CURRICULUM AND EDUCATION   **Open Access**

Development and pilot testing of a three-dimensional, phenomenon-based unit that integrates evolution and heredity

Sheila A. Homburger[1], Dina Drits-Esser[1]*, Molly Malone[1], Kevin Pompei[1], Kagan Breitenbach[1], Ryan D. Perkins[1], Pete C. Anderson[1], Nicola C. Barber[1], Amy J. Hawkins[1], Sam Katz[1], Max Kelly[1], Harmony Starr[1], Kristin M. Bass[2], Jo Ellen Roseman[3], Joseph Hardcastle[3], George DeBoer[3] and Louisa A. Stark[1]

Abstract

To realize the promise of the Next Generation Science Standards, educators require new three-dimensional, phenomenon-based curriculum materials. We describe and report on pilot test results from such a resource—*Evolution: DNA and the Unity of Life*. Designed for the Next Generation Science Standards, this freely available unit was developed for introductory high school biology students. It builds coherent understanding of evolution over the course of seven to 8 weeks. Based around multiple phenomena, it includes core ideas about evolution, as well as pertinent core ideas from heredity. The unit integrates relevant crosscutting concepts as well as practice in analyzing and interpreting skill-level-appropriate data from published research, and constructing evidence-based arguments. We report results from a national pilot test involving 944 grade nine or ten students in 16 teachers' classrooms. Results show statistically significant gains with large effect sizes from pretest to posttest in students' conceptual understanding of evolution and genetics. Students also gained skill in identifying claims, evidence, and reasoning in scientific arguments.

Keywords: Evolution, NGSS, Curriculum, Secondary, Teaching, Molecular genetics, Argumentation, Data analysis

Introduction

The *Framework for K-12 Science Education* (National Research Council 2012) and the *Next Generation Science Standards* (NGSS) (NGSS Lead States 2013) derived from the *Framework* delineate a vision for K-12 science education that integrates disciplinary core ideas, science practices, and crosscutting concepts. Our project team has responded to the *Framework's* call for new curriculum materials and assessments on evolution that integrate these three dimensions. The materials are freely available and easily accessible online at http://teach.genetics.utah.edu/content/evolution/.

Evolution is fundamental to understanding biology (Dobzhansky 1973; National Research Council 2012), and it is widely accepted as a unifying, cross-disciplinary concept in science (Gould 2002). According to Glaze

and Goldston (2015), "For a person to be truly scientifically literate and able to make logical choices based on an understanding of scientific concepts, they must understand and be able to apply the concepts of evolution directly and indirectly to problems. Evolution is in essence the defining feature of living things that differentiates us from the nonliving matter of the universe" (p. 501). The *NGSS* similarly consider evolution to be foundational in biology and incorporate aspects of evolution across grade levels (Krajcik et al. 2014; NGSS Lead States 2013).

Yet elementary through postsecondary students, and the general public, have a poor grasp of this essential science idea (reviewed in Gregory 2009). Research has documented that evolution is difficult to teach and learn (Borgerding et al. 2015). A national assessment of students' ideas about evolution and natural selection found that misconceptions related to common ancestry were among the most prevalent (Flanagan and Roseman 2011). Barnes et al. (2017) found that cognitive biases

---

*Correspondence: dina.drits@utah.edu
[1] Genetic Science Learning Center, University of Utah, Salt Lake City, USA
Full list of author information is available at the end of the article

Homburger *et al. Evo Edu Outreach*   (2019) 12:13

Page 2 of 17

significantly interfere with student learning of concepts in evolution. Specifically, teleological reasoning impairs student understanding of natural selection. Students have a poor understanding of evolutionary time (Catley et al. 2010), and they misinterpret evolutionary trees (Meir et al. 2007). They also have difficulty applying their knowledge of evolution to everyday issues (Catley et al. 2004). The most common student-held alternative conceptions about natural selection are rooted in misunderstandings about heredity (Bishop and Anderson 1990; Kalinowski et al. 2010; Nehm and Schonfeld 2008). The genetic mechanisms of mutation and random variation—key to understanding evolution—are particularly difficult for students to grasp (Morabito et al. 2010). Therefore, researchers have called for a stronger genetics component in students' study of evolution (Catley et al. 2010; Dougherty 2009).

Research (two studies with high school and one with undergraduate students) on curricula that integrate genetics and heredity suggests that this approach reduces students' alternative conceptions about evolution (Banet and Ayuso 2003; Geraedts and Boersma 2006; Kalinowski et al. 2010). Other research has shown that teaching genetics before evolution significantly increased high school students' evolution understanding compared to when genetics was taught after evolution (Mead et al. 2017). This difference was especially evident in lower-achieving students, where evolution understanding improved only when genetics was taught first. Some literature has described practitioners integrating these topics in their classroom (e.g., Brewer and Gardner 2013; Heil et al. 2013). Yet few widely available curriculum materials foster this integration, preventing students from easily making conceptual connections (e.g., Biggs et al. 2009; Miller and Levine 2008; Hopson and Postlethwait 2009).

Researchers have advocated for evolution instruction that not only integrates genetics, but also includes science practices, such as analyzing and interpreting data (Catley et al. 2004; Beardsley et al. 2011; Bray et al. 2009) and arguing from evidence, to foster student learning. Several studies have shown that students' content understanding increases when argumentation is an explicit part of instruction (Asterhan and Schwarz 2007; Bell and Linn 2000; Zohar and Nemet 2001).

Finally, researchers in science education have called for embedded formative assessments in curriculum materials (Achieve, Inc. 2016). Teachers can use these assessments to uncover student thinking and inform further instruction (Ayala et al. 2008; Furtak et al. 2016). The well-documented benefits of formative assessments in supporting student learning (e.g., Kingston and Nash 2011) include narrowing achievement gaps between high and low performing students

(Black and Wiliam 1998). Performance-based formative assessment tasks provide opportunities for students to explain their thinking though written activities (Kang et al. 2014). They can take many forms, including constructed response (Ayala et al. 2008) and multiple choice with written justification (Furtak 2009), among others.

Research has shown that high-quality curricular interventions play an important role in student learning. In a review of 213 studies on evolution teaching and learning, researchers found that curricula that provide students (and teachers) with appropriate conceptual connections and opportunities to use science practices positively impact student understanding (Glaze and Goldston 2015).

In response to the calls for new curricula that integrate the three major dimensions of NGSS, and for materials that address widespread misunderstandings related to biological evolution, the project team has developed and pilot tested an evolution curriculum unit for introductory high school biology. The unit fosters coherent student understanding of evolution through the integration of pertinent heredity core ideas, relevant crosscutting concepts, opportunities to analyze and interpret skill-level-appropriate data from published scientific research, and opportunities to construct evidence-based arguments. Further, the unit uses high-quality multimedia pieces to bring molecular-scale process and other difficult-to-understand concepts to life. Key molecules, such as DNA, mRNA, and proteins, are illustrated in a similar visual style across the module's materials. This consistent visual language adds a level of cohesion, helping students make conceptual connections across topics.

This article describes the *Evolution: DNA and the Unity of Life* unit (Genetic Science Learning Center 2018a, b), and outlines the unit's development and national pilot testing processes. The curriculum pilot test corresponds to the Design and Development phase of educational research (IES and NSF 2013) requiring a theory of action, articulation of design iterations, and initial evidence of effectiveness (i.e., To what extent does the new unit show promise for increasing student achievement?). The primary goals of the pilot test were to

1. Evaluate and improve the usability of the materials for teachers and students;
2. Gauge teachers' perceptions of the educational value of this unit compared to the evolution curriculum materials they have used in the past; and
3. Gather initial evidence of student learning gains from the unit.

Homburger *et al. Evo Edu Outreach*   (2019) 12:13

Page 3 of 17

This work sets the stage for further field testing of the unit using a randomized controlled trial, which is beyond the scope of this paper.

The pilot testing process, including iterative revisions and re-testing, is an essential component of our curriculum development process. The feedback from each goal informed curriculum revisions, most of which we re-tested with a different group of students and teachers in the second half of the school year. Here, we describe the curriculum experiences of 20 pilot teachers (16 of whom completed all research requirements), and present assessment results from 944 students.

Evolution: DNA and the Unity of Life curriculum unit
Unit overview

Evolution: DNA and the Unity of Life is a 7- to 8-week, comprehensive curriculum unit. Available for free, the unit's paper-based and interactive multimedia lessons were designed for the NGSS. Namely, they engage students in high-interest phenomena and provide opportunities for students to ask scientific questions, use models, analyze skill-level-appropriate data from published scientific studies, and construct evidence-based arguments. The unit incorporates the crosscutting concepts of patterns, systems and system models, and cause and effect.

Lessons are organized into five modules, each structured around a guiding question and age-appropriate phenomena. Table 1 outlines this structure, as well as the components of the NGSS featured in each module.

The disciplinary core ideas (DCIs) listed there are the ones whose components are most strongly featured. In some cases, to better integrate heredity and evolution concepts and to accommodate the featured phenomena, we unpacked the components of each DCI and arranged them more fluidly across several modules.

While the unit does not directly address the NGSS performance expectations (PEs) for LS4, Biological Evolution, it does incorporate most of the relevant DCIs, science practices (SEPs), and crosscutting concepts (CCs) contained within those PEs—as well as those from LS3, Heredity. Thus, the unit should help to progress students toward being able to complete the PEs. One reason we decided to address the Biological Evolution PEs indirectly was that they did not integrate concepts from heredity as fully as we set out to do in our unit. We decided that this indirect fulfillment of the PEs would make the unit consistent with NGSS while also maintaining its flexibility for teachers in states that have not adopted NGSS. We also anticipated that this will help to maintain the unit's relevance in the coming years as teaching standards and practices continue to change.

