

Positioning Genomics in Biology Education: Content Mapping of Undergraduate Biology Textbooks †

Naomi L. B. Wernick^{1, 2 *}, Eric Ndung'u³, Dominique Haughton³, and Fred D. Ledley¹

¹Department of Natural and Applied Sciences, Bentley University, Waltham, MA 02452, ²Department of Biological Sciences, University of Massachusetts Lowell, MA 01854, ³Department of Mathematical Sciences, Bentley University, Waltham, MA 02452

Biological thought increasingly recognizes the centrality of the genome in constituting and regulating processes ranging from cellular systems to ecology and evolution. In this paper, we ask whether genomics is similarly positioned as a core concept in the instructional sequence for undergraduate biology. Using quantitative methods, we analyzed the order in which core biological concepts were introduced in textbooks for first-year general and human biology. Statistical analysis was performed using self-organizing map algorithms and conventional methods to identify clusters of terms and their relative position in the books. General biology textbooks for both majors and nonmajors introduced genome-related content after text related to cell biology and biological chemistry, but before content describing higher-order biological processes. However, human biology textbooks most often introduced genomic content near the end of the books. These results suggest that genomics is not yet positioned as a core concept in commonly used textbooks for first-year biology and raises questions about whether such textbooks, or courses based on the outline of these textbooks, provide an appropriate foundation for understanding contemporary biological science.

INTRODUCTION

The emergence of genomics over the past 20 years has led to dramatic discoveries in biology and fundamentally changed our way of understanding biological systems. Biological thought is increasingly organized around the role of the genome in creating and regulating complex biological systems; explaining the diversity of life, evolution, and ecology; maintaining health; and contributing to disease. It has become common to describe this era of biological investigation and understanding as “post-genomic,” reflecting the fundamental changes afforded by the genomics paradigm. Thus, “information flow, exchange, and storage” of genetic information, topics synonymous with genomics, are identified as core concepts in Vision and Change (1). In this paper, we consider whether the emerging understanding of genomics as a core principle of biological understanding is reflected in the positioning of genomics within introductory, undergraduate biology curricula.

It has been argued that in the post-genomic era, the genome occupies the same position in biology as the periodic table in chemistry (10, 13, 21). In this context, understanding the genome becomes a predicate to understanding biological systems at the level of cells, organs, organisms, and even ecosystems in the same way that the periodic table paves the way to understanding molecules, chemical reactions, and materials.

In an educational context, the order of instruction impacts student learning. The constructivist approach to learning, for example, posits that learning builds not only on prior knowledge, but also on broad cognitive frameworks in which new concepts and observations can be anchored (4). Furthermore, the connections that students form between different concepts are fundamental to their learning and represent a key difference between meaningful and rote learning (27). These notions are supported by scholarship demonstrating that the sequence of instruction in core scientific concepts, for example, in the form of learning progressions in K–12, is important for the design of effective curriculum and for students’ abilities to understand more advanced concepts (7, 12, 29, 34). In this view, a student may achieve a subjectively different understanding of development, homeostasis, evolution, or health coming to these topics with a core foundation in areas other than genomics.

Recognizing the foundational role of genomics in our contemporary understanding of biology and medicine, and

*Corresponding author. Mailing address: Department of Biological Sciences, University of Massachusetts Lowell, One University Ave., Lowell, MA 01854. Tel: 978-934-3506. Fax: 978-934-3044. E-mail: Naomi_Wernick@uml.edu.

†Supplemental materials available at <http://jmbe.asm.org>

how the sequence of instruction in core scientific concepts can influence student learning of advanced science topics, this paper explores the positioning of genomics in undergraduate biology education. We ask whether genomics is positioned as a core concept in the teaching of undergraduate biology. We approached this question by identifying and analyzing the order in which key biological concepts are first introduced in textbooks for use in college-level biology courses.

There is extensive literature describing text analysis of biology textbooks, and many different methods have been used. For example, text analysis was used to ascertain whether evolution is presented as a unifying theme in instruction on advanced biological topics by analyzing the number of words written on topics related to evolution (30, 31, 32, 33). Other works mapped the location of text related to evolution within textbooks, asking whether these words are more prevalent toward the front or back of the textbook (31), and whether they are used in chapters on different subjects that might reference evolution as a unifying principle (26, 33). The observation that text describing evolution is most often associated with chapters on evolution or diversity, and is less often included in chapters on other topics, suggested that textbooks are not recurrently referring to evolution as a unifying theme of biology (26). The methods used in these works involved determining the frequency and distribution of specific terms throughout the text (26, 30, 31, 32, 33).

