BL, Beardsley PM, Bertka CM, Watson WA. Using human case studies to teach evolution in high school A.P. biology classrooms. EEO. 2018;11:3.

Population Division. World Urbanization Prospects: the 2018 Revision. United Nations, Department of Economic and Social Affairs. New York: United Nations Secretariat; 2018. https://population.un.org/wup/Publications/Files/WUP2018-KeyFacts.pdf. Accessed 1 Nov 2018.

Price RM. How we got here: evolutionary changes in skull shape in humans and their ancestors. Am Biol Teach. 2012;74(2):106–10.

Rhodes M, Gelman SA. A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. Cogn Psychol. 2009;59:244–74.

Sabella MS, Redish EF. Knowledge organization and activation in physics problem solving. Am J Phys. 2007;75:1017–29.

Sawyer RK, Greeno JG. Situativity and learning. In: Robbins P, Aydede ME, editors. The Cambridge handbook of situated cognition. Cambridge: Cambridge University Press; 2009. p. 347–67.

Scharmann LC. Teaching evolution: designing successful instruction. J Sci Teach Educ. 1994;5:122–9.

Schmiemann P, Nehm RH, Tornabene RE. Assessment of genetics understanding: under what conditions do situational features have an impact on measures? Sci Educ. 2017;266:1161–91.

Schrein CM. Evolution acceptance among undergraduates in the south. In: Lynn C, Glaze A, Evans W, Reed L, editors. Evolution education in the American south. New York: Palgrave Macmillan; 2017. p. 121–33.

Seely Brown J, Collins A, Duguid P. Situated cognition and the culture of learning. Educ Res. 1989;18:32–42.

Seoh KHR, Subramaniam R, Hoh YK. How humans evolved according to grade 12 students in Singapore. JRST. 2016;53:291–323.

Settlage J. Conceptions of natural selection: a snapshot of the sense-making process. JRST. 1994;31:449–57.

Shtulman A. Qualitative differences between naïve and scientific theories of evolution. Cogn Psychol. 2006;52:170–94.

Shtulman A, Schulz L. The relation between essentialist beliefs and evolutionary reasoning. Cogn Sci. 2008;32:1049–62.

Sinatra GM, Brem SK, Evans EM. Changing minds? Implications of conceptual change for teaching and learning about biological evolution. EEO. 2008;1:189–95.

Singmann H, Bolker B, Westfall J, Aust F, Højsgaard S, Fox F, et al. Analysis of factorial experiments (afex). 2018.v. 0.21-2.

Smith MU. Current status of research in teaching and learning evolution: II. Pedagogical issues. Sci Educ. 2010;19:539–71.

Spiro RJ, Feltovich PJ, Coulson RL, Anderson DK. Multiple analogies for complex concepts: antidotes for analogy-induced misconception in advanced knowledge acquisition. In: Vosniadou S, Ortony A, editors. Similarity and analogical reasoning. Cambridge: Cambridge University Press; 1989. p. 498–531.

Vosniadou S, Vamvakoussi X, Skopeliti I. The framework theory approach to the problem of conceptual change. In: Vosniadou S, editor. International handbook of research on conceptual change. New York: Routledge; 2008. p. 3–34.

Wellman HM, Gelman SA. Cognitive developments: foundational theories of core domains. Annu Rev Psychol. 1992;43:337–75.

Wilson DS. Evolution for everyone: how to increase acceptance of, interest in, and knowledge about evolution. PLoS Biol. 2005;3:2058–65.

Wilson RA, Clark A. How to situate cognition: letting nature take its course. In: Robbins P, Aydede ME, editors. The Cambridge handbook of situated cognition. Cambridge: Cambridge University Press; 2009. p. 55–77.