Rather than taking a historical perspective, the unit begins with some of the newest, strongest, and most compelling evidence of shared ancestry: all life on earth shares a set of genes and processes required for basic life functions. The unit's lessons continue to revisit the molecular basis of observable phenomena, highlighting

Table 1 Guiding questions, phenomena, and NGSS connections for each module

Module	Driving question	Phenomena used	DCIs	SEPs	CCs
Shared biochemistry	What shapes the characteristics of all living things?	Glowing fish Firefly tails	HS-LS1.A HS-LS3.A HS-LS4.A	Using a model Analyzing and interpreting data Argumentation from evidence	Systems and systems models Patterns
Common ancestry	What is the evidence that living species evolved from common ancestral species?	Cetacean fossils, anatomy, embryos, and DNA Number of shared genes between organisms	HS-LS4.A	Analyzing and interpreting data Argumentation from evidence	Patterns Cause and effect
Heredity	How do the differences arise in DNA that lead to differences in characteristics?	Canine trait similarities and differences Highly variable pigeon traits	HS-LS3.A HS-LS3.B	Using a model Argumentation from evidence	Cause and effect
Natural selection	How do species change over time?	Rock pocket mice coloration Variable lateral plate number in stickleback fish	HS-LS4.B HS-LS4.C	Analyzing and interpreting data Argumentation from Evidence	Patterns Cause and effect
Speciation	How does natural selection lead to the formation of new species?	Species is a human construct Hawthorn fly speciation	*HS- LS4.C*	Analyzing and interpreting data Argumentation from evidence	Cause and effect

Homburger *et al. Evo Edu Outreach* (2019) 12:13

Page 4 of 17

the connections between DNA, protein synthesis, and inherited traits. Thus, the unit explicitly connects these causative mechanisms with the types of observations and inferences that scientists began making in the 1850s. It features DNA as both a source and a record of the unity and the diversity of life.

The modules, and most lessons within, can be used individually or together in sequence (Table 1). With the exception of Shared Biochemistry, each module features one phenomenon that students explore in depth. To illustrate that the principles apply broadly, each module incorporates several additional examples.

When used in sequence, the modules first establish DNA as a blueprint for all living things, and then carry the DNA theme throughout. Later modules highlight DNA's underlying role in variations in heritable traits, which are shaped through natural selection into diverse life forms. So that the materials would be widely usable across student and teacher populations, the modules on common ancestry, natural selection, and speciation focus on non-human examples—though they leave room for human examples, should teachers feel comfortable using them. Throughout the unit, a scaffolded claims-evidence-reasoning framework (Berland and McNeill 2010; Kuhn 2015; Osborne 2010; Toulmin 1958) is designed to gradually build students' skills in constructing arguments from evidence. The descriptions below offer a general outline of the conceptual flow of the modules and describe sample lessons.

Shared biochemistry: what shapes the characteristics of all living things?

The unit's first module, *Shared Biochemistry*, establishes DNA and the process of protein synthesis as common and essential to all life. The module's lessons address the universal structure and function of DNA and proteins. A series of online and paper-based lessons engage students in modeling the process of protein synthesis at three different levels of detail (two of these are shown in Fig. 1). After establishing that all living things make proteins the same way, lessons task students with comparing amino acid sequences from a variety of organisms. Students identify patterns in the sequence data to reveal that even vastly different living things have proteins in common. Finally, this module introduces argumentation. A video describes scientific argumentation as a method for combatting natural human cognitive biases, and it introduces the claim, evidence, and reasoning components of an argument. Students compare and contrast sample arguments, one well-written and one poorly written, for each of two bioengineering phenomena: whether insulin is better medicine for people with diabetes when it is isolated from animals or bioengineered in bacteria or yeast,
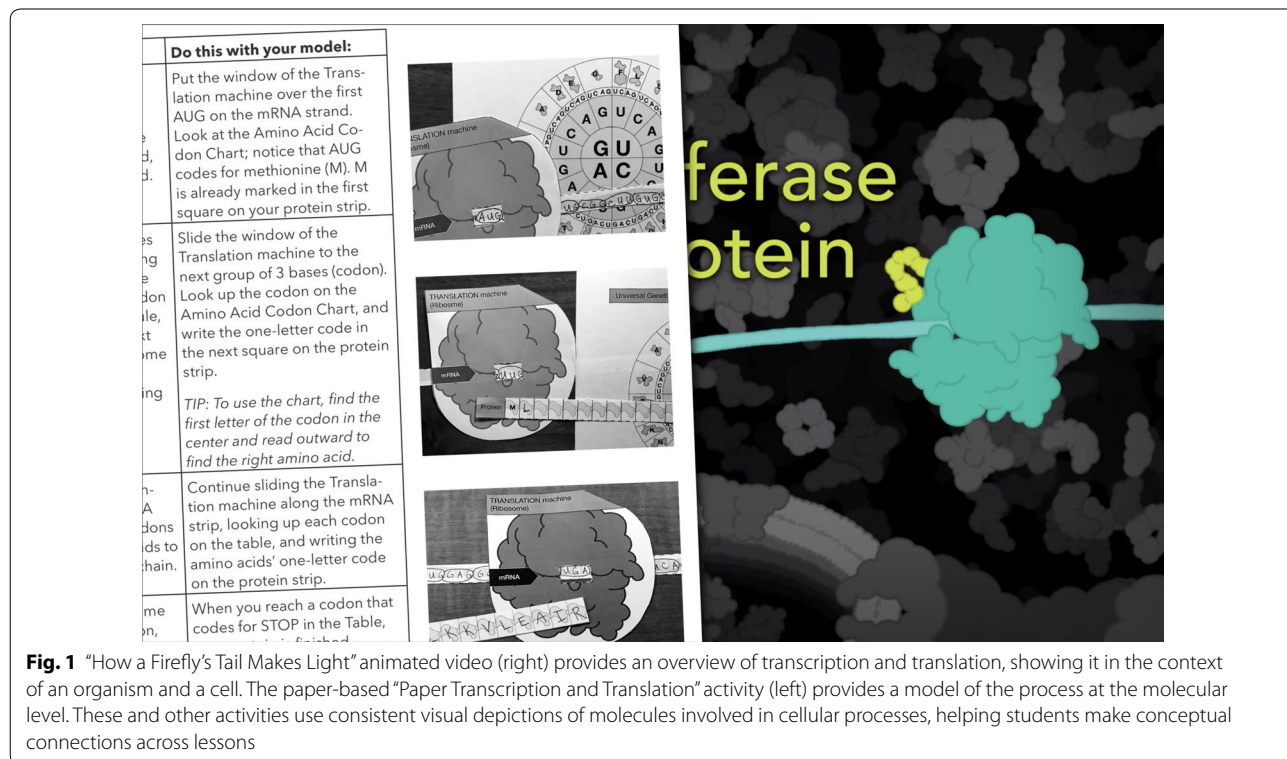

Fig. 1 "How a Firefly's Tail Makes Light" animated video (right) provides an overview of transcription and translation, showing it in the context of an organism and a cell. The paper-based "Paper Transcription and Translation" activity (left) provides a model of the process at the molecular level. These and other activities use consistent visual depictions of molecules involved in cellular processes, helping students make conceptual connections across lessons

and whether mouse cells can make functional firefly luciferase protein. Students practice identifying each component in the sample arguments and evaluate the merit of the arguments according to the inclusion or exclusion of these components. By the end of the module, students should understand that living things are similar at the molecular level, and that these similarities are rooted in DNA—strong evidence that all living things share a common ancestor.

Common ancestry: what is the evidence that living species evolved from common ancestral species?

The next module, *Common Ancestry*, explores the four lines of evidence for common ancestry, as specified in the NGSS: fossils, anatomy, embryos, and DNA. Through a comprehensive case study (Fig. 2), students analyze data from each line of evidence to deduce the ancestry of cetaceans (whales, dolphins, and porpoises). DNA is presented as underlying all of the other lines of evidence. Within the case study, students continue building argumentation skills as they practice identifying the evidence that supports claims and reasoning about cetacean ancestry. The lessons introduce tree diagrams as a

system for organizing information and hypotheses about relationships. Finally, students use an interactive phylogenetic tree (Fig. 2) to identify patterns in genetic data that help indicate the relationships among sample organisms. Through this module, students learn that multiple lines of evidence corroborate hypotheses about common ancestry, similarities among organisms suggest relatedness, and DNA underlies the similarities and differences in each line of evidence.

Heredity: how do the differences arise in DNA that lead to differences in characteristics?

The *Heredity* module examines the genetic processes that generate variation among individual organisms. Focusing first on the source of variation in genes, multimedia presentations introduce the process and outcomes of DNA mutations. Next, students use a paper-based model to make a random mutation in the Human Leukocyte Antigen-B (HLA-B) gene (Fig. 3). They learn how the mutation affects the gene's protein product, and they compare their mutation to known variants. Having established how mutation generates different versions of genes (i.e., alleles), the module lessons next demonstrate how
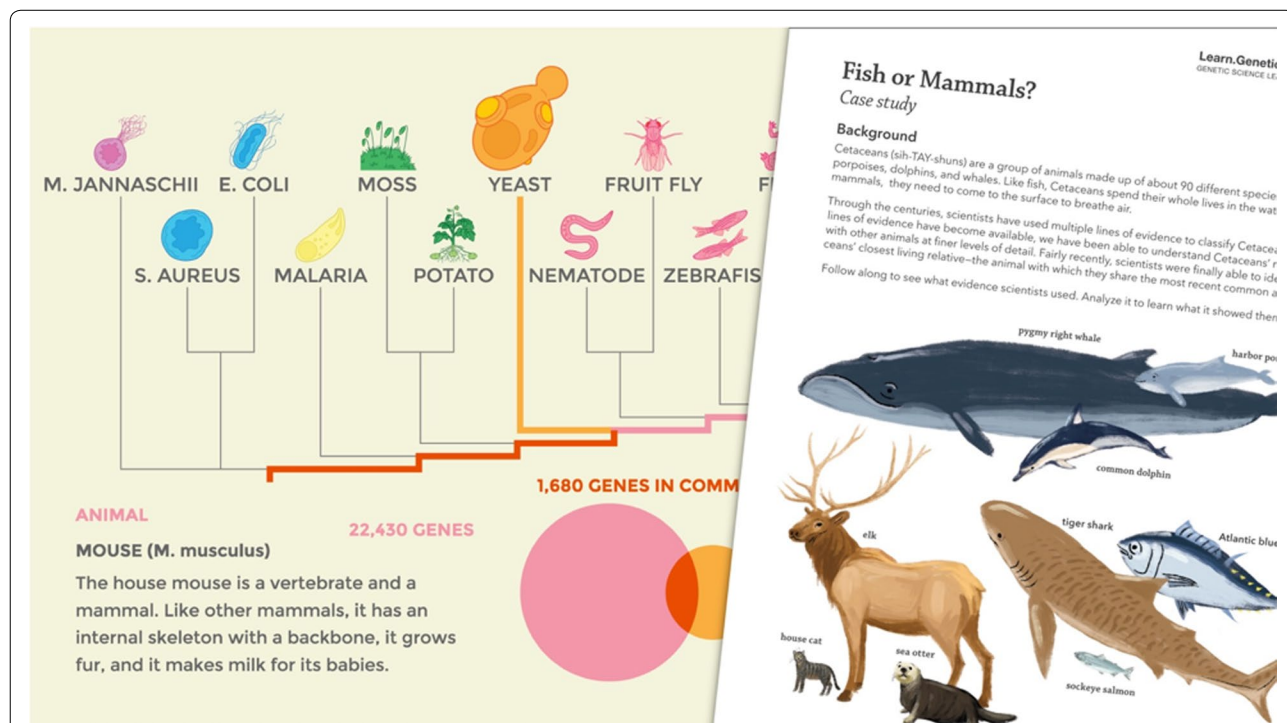