Most recently, text analysis of high school biology textbooks has been performed as part of AAAS Project 2061 (29). This study acknowledges the important role of textbooks in guiding students toward making proper connections and forming deeper understandings within the context of a greater scientific picture. While its approach of deeply assessing one particular topic and all the connections between the underlying concepts is different from our own, the overarching premise resembles our belief that essential concepts must be positioned to serve as a fundamental organizing principle.

The question we ask in this work is not whether diverse topics in biology are explicitly explained with reference to genomics, but whether genomic concepts are introduced early enough in the undergraduate biology curriculum for students to use them for understanding other biological concepts. To do this, we analyzed the content of 25 widely used textbooks for introductory undergraduate biology. The analysis was based on a lexicon of metaterms representing core biological concepts and the locations of where these concepts were first introduced in sufficient detail to serve as a foundation for subsequent learning in textbooks. We found that general biology textbooks designed for both majors and nonmajors introduce concepts related to genomics after text introducing cell structure and function as well as biological chemistry, but before text introducing higher-order biological processes. In contrast, genome-related metaterms in human biology textbooks were often positioned near the end of these books. This analysis raises questions about the

positioning of genomics in “introductory” undergraduate biology courses and whether such courses provide appropriate foundations for learning contemporary biological science.

METHODS

Textbooks

Textbooks for courses listed as general biology or human biology, for majors or nonmajors, were identified on the websites of major publishing companies including Brooks/Cole Cengage Learning, Bartlett Learning, Benjamin Cummings-Pearson Education, McGraw Hill, W.H. Freeman, and Wiley. Books were categorized as majors general biology (MGB), nonmajors general biology (NMGB), or human biology (HB) according to the publisher’s designation. Several publishers offered multiple titles by the same authors with much overlapping content. To avoid statistical bias from over-representation of multiple books by one author, texts were eliminated within each category that had common authors and analogous content, even if they had different titles. All books used were the most recent editions available when this study was initiated and had publication dates after 2004. Twenty-five books (Appendix 1) were analyzed, which included 7 human biology, 7 majors general biology and 11 nonmajors general biology textbooks.

Metaterms and data collection

A lexicon of 79 metaterms was generated representing core biological concepts (Fig. 1A). The lexicon was populated by first collecting chapter titles from each book. A standard nomenclature for the chapter titles was developed, and a list of key concepts associated with each chapter title was generated from chapter subheadings and expert knowledge of the field. A final set of 79 metaterms was selected from the list of chapter titles and associated concepts. While microbial and plant genomics are crucial to understanding biological systems, these terms do not appear in human biology textbooks and were eliminated to enable statistical comparison of textbooks for general biology with those for human biology.

We define the “location” or “position” of a metaterm in a book as the page number where related concepts are introduced in sufficient detail to serve as a foundation for future learning. Page numbers were first identified in the index and the text at that location was reviewed. When the term was discussed in multiple locations within the book, the designated page number was the page where the content associated with that metaterm was first discussed in detail. When metaterms did not appear in the index, synonyms or subsumed terms, based on the original chapter titles, subtitles, or subsections as well as expert knowledge of the field, were used to identify pages where related content was introduced in the text. When the location of content could not be determined directly from the metaterms, a senior author reviewed the book to identify locations where related

content was introduced. In some instances, no related content could be identified. All page assignments were confirmed by at least two authors of this paper.

Statistical analysis and Kohonen Self-Organizing Maps

Clusters were identified using Kohonen Self-Organizing Maps (SOM), an application of a competitive neural network algorithm (16, 18, 19). Algorithms and applications of SOM methods have been described elsewhere (9, 17). Analysis was performed using a standard statistical package, MATLAB 6.1.