Fig. 2 *Common Ancestry*'s paper-based series "Fish or Mammals?" (right) leads students on a data-based exploration of the four lines of evidence for common ancestry: fossils, anatomy, embryos, and DNA. Each new piece of evidence leads to a more-detailed understanding of cetaceans' relationship with other species, finally revealing that their closest living land-dwelling relative is the hippopotamus. In the online "Interactive Phylogenetic Tree" (left), students explore DNA, which is both a source and a record of evolutionary relatedness. Students choose pairs of organisms on the tree to reveal the number of genes they share (based on published data). This activity reveals the pattern that more closely related organisms, which share a more recent common ancestor, have more genes in common than more distantly related ones
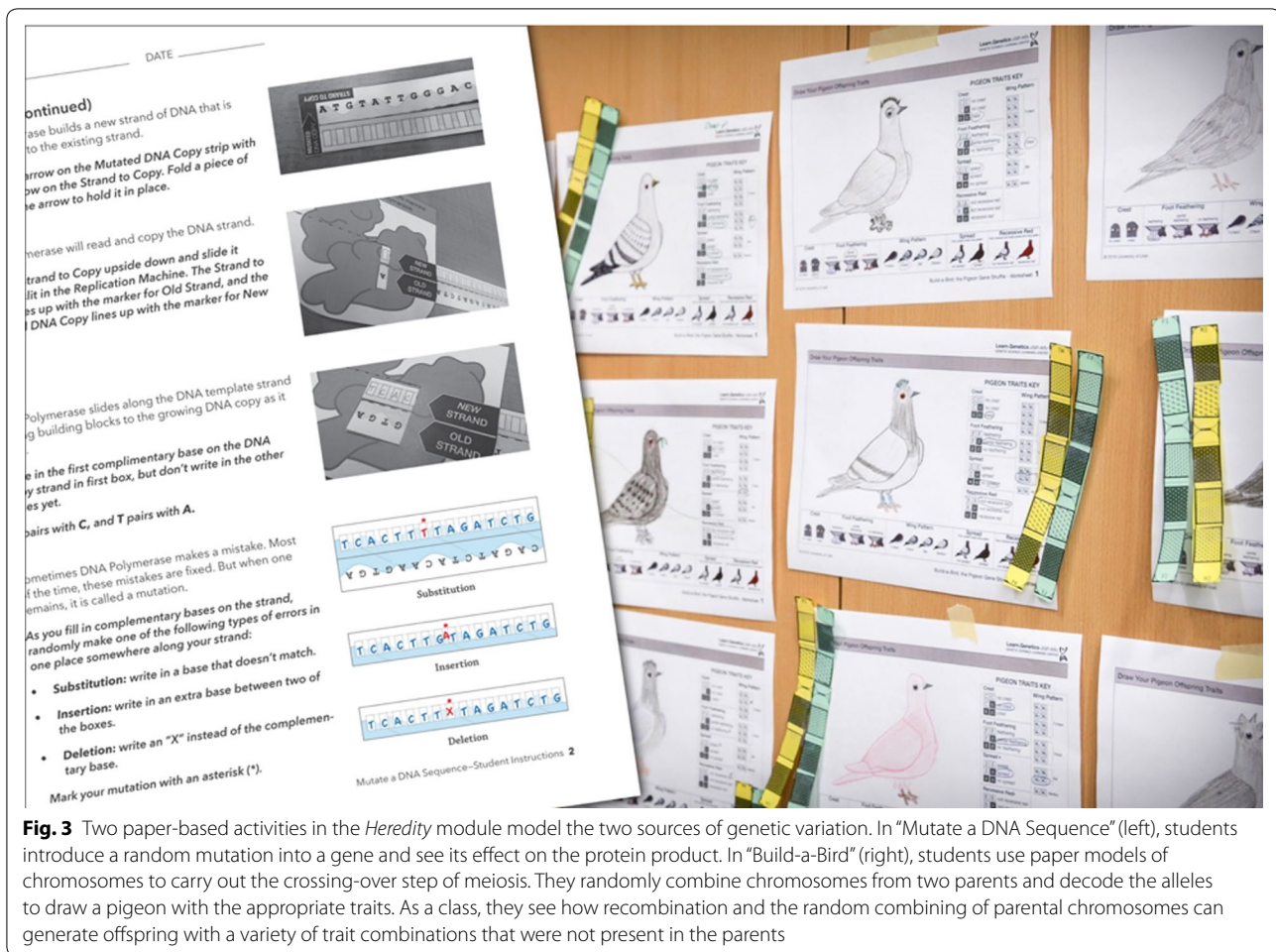
Homburger *et al. Evo Edu Outreach* (2019) 12:13

Page 6 of 17

Fig. 3 Two paper-based activities in the *Heredity* module model the two sources of genetic variation. In "Mutate a DNA Sequence" (left), students introduce a random mutation into a gene and see its effect on the protein product. In "Build-a-Bird" (right), students use paper models of chromosomes to carry out the crossing-over step of meiosis. They randomly combine chromosomes from two parents and decode the alleles to draw a pigeon with the appropriate traits. As a class, they see how recombination and the random combining of parental chromosomes can generate offspring with a variety of trait combinations that were not present in the parents

the shuffling of alleles during sexual reproduction generates variation in a population. Students use a paper-based model of this process in pigeons (Fig. 3), generating a population of birds with an array of characteristics based on known alleles. The model concentrates specifically on recombination and the random combining of gametes rather than the mechanics of meiosis, focusing on the points at which variation is generated. The argumentation practice built into this module tasks students with identifying the appropriate reasoning that links evidence to claims about the source of genetic variation. From this module students learn that two processes increase genetic variation in a population: mutation gives rise to variations in genes (alleles), and sexual reproduction generates new combinations of these alleles.

Natural selection: how do species change over time?

The *Natural Selection* module focuses on the process by which genetic traits become more or less frequent over time, gradually leading to changes in the characteristics

of a population. As species-level changes come about through the same mechanisms, this population-level view prepares students for learning about speciation later. A simulation demonstrates an intuitive example: selection of coat color variants in rock pocket mice in two different environments. Several lessons are centered around a real population of stickleback fish in which researchers have observed a change in body armor. Beginning at a virtual lake (Fig. 4) based on the actual lake), the web-based interactive and associated lessons guide students in analyzing published scientific data. Lessons introduce three criteria for natural selection: variation, heritability, and reproductive advantage. Students analyze relevant data, and then evaluate the extent to which the observed change in the stickleback population meets these criteria. Students organize evidence on a checklist (Fig. 4), which they use to write a supported argument. As reinforcement, students evaluate other examples of changes in characteristics over time. They analyze data, then apply the same three criteria to decide whether the examples meet the requirements for natural selection (some do
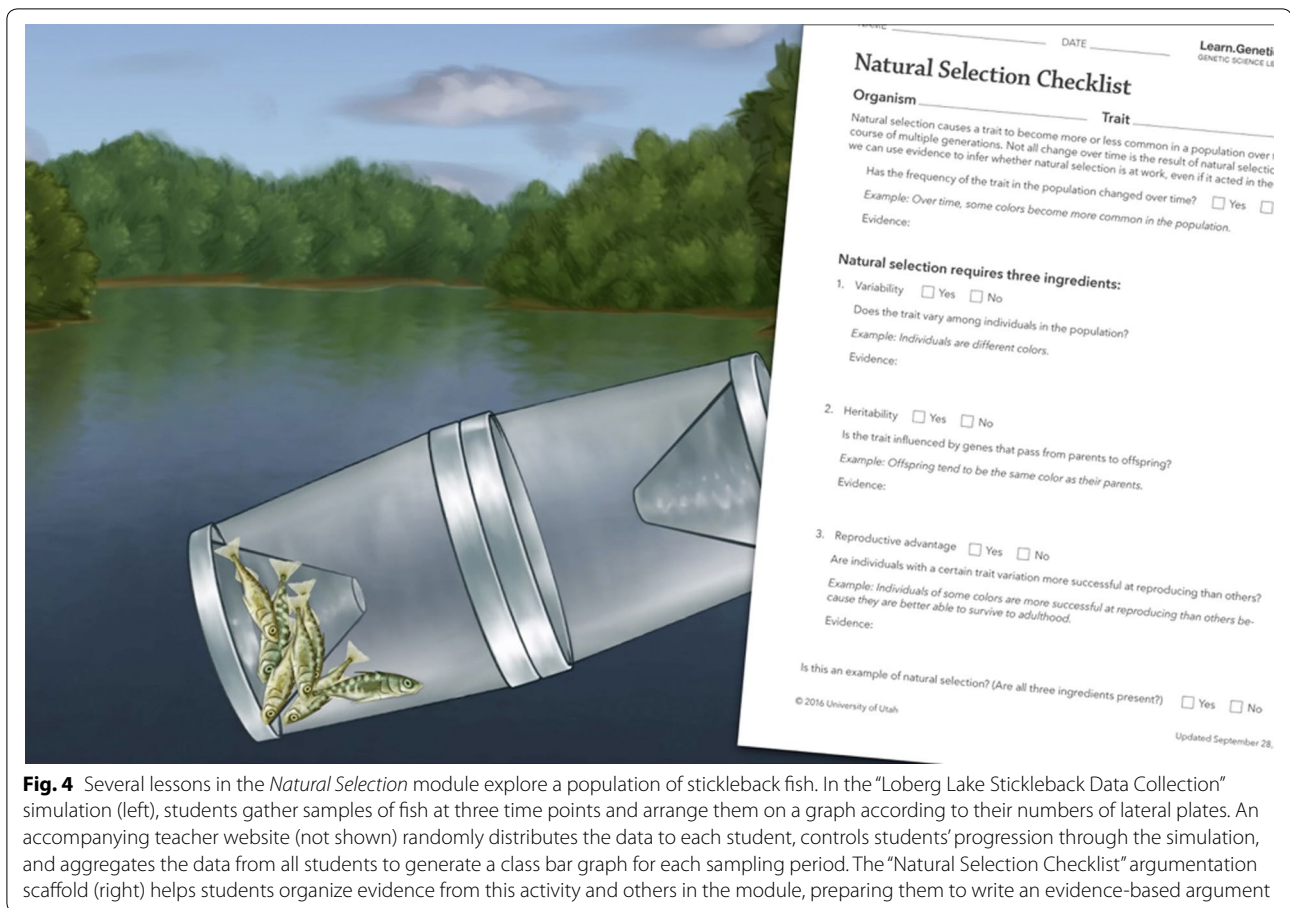
Fig. 4 Several lessons in the *Natural Selection* module explore a population of stickleback fish. In the "Loberg Lake Stickleback Data Collection" simulation (left), students gather samples of fish at three time points and arrange them on a graph according to their numbers of lateral plates. An accompanying teacher website (not shown) randomly distributes the data to each student, controls students' progression through the simulation, and aggregates the data from all students to generate a class bar graph for each sampling period. The "Natural Selection Checklist" argumentation scaffold (right) helps students organize evidence from this activity and others in the module, preparing them to write an evidence-based argument

and others do not). At the module's conclusion, students should understand that natural selection acts on existing heritable trait variations that confer a reproductive advantage, and that this process causes a DNA-based variation to become more or less frequent in a population over time.

Speciation: how does natural selection lead to the formation of new species?

The final module, *Speciation,* investigates what happens when natural selection acts on genetic variation in isolated populations over longer time scales. The module begins by introducing the concept of "species" as a human construct, with a definition that varies according to what scientists are studying and for what purpose. Through the lens of the biological species concept, which focuses on reproductive isolation, students explore several ambiguous examples. These examples demonstrate that species are not always distinct, nor are they fixed—setting the stage for students to understand speciation as a process. Next, students delve into a data-rich

case study of *Rhagoletus* flies, again based on published research (Fig. 5). These flies may be diverging into two species based on their preferences for different host fruit: apples or hawthorn berries. Students analyze data about the flies' life cycles, allele frequencies, and host fruit preferences.

An organizing worksheet guides students in compiling the various lines of evidence, helping them decide whether the flies are reproductively isolated, and whether different heritable characteristics are being selected for in each population. Weighing the evidence, students determine where the populations fit on a continuum between "same species" and "different species." Using their organized evidence, students write a supported argument that justifies their chosen placement along the continuum. The module (and unit) concludes with a video that connects multiple processes—genetic variation, natural selection acting on multiple traits over many generations, and reproductive isolation—to explain the continuous branching of genetic lineages and the divergence of life over time. Through this module, students should understand the processes that cause characteristics of
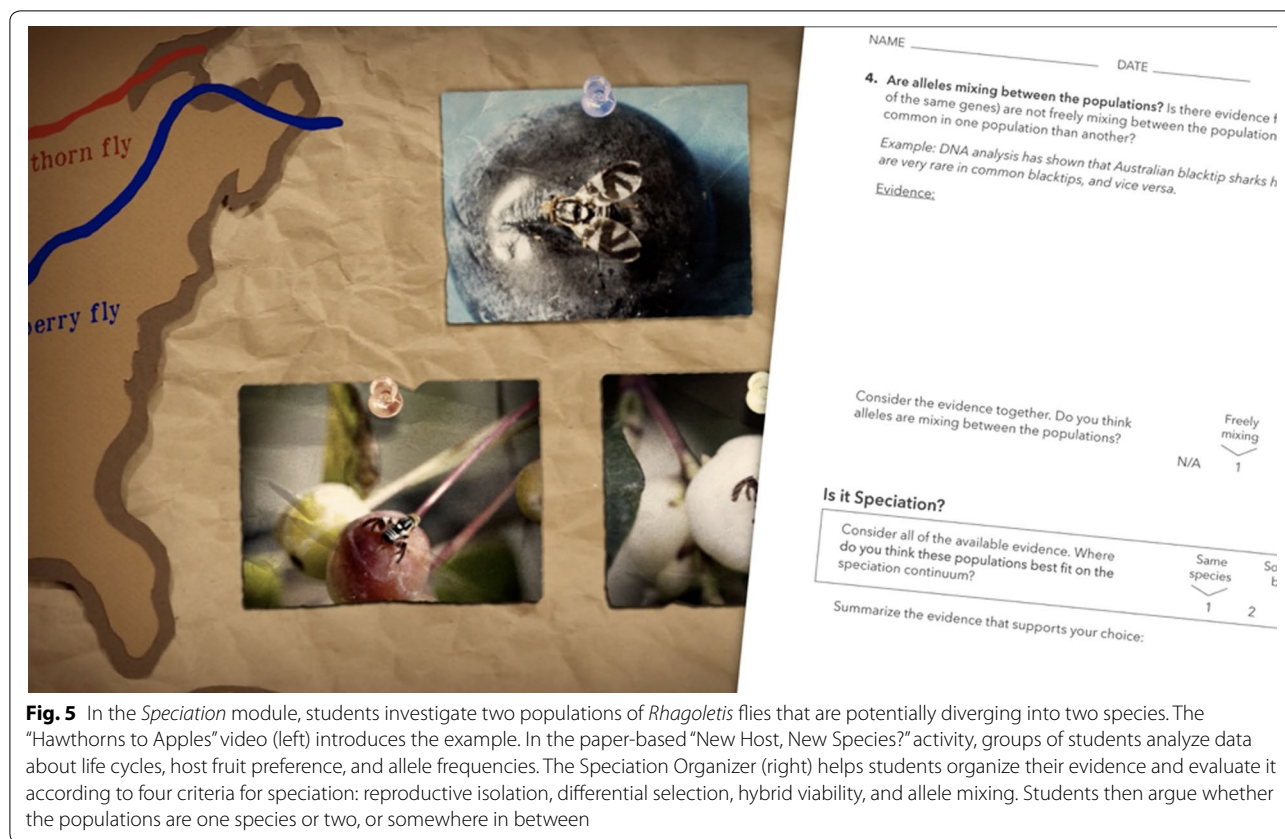
Fig. 5 In the *Speciation* module, students investigate two populations of *Rhagoletis* flies that are potentially diverging into two species. The "Hawthorns to Apples" video (left) introduces the example. In the paper-based "New Host, New Species?" activity, groups of students analyze data about life cycles, host fruit preference, and allele frequencies. The Speciation Organizer (right) helps students organize their evidence and evaluate it according to four criteria for speciation: reproductive isolation, differential selection, hybrid viability, and allele mixing. Students then argue whether the populations are one species or two, or somewhere in between

living things to diverge, and that species differ from one another across multiple heritable traits.

Built-in assessments

Formative assessments (Fig. 6) are embedded in the lesson sequence of each module. The tasks provide opportunities for students to explain their thinking though written activities and other forms of work, eliciting and revealing complex student cognitions (Coffey et al. 2011; Kang et al. 2014). The assessments are designed to help teachers quickly and efficiently evaluate students' progress and refocus instruction as needed. The highly visual tasks use short writing prompts and multiple-choice items with written justification. They evaluate students' conceptual understanding, data analysis and interpretation skills, and argumentation skill. At the end of the unit, teachers may administer one of two optional open-ended summative assessments, both of which ask students to reflect on their understanding of evolution using evidence-based justifications for their responses. One of the assessment options uses two items from the ACORNS instrument (Nehm et al. 2012), which assesses students' written explanations of evolutionary change and can be scored using the related online, free EvoGrader tool (Nehm 2011).

Accessing the unit

The unit's materials are freely available and hosted on two parallel websites: one for students (http://learn.genetics.utah.edu/content/evolution/), and the other an enhanced version for teachers (http://teach.genetics.utah.edu/content/evolution/). The teacher site contains a wealth of support materials. It includes guiding questions and learning objectives; short videos that summarize each module; at-a-glance lesson summaries that include connections to NGSS SEPs and CCs; in-depth guides with suggestions for implementation; copy masters; answer keys; and discussion questions. Video guides support teachers in implementing some of the more complex lessons.

The suggested lesson sequence and implementation instructions are consistent with the NGSS topic arrangements. But because education standards vary by state, the unit's lessons were designed to be used flexibly. They may be used in whole or in part, with or without the addition of outside materials. The unit's lessons are designed to be easily accessible and cost effective. Hands-on activities use only low-cost materials that are readily available in most classrooms. Teacher instructions include tips for minimizing and re-using material resources. Nearly all of the online components work across platforms, including on tablets and smartphones.