Briefly, SOM provides a non-biased, statistical method for quantifying the pairwise distances between each of the 79 metaterms in the 25 textbooks and visualizing these distances on a two-dimensional lattice. The SOM algorithm considers the position of the 79 metaterms within each book to constitute a unique vector in 79 dimensions. Each of the 25 textbooks constitutes a separate 79 dimensional vector. The SOM method involves calculation of a non-linear data projection of these vectors onto a two-dimensional lattice with independent weighting of each of the 79 terms. This is done through an iterative, neural networking process during which each vector is forwarded from a single input node to possible nodes in the output layer, and a weighted scoring function is calculated that minimizes the (Euclidean) distances traveled by the vectors. The output node that provides the minimal calculated distance becomes the new input node and the center of a neighborhood of vectors whose weighted distances are adjusted through subsequent iterations (8, 18, 19, 22). The algorithm converges, achieving maximum parsimony, when little or no change occurs in the weight vector between sequential iterations, at which point each of the 79 metaterms can be mapped to a specific node on the two-dimensional lattice based on the weighting function. The results are visualized in a classic Kohonen map (Fig. 1B) in which the 79 dimensional metaterms and their nodal positions are represented by a hexagonal lattice where individual nodes are represented by hexagons separated by intervening hexagons whose color reflects the distance between the adjacent nodes.

In this representation, the metaterms that are most often closely associated are positioned within a single node. The nodes are positioned in the Kohonen map such that the distance between the metaterms in the set of books examined is represented by the geometric distance between nodes and also by the color of the intervening hexagons (16). Thus, nodes on opposite sides of the Kohonen map are most distantly associated in the textbooks, while those that are in adjacent regions are more closely associated. In addition, the calculated distance between the metaterms contained in adjacent nodes is indicated by the color of the intervening hexagon using a rainbow scale in which red hexagons indicate a greater distance and blue hexagons indicate the least distance. In this analysis, the clustering of concepts in the textbooks is reflected by the juxtaposition of nodes

within the lattice connected by hexagons on the blue end of the spectrum. The boundaries of a cluster can be recognized by hexagons on the red end of the spectrum, which indicate greater distance from nodes elsewhere in the lattice.

Textbook usage data

BOWKER'S PUBTRACK database of textbook adoption in introductory biology courses was purchased from the publisher (3). These data describe textbooks used by over 11 million students, representing ~60% of the ~14.5 million students enrolled in two- or four-year colleges or universities in the US in 2005 (25). Data on textbooks for courses in human, majors, nonmajors, and "all others" were included in the analysis. For this analysis, a list of "unique books" was generated to eliminate redundancies in the database. Each "unique book" combined all of the editions of a text with the same title, editions by the same authors and within the same category (MGB, NMGB, or HB) that have different titles (often expanded editions that include specialized chapters), and these texts sold as packages with supplementary material or laboratory manuals, which appear separately in the BOWKER'S PUBTRACK database. We did not analyze laboratory manuals.

RESULTS

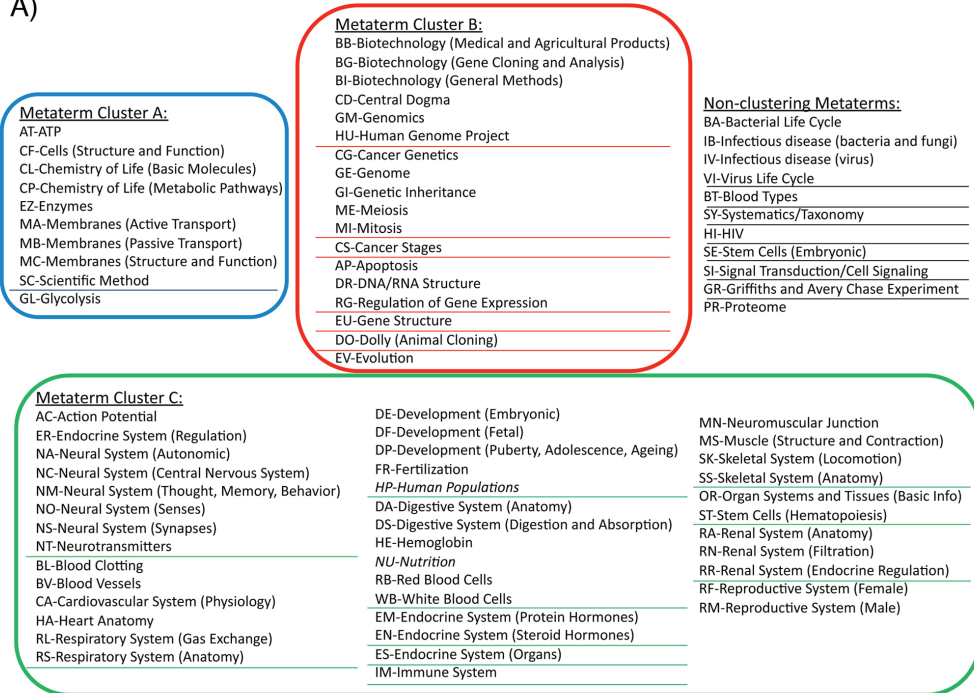
Seventy-nine metaterms (Fig. 1A) were selected. These terms describe content in general biology, human biology, and genomics. While plant and microbial biology is a central component of general biology education, these topics are not covered in human biology textbooks and thus were excluded from our statistical analysis. Twenty-three of the textbooks (92%) contained at least 90% of the metaterms. All 25 textbooks contained at least 55% (44 of 79) of the metaterms.

The average length of the 25 books was 817 ± 307 (SD) pages. For the 11 NMGB textbooks, the average length was 702 ± 145 (SD) pages. For HB textbooks, the average length was 549 ± 51 (SD) pages. For MGB textbooks, the average length was 1266 ± 23 (SD) pages.

Clustering of metaterms

Statistical analysis was performed to compare the positioning of metaterms (represented as page numbers denoting where content represented by that metaterm was first sufficiently described to serve as a foundation for future learning) in 25 textbooks for undergraduate biology. SOM algorithms were used to identify clustering of concepts based on the location of metaterms within the 25 books. Three clusters were recognized as regions of blue hexagons, indicating relatively little distance between the location of metaterms in adjacent hexagons, separated by "walls" of yellow or red hexagons indicating greater distance between the locations of metaterms in adjacent hexagons (Fig. 1B). The metaterms associated with the three clusters

A)



B)

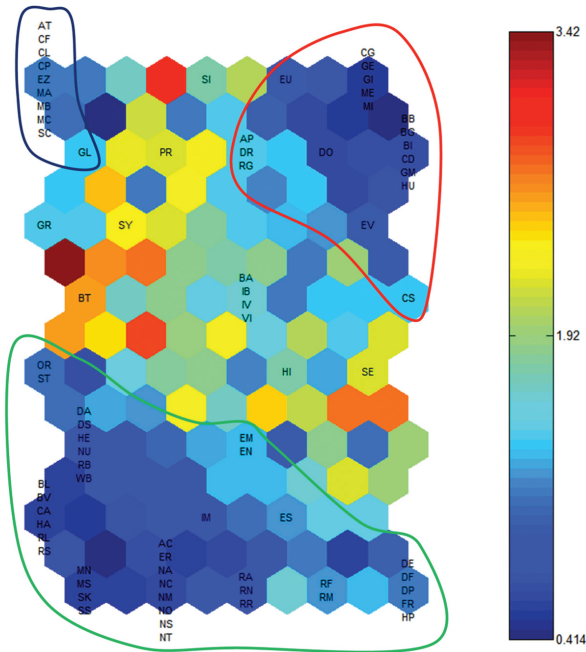


FIGURE 1. Kohonen map showing clustering of 79 metaterms in 25 introductory biology textbooks. A) This set of 79 metaterms represents concepts commonly found in first courses of undergraduate biology, with an emphasis on terms related to human biology and genomics. Three clusters of metaterms were identified through SOM analysis of 25 introductory biology textbooks. Metaterms circled in blue (Cluster A) include mostly introductory content. Metaterms circled in red (Cluster B) include mostly genome-related content. Metaterms circled in green (Cluster C) include content describing organ systems. Lines within each cluster are included to denote metaterms that mapped to distinct nodes. Terms that did not map within any of the three major clusters are indicated as non-clustering metaterms. The two-letter code for each metaterm is included (Fig. 1B). B) This Kohonen map consists of alternating hexagons denoting nodes that represent individual or groups of metaterms, as well as intermediate hexagons that represent the distance between nodes. The color of each hexagon indicates the distance between adjacent nodes, with blue indicating close proximity and red indicating maximal distance. Metaterms, indicated by their two-letter code, are positioned by the SOM analysis so as to minimize the distances between terms in a projection of the 79 dimensional vectors representing each book onto the two dimensional lattice. Three clusters of closely linked nodes are highlighted. Those circled in blue (Cluster A) include metaterms associated with introductory content. Those circled in red (Cluster B) include genome-related content. Those circled in green (Cluster C) consist of content describing organ systems. The full designation of each metaterm is shown in Figure 1A.

are listed together in Figure 1A. Within each cluster, those metaterms associated with the same node, indicating the greatest proximity, are grouped in sections separated by lines (Fig. 1A).