Scientists can put spider genes into yeast. The yeast can decode the spider genes to make spider silk protein. Why is this so? Evaluate the models.
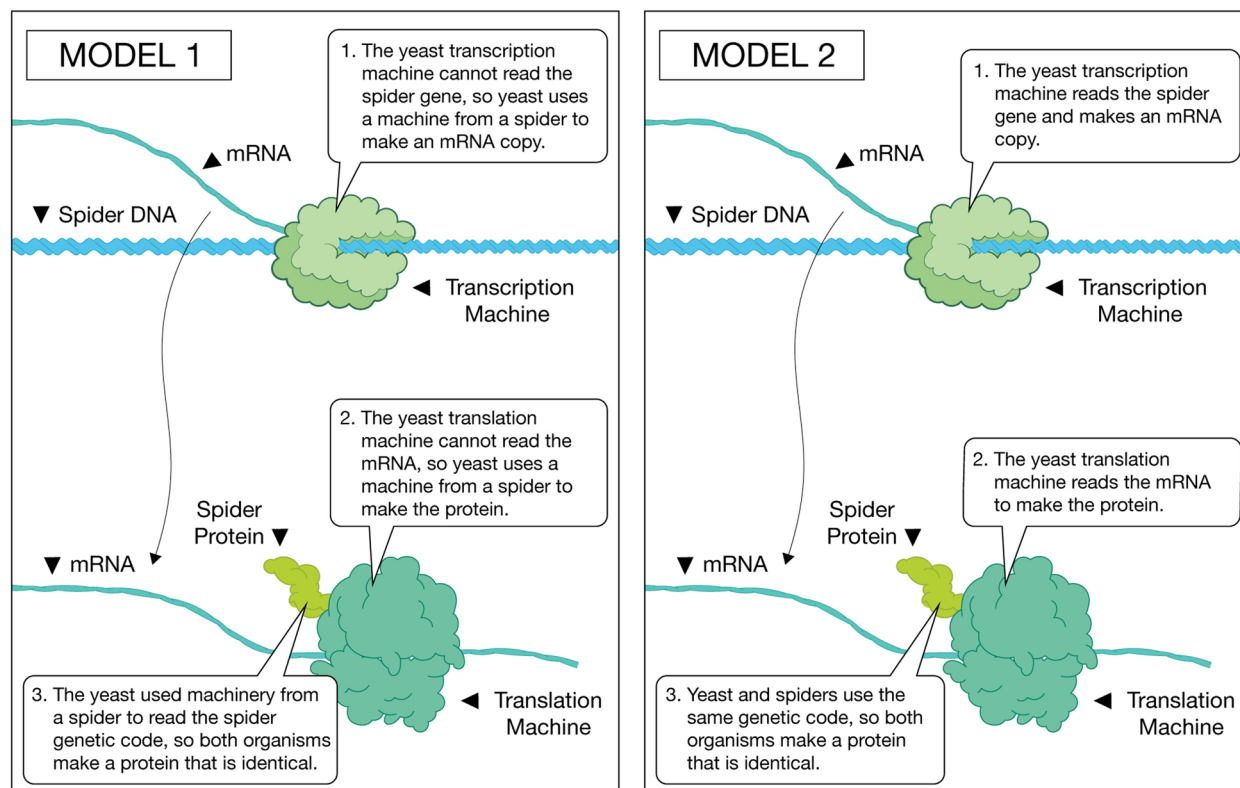
MODEL 1

1. The yeast transcription machine cannot read the spider gene, so yeast uses a machine from a spider to make an mRNA copy.

▲ mRNA

▼ Spider DNA

◄ Transcription Machine

2. The yeast translation machine cannot read the mRNA, so yeast uses a machine from a spider to make the protein.

Spider Protein ▼

▼ mRNA ▸

3. The yeast used machinery from a spider to read the spider genetic code, so both organisms make a protein that is identical.

◄ Translation Machine

MODEL 2

1. The yeast transcription machine reads the spider gene and makes an mRNA copy.

▲ mRNA

▼ Spider DNA

◄ Transcription Machine

2. The yeast translation machine reads the mRNA to make the protein.

Spider Protein ▼

▼ mRNA ▸

3. Yeast and spiders use the same genetic code, so both organisms make a protein that is identical.

◄ Translation Machine

Fig. 6 In this assessment task, students choose a model that best describes why yeast can decode spider genes to make spider silk protein. The teacher website (not shown) includes other ideas for assessments, which teachers may choose if they have more time available or if their students need extra practice

Unit development and early testing

The *Evolution: DNA and the Unity of Life* unit was developed by the Genetic Science Learning Center (GSLC) at the University of Utah. The team included curriculum developers, instructional designers, biology education specialists, science writers, multimedia producers, visual designers, animators, computer programmers, videographers, a music composer and audio engineer, web developers, and education researchers, along with significant input from teachers and scientists with relevant expertise. Pre/post assessments for evaluating student learning of the target science ideas were developed by AAAS Project 2061.

Theoretical framing of the curriculum

Each stage of unit development was informed by the GSLC team's theory of change. We posited that students will better understand the disciplinary core ideas about biological evolution when curriculum materials and instruction:

- Integrate pertinent topics in heredity;
- Provide opportunities to analyze and interpret data;
- Engage students in argument from evidence;
- Include consistent visual depictions of key molecules and processes.

Our development framework drew on constructivist, conceptual change, and situated cognition theories of learning. The curriculum guides students in constructing knowledge about evolution through a process of hypothesis testing and interacting with phenomena (Driver 1995). During these processes they have opportunities to access their current understandings and evaluate them in light of the learning experience(s) in which they are engaged. The resulting cognitive dissonance supports students in modifying their conceptual structures (Strike and Posner 1992). Social interactions and communication with other students that involves explicating, exploring, and exchanging ideas contribute to this process and reinforce learning that is congruent

Homburger *et al. Evo Edu Outreach*   (2019) 12:13

Page 10 of 17

with the scientific ideas and theories that have been socially constructed by the scientific community. Students use authentic scientific tools and practices to gain new knowledge and skills, while their teachers provide scaffolds to support student learning (Brown et al. 1989).

Our development framework was informed by several learning progressions. Catley et al. (2004) developed an evolution learning progression for elementary and middle school grades that "unpacks" AAAS Benchmarks (1993). While they did not extend their learning progression to the high school level, we reviewed the progression they developed for earlier grades and attended to their assertion that evolution education needs to focus on "big ideas" that integrate across the multiple disciplines. As they recommend, we decided to engage students in analyzing data and in constructing evidence-based arguments, making these the two primary SEPs for the unit.

We also consulted the genetics learning progression developed by Duncan et al. (2009), and identified the core ideas for high school that are relevant to understanding evolution. In addition, we looked at the core ideas for middle grades and considered ways to briefly review and remind students about these ideas. While developing the unit SEPs, we considered Berland and McNeill's scientific argumentation learning progression (Berland and McNeill 2010). Our alpha testing of the Natural Selection module showed that most students needed more scaffolding for learning how to construct evidence-based arguments. We therefore incorporated a scaffolded approach to constructing arguments using the claims, evidence, and reasoning framework, taking into account the components of the learning progression.

Unit development and early testing

Development and testing of the unit followed an iterative, multi-step, multi-year process. The Natural Selection module was developed first, and underwent several rounds of development, classroom testing, and revision. It was then beta-tested with over 1200 students taught by seven teachers across the U.S. and revised again (Stark et al. 2016).

We next developed the outline and sequence for the remaining four modules. We identified appropriate, engaging phenomena and associated published data to draw from. The unit-wide argumentation scaffold was drafted, along with paper and multimedia lessons and activities for two of the modules. These were tested locally in one teacher's classroom. Researcher observations, teacher interviews, and student informal interviews provided data for lesson revisions. They also provided proof-of-concept evidence for the evolving unit's conceptual flow, classroom utility, and

effectiveness for learning. We completed drafts of lessons and activities for the remaining modules, along with drafts of embedded formative assessments. To establish the degree of alignment to the NGSS, an external reviewer (AAAS Project 2061) conducted an alignment evaluation of components of the unit using the Educators Evaluating the Quality of Instructional Products (EQuIP) rubric (Achieve Inc. 2016). The analysis provided feedback on parts of the curriculum that claimed to have alignment to specific science practices and crosscutting concepts but were insufficient for robust alignment. We removed these claims of alignment. This process prompted us to make more explicit the parts of the materials that did have robust alignment.

Unit pilot testing
Participants and professional development
We conducted the curriculum unit pilot test in the 2016–2017 school year to evaluate the unit's classroom utility, usability, and effectiveness for student learning. We invited teachers to submit an application to participate in the pilot study through the GSLC's email list of over 24,000 educators. From the 372 applicants, we recruited 20 biology teachers from 11 states (AR, CA, KS, LA, OH, OR, MD, PA, NJ, NM, UT) and Canada. Inclusion criteria included teaching at least two sections of introductory or honors biology (grades nine and ten). Selected teachers represented broad ranges of students across ethnic, socioeconomic, and geographic categories. The sample included special education, honors, and general education students. Teachers represented both public and private schools in urban, suburban, and rural settings, block and daily instruction schedules. Years of teaching experience ranged from 6 to 31. Five local teachers were recruited to allow for in-person classroom observations.

The demographics for student participants (the students of the pilot teachers) were as follows: 54% of the sample were female; English was not the primary language for 6%; 4% were special education students; and 49% were eligible for free or reduced lunch. Racial and ethnic demographics were 54% White, 13% Hispanic or Latin American, 8% Black/African American, 7% Other, 6% Asian, 5% American Indian or Alaskan Native, and < 1% Native Hawaiian or Pacific Islander.

In summer 2016, the teachers came to the University of Utah for a 3.5-day in-person training institute. They practiced using the draft lessons, received instruction in implementation, and provided feedback. This feedback informed unit revisions and further development. Of note, the majority of these teachers told us that they felt there were significant barriers to their using human examples in evolution instruction. Thus, we decided to

focus our efforts on non-human examples that everyone could use. We included optional human examples in some lessons, and there is room for teachers to add their own examples.

Pilot test data collection and results

The remainder of this section describes the data collection and results around each of the goals of the pilot study:

1. Evaluate and improve the usability of the materials for teachers and students.
2. Gauge the perceived educational value of this unit compared to the evolution curriculum materials teachers have used in the past.
3. Gather initial evidence of student learning gains from the unit.

Goal 1: Classroom usability

After the summer training, the 20 teachers implemented the unit in their introductory biology classrooms (2016–2017 school year). GSLC staff conducted daily observations in 5 classrooms in local schools and had conversations with the teachers. To capture implementation data from the remaining classrooms and additional reflections from the observed teachers, the GSLC's internal and external evaluators developed logs for the teachers to complete after each day of teaching the unit. GSLC staff and pilot test teachers vetted the instruments, and each was revised by the evaluators. We used the data to gauge teachers' classroom experiences with the materials, including issues or problems. Daily log questions included the following:

- Regarding implementation, student engagement, timing, or instructions:

  - What worked well today?
  - Did you encounter any unforeseen problems?
  - Do you have any suggestions for improvement?

Evaluators received 365 total logs from the 20 teachers (range 11–29 logs per teacher, average = 18.25). Three teachers completed most but not all of the unit, due to time constraints. Two teachers completed approximately half of the unit; one could not be reached for follow-up and the other indicated the reading level was challenging for his special education students. Evaluators sent the pertinent teacher feedback to curriculum developers daily to inform revisions. Further, the evaluators together reviewed teacher logs to develop initial patterns and

themes (Miles and Huberman 1994). We used the classroom observation data to provide support for the themes.