Cluster A (Fig. 1, blue circle) is comprised of metaterms related to introductory content, including: scientific method, cells, membrane functions, transport, basic chemistry, chemistry of life, and energy metabolism. Cluster B (Fig. 1, red circle) was revealed to include mostly genome-related metaterms, including: Central Dogma, DNA, gene structure, regulation of gene expression, and biotechnology as well as mitosis, meiosis, inheritance, cancer genetics, and evolution. Cluster C (Fig. 1, green circle) is comprised predominantly of metaterms related to various organ systems. Some, listed as “non-clustering metaterms” (Fig. 1), did not map to clusters; these include: infectious diseases, taxonomy, stem cells, and signal transduction.

Distribution of metaterms

Figure 2 shows the relative location of metaterms within the three categories of books and plots the rank order of metaterms against their average relative position within each category. A relatively uniform distribution is observed in HB books, as evidenced by the slope of the curve in this visualization. In MGB books, there is increased spacing between the locations of metaterms in the central core of the book, reflecting the absence of metaterms associated with plant and microbial biology. In NMGB books, the increased spacing between the locations of metaterms is still evident, but is less pronounced than for MGB books. When the location of metaterms was ranked in order of average position within all textbooks, only 5 of 78 adjacent metaterms were separated by greater than 5% of the total length of the book. This occurred in sections of the book that introduced metaterms related to non-mammalian species, plants, and microorganisms, and in chapters with proportionally greater detail and illustrations. Two metaterms, “proteome” and “Griffiths or Avery Chase experiments,” were missing from 13 of 25 textbooks (52%) and 12 of 25 textbooks (48%), respectively.

Positioning of metaterms

To compare the order of metaterms in MGB and NMGB books, a regression was performed between the average position of each metaterm in MGB books compared to NMGB books (Fig. 3A). While some local variation was observed, a very high correlation was found ($R^2 = 0.93$, $p < 0.01$) between the positioning of metaterms in MGB and NMGB books. In contrast, the average position of metaterms in HB compared to MGB books (Fig. 3B) shows that there is no statistical correlation between the position of metaterms in these books ($R^2 = 0.01$, $p = \text{NS}$). Two distinctly non-linear clusters of metaterms are visible in this regression. The first, reflecting terms that appear early in MGB books and late in HB books (circled in red), was comprised predominantly

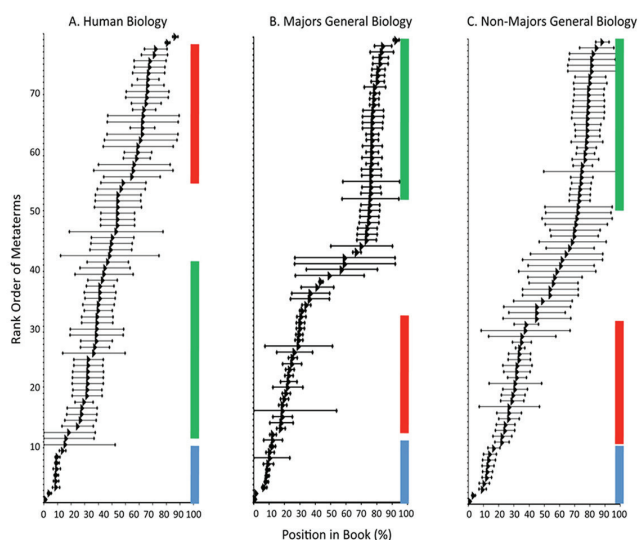


FIGURE 2. Position of metaterms in textbooks for first courses in undergraduate biology. A) Position of metaterms in textbooks for human biology (HB). B) Position of metaterms in textbooks for majors general biology (MGB). C) Position of metaterms in textbooks for nonmajors general biology (NMGB). In each panel, the position of each metaterm is given as the average page number where content associated with that metaterm is introduced divided by the total number of pages in the book. Rank order is determined based on ordering of average metaterm positions. Blue bars, red bars and green bars indicate the position of cluster A, cluster B, and cluster C, respectively. In panel B (majors general biology) and C (nonmajors general biology) bars cover greater than 95% of metaterms in the cluster. In panel A, the green and red bars cover 70% of metaterms in those clusters.

of metaterms associated with cluster B (genome-related material). The second, reflecting terms that appear early in HB books and late in MGB books (circled in green), is comprised predominantly of terms associated with cluster C (organ systems).