Based on this feedback, we revised many lessons (sometimes substantially), removed a few and made some optional, and developed new lessons. For example, in response to teachers' feedback that their students seemed to be getting bored with the cetacean and stickleback fish lessons, which extended across several class periods, we streamlined some of these lessons significantly by making them more concise. Other examples include revising the estimated implementation time of activities; reducing the number of worksheets; making some of the formative assessments more visual to decrease reading and scoring time for teachers; adding alternative paper-based versions of some web-based activities; and adjusting lesson sequences.

Ten teachers implemented the lessons in the fall and the other ten implemented in the spring. This allowed for re-testing modified activities, testing new activities, and development and testing of some of the teacher support materials. On average, the fall teachers spent 10 weeks teaching the unit. Our primary revisions were streamlining and trimming materials while keeping the key, integral aspects of each activity. Therefore, the unit maintained the key aspects of each activity for spring testing. The spring teachers spent approximately 6.5 weeks on the unit. We present student gain results comparing fall students to spring students in the Student assessment results section.

Additional teacher support materials were developed after the spring pilot testing, including instructional videos and additional formative assessment items. These support materials were informed by pilot teacher feedback, and they aimed to clarify suggested implementation instructions in the places where teachers had the most questions and challenges. In many cases, the draft teacher support materials did include all of the necessary information, but teachers were either not reading it or not recalling it at key moments. To address this issue, we made several changes, including moving copy instructions from teacher guides or online text to the pdf documents to be copied, trimming peripheral information from teacher guides in order to emphasize key details, re-writing and formatting instructions to make them easier to scan, and arranging instructions so that teachers would see key information closer to the time that they would need to implement it.

Goal 2: Educational value

The evaluators created an end-of-implementation survey for teachers to complete on the final day of pilot testing. We used the survey data to assess the overall appeal

Homburger *et al. Evo Edu Outreach* (2019) 12:13

Page 12 of 17

of the unit and teachers' perceptions of the educational value of the unit compared to current practices. As with the teacher log, GSLC staff and pilot test teachers vetted the instruments, and each was revised by the evaluators. Questions included the following:

- What did you like best and least about the unit?
- Do you plan to use this unit or parts of this unit in future years?
- How did the unit compare to other units you have used to teach similar content?

The evaluators reviewed the surveys independently and identified broad themes that focused on initial patterns and perceptions of critical issues (Miles and Huberman 1994). Next, we engaged in a cooperative, cyclical process of analyzing the data, "refining and modifying the data at multiple levels of complexity in order to locate the main essence or meaning" (Stake 2005, p. 389). We narrowed our themes and used the teacher log data and informal conversations with teachers during classroom observations to provide further support for the findings. Eighteen teachers completed the survey (the two who did not complete the survey were not available for follow up).

The data showed that twelve teachers (66.7% of respondents) reported that the unit was better than curriculum materials they had used in the past and three (16.67%) noted that it was as good as their current materials. The remaining three (16.7%) indicated that some parts of the unit were better than materials they had used in the past, and that some parts were not as good. Teachers indicated that the unit was superior to others they have used in the following ways: the use of real-world data, the CER scaffold and opportunities to build the practice of argumentation, unit design that allows students to take ownership over their learning, and the scientific research that went into designing the activities. Teachers preferred other materials for their lower reading levels, which they said were more appropriate for their special-education and low-achieving students. Several of these teachers, however, indicated that the materials are straightforward enough to modify to a lower reading level.

Among the aspects that teachers liked most about the unit were that it builds conceptual understanding of evolution by starting with the biochemistry underlying evolution and ending with speciation, that the unit was thoughtfully and carefully designed to tell the story of evolution in a way that resonated with students, and that students were engaging with phenomena and analyzing data from published scientific research studies. Further, every teacher who completed the survey indicated appreciation for the argumentation framework and the scaffolding used in the unit. Comments included that it simplified and structured what could be a very complicated process, it built students' capacity to argue from evidence, and it provided opportunities to hear other students' perspectives. As one teacher explained, *The area that I think the students grew the most in was the CER—claim, evidence and reasoning technique. This really allowed them to start to think more for themselves.*

Key challenges reported were that the unit was longer than they typically spend teaching evolution (particularly fall semester teachers who used the unit before we modified the length), that the amount and level of reading proved especially challenging for some students (as described earlier), and the large number of worksheets and the associated printing and reading required. For example, *It was too long—most of our units last a maximum of 2–3 weeks because of all the topics we have to cover during the year*; *Some of the reading examples were difficult for some of the students, especially those with learning disabilities and for English Language Learners*; and *I did not like how much of the unit was done* via worksheets.

In spite of these concerns, all 18 teachers indicated that they would use all or parts of the unit in the future. Nearly half ($n=8$) planned to teach the unit in sequence, but add labs or other hands-on activities. One-third ($n=6$) would teach select elements of the unit. Three of those teachers planned to teach all of the modules, but not all of the activities in each. One teacher expected to use all of the materials except for the heredity module, *This is only because I usually cover much of this earlier in the year, and go into a lot more detail with my students.* The remaining two teachers planned to teach the Natural Selection and Speciation, and the Shared Biochemistry and Natural Selection modules, respectively. Overall, results from the data sources illustrate the feasibility and perceived educational value of the curriculum materials.

Goal 3: Initial evidence of student learning

Multiple-choice student assessment items were created in parallel to the curriculum by AAAS Project 2061. The assessment items were written to be aligned to the same NGSS DCIs and SEPs as the curriculum. Items were not written to be directly aligned with the curriculum but rather indirectly through the NGSS learning goals that the curriculum was addressing. For most items, students were expected to apply their knowledge of basic science ideas to phenomena that were different from what they experienced in the curriculum. Thus, the items were more "distal" to the curriculum than the items that characterize most classroom tests. The assessment items were pilot tested nationally with 4588 middle and high school students. Based on student answer choice selection and

written pilot test feedback, 84 items were judged to be acceptable for assessing students' understanding of the ideas and practices targeted in the unit.

Items assessing the argumentation practice were limited to assessing students' ability to identify claims, evidence, and reasoning in the context of evolution. In the topic-level summaries of learning gains, students' scores on the argumentation items were counted toward both argumentation and the relevant evolution sub-topic. Items assessing the practice of data analysis did so in conjunction with assessing evolution content knowledge and were limited in their number; therefore, we do not report results on student's understanding of this practice. See Additional file 1 for sample assessment items.

To evaluate the pilot curriculum, the 84 items were distributed across four test forms. Each test contained 25 items, including seven linking items. Items were distributed such that each test had a similar number of items per topic (i.e., Shared Biochemistry, Common Ancestry, Natural Selection, etc.), and equivalent average test difficulties. The pre- and posttests were administered online, and students in a given classroom were randomly assigned one of the four test forms so that results from all forms were available from each classroom. On the posttest, each student received a different form than their pretest, to minimize test–retest effects. Teachers were asked to administer the pretest immediately before starting the pilot test and the posttest immediately after ending the pilot test.

Rasch modeling using WINSTEPS (Linacre 2016) was used to examine test, person, and item reliability in order to assess the reliability of the assessment instrument. Overall test and person reliability were high (.97 and .79 on the pretest and posttest, respectively), and each item had positive point-measure correlations and acceptable fit (between .7 and 1.3) to the Rasch model (Bond and Fox 2013). All items were modeled together to measure the students' overall knowledge of evolution. A Principle Component Analysis (PCA) (Linacre 1998) of the fit residuals did not show significant loading on multiple dimensions, suggesting the test was substantively unidimensional and could be treated as measuring a single trait (i.e., evolution). These results, in combination with care taken in developing and aligning the assessments to the pertinent NGSS learning objectives, provide evidence that the pre/posttest assessments were a reliable and valid measure of students' understanding of evolution.
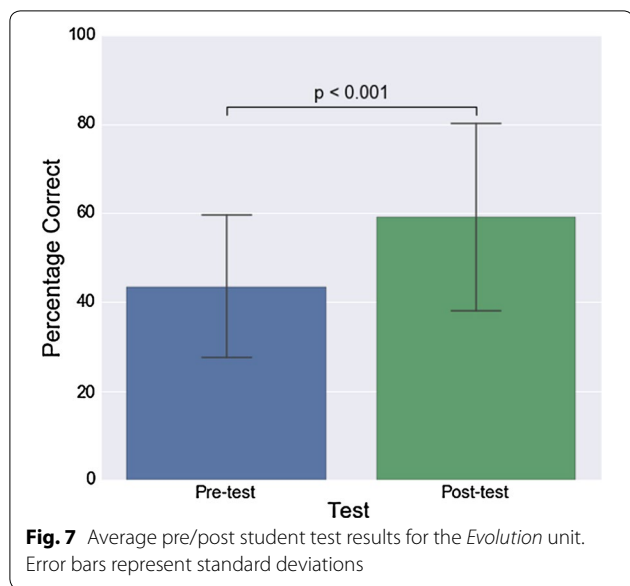
Student assessment results   Assessment data from the curriculum pilot test represent 944 students who completed both pretests and posttests (Table 2). An additional 120 students experienced the curriculum but did not complete their assessments.

Table 2 Pilot teachers' (*n* = 16) classroom demographics and pre/post gains

Teacher	Gender (%)		Primary language (%)		Ethnicity (%)							Pre/post gain (%)	
	Female	Male	English	Non-english	AI	AS	BL	HI	PI	OT	WH	Mean	Standard deviation
A	52	48	86	14	0	17	2	29	2	10	40	10	15
B	68	32	96	4	0	2	2	0	0	7	89	26	13
C	20	80	40	60	0	0	0	40	40	0	20	29	9
D	59	41	93	7	1	11	2	18	2	16	50	15	16
E	39	61	96	4	0	2	78	4	0	4	13	4	14
F	69	31	95	5	93	0	0	2	0	2	2	10	14
G	42	58	94	6	2	2	1	8	0	6	80	14	17
H	63	37	100	0	0	0	95	0	0	5	0	5	13
I	73	27	97	3	2	2	0	8	0	3	86	20	13
J	41	59	80	20	0	2	12	57	0	10	19	15	15
K	51	49	94	6	3	3	3	17	0	3	71	26	14
M	52	48	93	7	2	2	7	18	0	14	57	6	13
N	55	45	98	2	2	5	2	12	0	2	77	10	15
O	69	31	98	2	0	3	3	2	3	3	85	22	14
P	50	50	86	14	0	14	0	0	0	36	50	22	16
Q	54	46	100	0	1	8	3	3	0	14	71	21	15

Four teachers were not included in our final analysis because they either did not administer the post assessment, they administered the assessments incorrectly, or we did not receive permission to share their assessment results (see Additional file 3 for a discussion of the variance in gains between teachers)

AI, American Indian or Alaskan Native; AS, Asian; BK, Black/African American; HI, Hispanic or Latin American; PI, Native Hawaiian or Pacific Islander; OT, other or multiple ethnicities selected; WH, white

Homburger *et al. Evo Edu Outreach*   (2019) 12:13

Page 14 of 17

Fig. 7 Average pre/post student test results for the *Evolution* unit. Error bars represent standard deviations
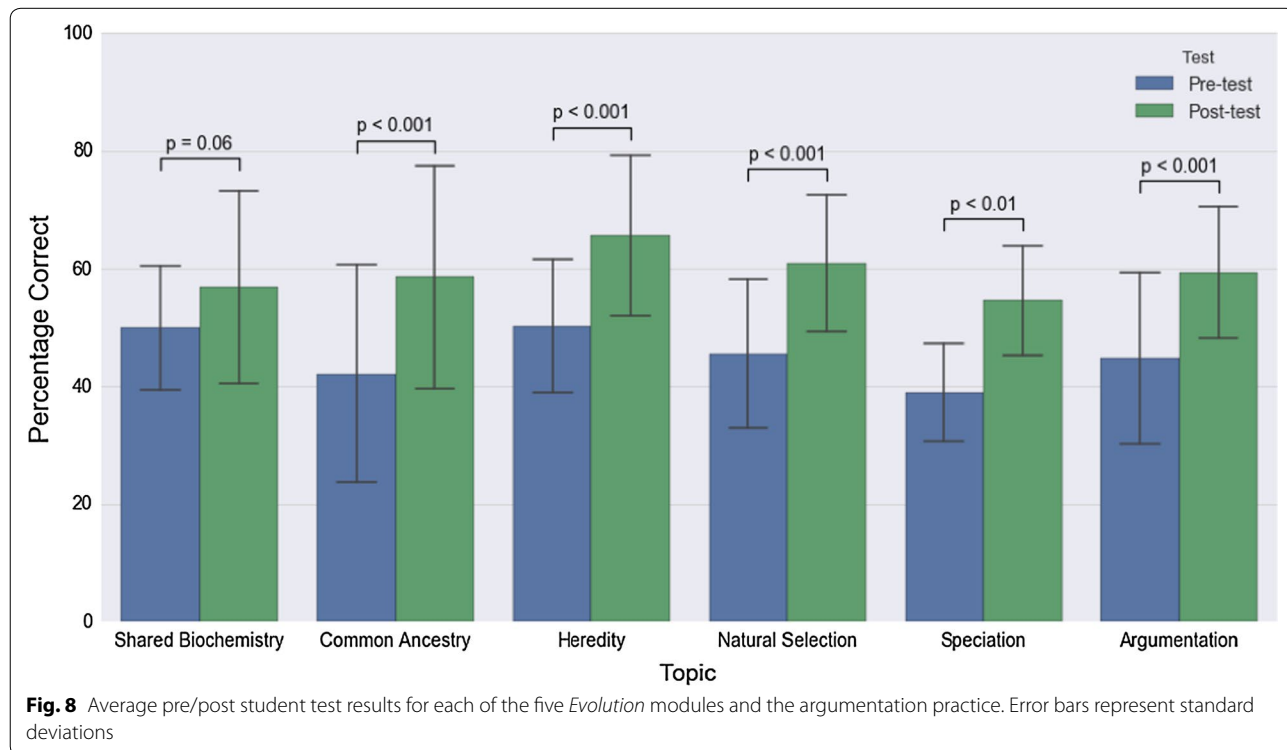
Bonferroni-adjusted paired *t* test results revealed a statistically significant increase in student scores from pretest to posttest (Fig. 7), with an average gain of 17% points: $t(943) = 29.6$, $p < .001$, Cohen's $d = .96$. We also observed an increase in the number of students getting a majority of the test items correct (see Additional file 2 for a histogram of students' percentage correct scores on the pre/posttests). An analysis of performance differences across demographic subgroups indicated that gender, primary language, and special education status did not result in statistically significant differences in improvement from pretest to posttest; however, small marginally significant effects on performance gains were found for some ethnicity comparisons (see Additional file 3 for demographic details).

Paired *t* tests on subscale results indicated statistically significant knowledge gains for four of the five modules ($p < .01$–.001) and for identifying the CER components of an argument ($p < .001$) (Fig. 8). The *p* value for the Shared Biochemistry module, at .06, was not statistically significant; we discuss possible reasons for this result in the limitations section. Students increased between 14 and 16% points from the pretest to the posttest on each module.

Even though the spring students spent on average 3.5 weeks less time on the unit, we found no statistical difference between the gains of students in the fall and spring ($p = .79$). These results suggest that our end-of-fall revisions that included streamlining and trimming were effective in keeping the integrity of each activity while reducing time spent on the unit. In other words, the materials we removed were not integral to student learning of the tested concepts from NGSS.

Fig. 8 Average pre/post student test results for each of the five *Evolution* modules and the argumentation practice. Error bars represent standard deviations

At the end of the testing year, AAAS Project 2061 provided the curriculum development team with a list of student misconceptions that were represented in the multiple-choice items, and the percentages of students who incorrectly chose them as answers on the pretests and posttests (see Additional file 4 for a list of misconceptions and percentage of students who chose them as answers on the posttest). The curriculum development team used this information to inform revisions of the lessons, making an effort to address the misconceptions that students chose at high frequency.

Conclusions

The goals of the curriculum pilot test, conducted in 2016–2017, correspond to the Design and Development phase of educational research (Institute of Education Sciences, U.S. Department of Education, National Science Foundation. Common Guidelines for Education Research and Development: A Report from the Institute of Education Sciences, U.S. Department of Education, and the National Science Foundation 2013) requiring a theory of action, articulation of design iterations, and initial evidence of effectiveness. We accomplished our three primary goals for this stage of curriculum development and testing. First, in fall pilot testing, we gathered and analyzed extensive teacher feedback through daily teacher logs and conversations, and made (sometimes substantial) revisions and refinements to the curriculum based on the feedback. Key revisions included streamlining some activities to reduce overall unit time and to enhance pacing, reducing text on teacher support materials and developing short teacher-support videos, and adding figures to the formative assessments to reduce writing requirements. We then re-tested the materials in the second half of the school year.

Second, teacher survey data provided us with an understanding of teachers' perceptions of the educational value of the materials. These findings showed teachers' appreciation for the unit's use of real-world data, the CER scaffold and opportunities to build this skill, the building of conceptual understanding of evolution, and student ownership over learning. The majority of teachers indicated that the unit is superior to others they have used in the past, despite their concerns over high reading levels that are challenging for some students. These findings illustrate that the unit is feasible for teachers to implement, and that teachers view it as having educational value. Third, results from student pre/posttesting revealed that students who experienced the unit learned the DCIs for evolution and heredity, and gained skill in identifying claims, evidence, and reasoning in scientific arguments.

Overall, this research suggests that teaching heredity and evolution in an integrated unit, combined with exposure to numerous sources of evidence and practice in constructing arguments, facilitated student understanding of evolution. This is consistent with our theory of change. We conclude that the *Evolution: DNA and the Unity of Life* is an example of a unit that was designed for the NGSS and that demonstrates initial evidence of effectiveness—which we defined at this stage as feasibility and usability for teachers, and statistically significant student learning gains.

The results reported here set the stage for a larger randomized controlled trial, which was conducted during the 2017/2018 school year. This trial compares learning gains made by students whose teachers were assigned to either the treatment (our unit) or control (NGSS-aligned "business as usual") condition. Because treatment teachers used only the online teacher support and received no additional training, it is also a test of those materials' effectiveness. Once data analysis is complete, the efficacy trial will enable us to explore new questions about the mediating factors that might influence the observed outcomes. It will contribute to knowledge of the critical components of effective instruction in evolution (Ziadie and Andrews 2018), which is a gap in educational research. In the meantime, educators can use the free *Evolution: DNA and the Unity of Life* curriculum with confidence in the materials' feasibility and educational value.

Limitations

This work had several limitations that should be acknowledged. First, regarding the student pre/post assessments, items were aligned to NGSS learning goals that the curriculum targeted, not to the unit directly. As such, some of the unique features of the unit that are not specifically mentioned in the NGSS were not assessed. For example, the curriculum developers saw transcription and translation as central to understanding the molecular underpinnings of evolution. But because this connection is not explicit in NGSS, it was not assessed. Thus, we do not know what students may have learned beyond what is included in NGSS. An additional limitation to the assessment is that the items were pilot tested along with the curriculum. Thus, some of the assessment items described here were still in draft form. In January of the pilot test year, the evaluators analyzed the alignment between the NGSS learning goals of the assessment items and the NGSS learning goals of the curriculum. Although the teams had developed the goals collaboratively at the onset of the project, results indicated that only a small number of assessment items satisfactory aligned to the learning goals targeted in the Shared Biochemistry module, in addition to other areas of incomplete alignment. This may explain why the Shared Biochemistry module

Homburger *et al. Evo Edu Outreach* (2019) 12:13

Page 16 of 17

did not show statistically significant learning gains at the $p < .05$ threshold. Subsequently, new items were developed and pilot tested to be used in the randomized controlled trial of the curriculum.

Regarding the curriculum, its learning objectives do not include every aspect of HS-LS4, Biological Evolution—namely human impacts on biodiversity (LS4.D). Additionally, the unit includes most of HS-LS3, Inheritance and Variation of Traits, but it excludes the pieces that are not necessary for understanding the connections between heredity and evolution—namely the influence on traits of the environment, the role of regulatory DNA sequences, and environmentally induced mutations. Furthermore, integrating pertinent heredity concepts in a way that supports understanding of core evolution ideas necessitated some re-arranging of concepts contained in the DCIs as outlined by the NGSS. Finally, while we recruited teachers from a diversity of contexts, they are a self-selected group that may not be representative of high school biology teachers as a whole. Participating teachers were open to using a new curriculum, and they were interested in implementing evolution curriculum materials that were NGSS-aligned, that integrated heredity and genetics, or both.

Additional files

Additional file 1. Sample assessment items.

Additional file 2. Histogram of students' percent correct on the pretest and posttest.

Additional file 3. Regression analysis of pre- to posttest gains by student demographic characteristics.

Additional file 4. Table of evolution misconceptions and the percentage of students on the posttest who chose a multiple-choice answer aligned to that misconception.

Authors' contributions
SAH, MM, KP, KB, RP, PA, NB, AH, SK, MK, and HS contributed to the curriculum development. LAS, DD, and KMB contributed to the study's research design. DD and KMB contributed to the collection and analysis of the teacher measures. JH, JR, and GB developed and analyzed the student assessment data. SAH, DD, MM, KMB, LAS, and JH drafted the manuscript. All authors read and approved the final manuscript.

Author details
[1] Genetic Science Learning Center, University of Utah, Salt Lake City, USA. [2] Rockman et al, San Francisco, USA. [3] American Association for the Advancement of Science, Project 2061, Washington, D.C., USA.

References
Achieve, Inc. EQuIP rubric for lessons & units: science, version 2.0. 2016. https://www.nextgenscience.org/sites/default/files/EQuIP%20Rubric%20for%20Science%20v2.pdf.
American Association for the Advancement of Science. Benchmarks for Science Literacy. Oxford University Press, 198 Madison Avenue, New York, NY 10016-4314. 1993.
Asterhan CS, Schwarz BB. The effects of monological and dialogical argumentation on concept learning in evolutionary theory. J Educ Psychol. 2007;99(3):626–39.
Ayala CC, Shavelson RJ, Araceli Ruiz-Primo M, Brandon PR, Yin Y, Furtak EM. From formal embedded assessments to reflective lessons: the development of formative assessment studies. Appl Measur Educ. 2008;21(4):315–34.
Banet E, Ayuso GE. Teaching of biological inheritance and evolution of living beings in secondary school. Int J Sci Educ. 2003;25(3):373–407.
Barnes ME, Evans EM, Hazel A, Brownell SE, Nesse RM. Teleological reasoning, not acceptance of evolution, impacts students' ability to learn natural selection. Evolution. 2017;10:1–12.
Beardsley PM, Stuhlsatz MA, Kruse RA, Eckstrand IA, Gordon SD, Odenwald WF. Evolution and medicine: an inquiry-based high school curriculum supplement. Evolution. 2011;4(4):603–12.
Bell P, Linn MC. Scientific arguments as learning artifacts: designing for learning from the web with KIE. Int J Sci Educ. 2000;22:797–817.
Berland LK, McNeill KL. A learning progression for scientific argumentation: understanding student work and designing supportive instructional contexts. Sci Educ. 2010;94(5):765–93.
Biggs A, Hagins WC, Holliday WG, Kapicka CL, Lundgren L, MacKenzie AH. Glencoe biology national geographic. Columbus: McGraw-Hill Companies Inc; 2009.
Bishop BA, Anderson CW. Student conceptions of natural selection and its role in evolution. J Res Sci Teach. 1990;27(5):415–27.
Black P, Wiliam D. Assessment and classroom learning. Assess Educ. 1998;5:7–74.
Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. London: Psychology Press; 2013.
Borgerding LA, Klein VA, Ghosh R, Eibel A. Student teachers' approaches to teaching biological evolution. J Sci Teacher Educ. 2015;26(4):371–92.
Bray E, Long TM, Pennock RT, Ebert-May D. Using Avida-ED for teaching and learning about evolution in undergraduate introductory biology courses. Evolution. 2009;2(3):415–28.

Homburger *et al. Evo Edu Outreach*   (2019) 12:13

Page 17 of 17

Brewer M, Gardner G. Teaching evolution through the Hardy–Weinberg principle: a real-time, active-learning exercise using classroom response devices. Am Biol Teach. 2013;75(7):476–9.

Brown J, Collins A, Duguid P. Situated cognition and the culture of learning. Educ Res. 1989;18(4):32–42.

Catley KM, Lehrer R, Reiser B. Tracing a prospective learning progression for developing understanding of evolution. National Academies Committee on Test Design for K-12 Science Achievement. 2004.

Catley KM, Novick LR, Shade CK. Interpreting evolutionary diagrams: when topology and process conflict. J Res Sci Teach. 2010;1:22.

Coffey JE, Hammer D, Levin DM, Grant T. The missing disciplinary substance of formative assessment. J Res Sci Teach. 2011;48(10):1109–36.

Dobzhansky T. Nothing in biology makes sense except in the light of evolution. Am Biol Teach. 1973;35:125–9.

Dougherty MJ. Closing the gap: inverting the genetics curriculum to ensure an informed public. Am J Hum Genet. 2009;85:6–12.

Driver R. Constructivist approaches to science teaching. In: Steffe LP, Gale J, editors. Constructivism in education. Hillsdale: Lawrence Erlbaum Associates; 1995. p. 385–500.

Duncan RG, Rogat AD, Yarden A. A learning progression for deepening students' understandings of modern genetics across the 5th–10th grades. J Res Sci Teach. 2009;46(6):655–74.

Flanagan JC, Roseman J. Assessing middle and high school students' understanding of evolution with standards-based items. The National Association for Research in Science Teaching. Orlando; 2011.

Furtak EM. Formative assessment for secondary science teachers. Thousand Oaks: Corwin Press; 2009.

Furtak EM, Kiemer K, Circi RK, Swanson R, de León V, Morrison D. Teachers' formative assessment abilities and their relationship to student learning: findings from a four-year intervention study. Instr Sci. 2016;44(3):267–91.

Genetic Science Learning Center. Evolution: DNA and the unity of life (student site). 2018a. https://learn.genetics.utah.edu/content/evolution/. Accessed 9 May 2019.

Genetic Science Learning Center. Evolution: DNA and the unity of life (teacher site). 2018b. https://teach.genetics.utah.edu/content/evolution. Accessed 9 May 2019.

Geraedts CL, Boersma KT. Reinventing natural selection. Int J Sci Educ. 2006;28(8):843–70.

Glaze AL, Goldston MJUS. Science teaching and learning of evolution: a critical review of the literature 2000–2014. Sci Educ. 2015;99:500–18.

Gould SJ. The structure of evolutionary theory. Cambridge: Harvard University Press; 2002.

Gregory TR. Understanding natural selection: essential concepts and common misconceptions. Evolution. 2009;2:156–75.

Heil CS, Manzano-Winkler B, Hunter J, Noor JK, Noor MA. Witnessing evolution first hand: a K–12 laboratory exercise in genetics and evolution using Drosophila. Am Biol Teach. 2013;75(2):116–9.

Hopson JL, Postlethwait J. Modern biology. Austin: Holt; 2009.

Institute of Education Sciences, U.S. Department of Education, National Science Foundation. Common guidelines for education research and development: a report from the institute of education sciences, U.S. Department of Education, and the National Science Foundation. 2013.

Kalinowski ST, Leonard MJ, Andrews TM. Nothing in evolution makes sense except in the light of DNA. CBE-Life Sciences Education. 2010;9:87–97.

Kang H, Thompson J, Windschitl M. Creating opportunities for students to show what they know: the role of scaffolding in assessment tasks. Sci Educ. 2014;98(4):674–704.

Kingston N, Nash B. Formative assessment: a meta-analysis and a call for research. Educ Meas. 2011;30(4):28–37.

Krajcik J, Codere S, Dahsah C, Bayer R, Mun K. Planning instruction to meet the intent of the Next Generation Science Standards. J Sci Teach Educ. 2014;1:1. https://doi.org/10.1007/s10972-014-9383-2.

Kuhn D. Thinking together and alone. Educ Res. 2015;44(1):46–53.

Linacre JM. Detecting multidimensionality: which residual data-type works best? J Outcome Meas. 1998;2:266–83.

Linacre JM. Winsteps® Rasch measurement computer program. Beaverton: Winsteps.com; 2016.

Mead R, Hejmadi M, Hurst LD. Teaching genetics prior to teaching evolution improves evolution understanding but not acceptance. PLoS Biol. 2017;15(5):2002255.

Meir E, Perry J, Herron JC, Kingsolver J. College student's misconceptions about evolutionary trees. Am Biol Teach. 2007;69(7):71–6.

Miles MA, Huberman AM. Qualitative data analysis: an expanded sourcebook. New York: Sage; 1994.

Miller KR, Levine J. Prentice Hall biology. Boston: Pearson Education, Inc.; 2008.

Morabito NP, Catley KM, Novick LR. Reasoning about evolutionary history: post-secondary students' knowledge of most recent common ancestry and homoplasy. J Biol Educ. 2010;44(4):166–74.

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

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.

Nehm RH. EvoGrader. http://evograder.org/index.jsp. Accessed 2018.

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.

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

Osborne J. Arguing to learn in science: the role of collaborative, critical discourse. Science. 2010;328(5977):463–6.

Stake RE. Qualitative case studies. The Sage handbook of qualitative research. 3rd ed. Thousand Oaks: Sage Publications Ltd; 2005. p. 443–66.

Stark LA, Barber NC, Bass KM, Malone MA, Roseman JE. Designing a new NGSS-aligned high school biology curriculum unit and associated teacher professional development. Washington: Paper presented at the American Education Research Association Annual Meeting; 2016.

Strike KA, Posner GJ. A revisionist theory of conceptual change. In: Duschl RA, Hamilton RJ, editors. Albany. New York: State University of New York Press; 1992.

Toulmin S. The uses of argument. Cambridge: Cambridge University Press; 1958.

Ziadie MA, Andrews TC. Moving evolution education forward: a systematic analysis of literature to identify gaps in collective knowledge for teaching. CBE-Life Sci Educ. 2018;17(1):10–1187.

Zohar A, Nemet F. Fostering students' knowledge and argumentation skills through dilemmas in human genetics. J Res Sci Teach. 2001;39(1):35–62.

Publisher's Note