The distinct position of metaterms in HB textbooks compared to MGB and NMGB textbooks is illustrated in a color-coded ordering of metaterms (Fig. 2). The blue bars indicate the position of metaterms associated with cluster A. The red bars indicate the position of metaterms associated with cluster B in MGB and NMGB books, and 70% of these terms in HB books. The green bar indicates the position of metaterms associated with cluster C in MGB and NMGB books, and 70% of these terms in HB books.

We had previously observed significant differences in the lengths of MGB and NMGB books (average length: MGB = 1266 pages, NMGB = 702 pages). To examine how differences in the amount of content were distributed through these texts, the mean and standard deviation of the absolute page numbers for metaterms in NMGB textbooks (Fig. 4, grey circles) were plotted against the rank order of the metaterms determined for MGB textbooks (Fig. 4, black squares). In this analysis, the number of pages dedicated to each metaterm and its surrounding content is visualized as

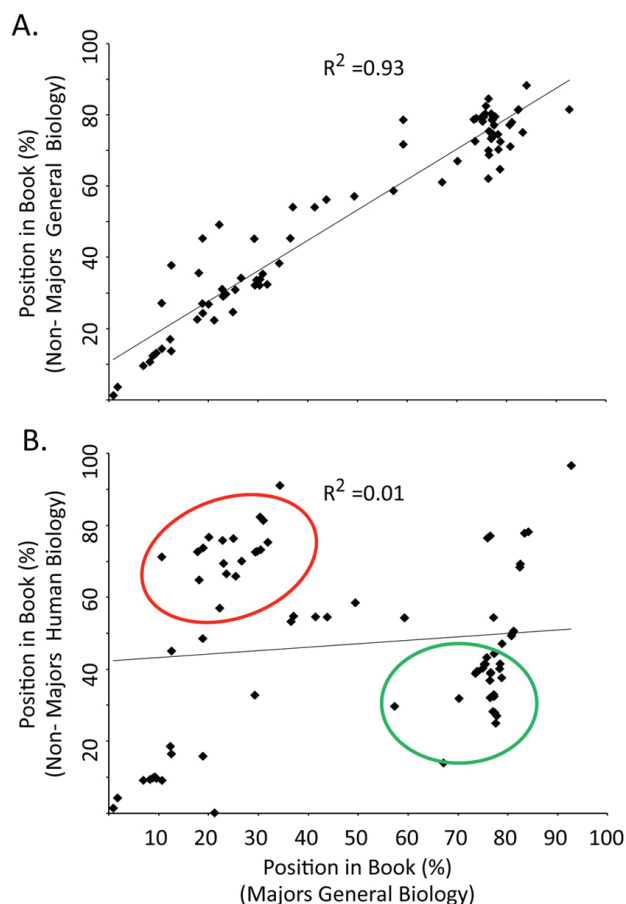


FIGURE 3. Correlation of metaterm positions in three categories of textbooks. A) Regression analysis comparing positioning of metaterms in MGB and NMGB textbooks. B) Regression analysis comparing positioning of metaterms in MGB and HB books. Two nonlinear clusters of metaterms highlighted (panel B) correspond to metaterms associated with cluster B (red circle) and metaterms associated with cluster C (green circle).

the slope of the curve. Relatively little difference is seen in the amount of content dedicated to introductory topics (Fig. 4, cluster A, blue bar). However, differences in the amount of content are seen in the slopes of the lines through metaterms comprising cluster B (Fig. 4, red bar), although this contributed only a small fraction to the variation of the books' overall lengths. No significant differences are seen in the slopes of the lines through metaterms comprising cluster C (Fig. 4, green bar). The greatest differences were observed for content describing plant and microbial life; these topics map to the central sections of the book and cover almost twice as much content in the MGB books as in the NMGB books.

Finally, our set of textbooks contained two pairs of books by identical authors in different categories. To compare the organization of these books by the same authors, regressions were performed between the locations of metaterms in each (Fig. 5). One set by the same author

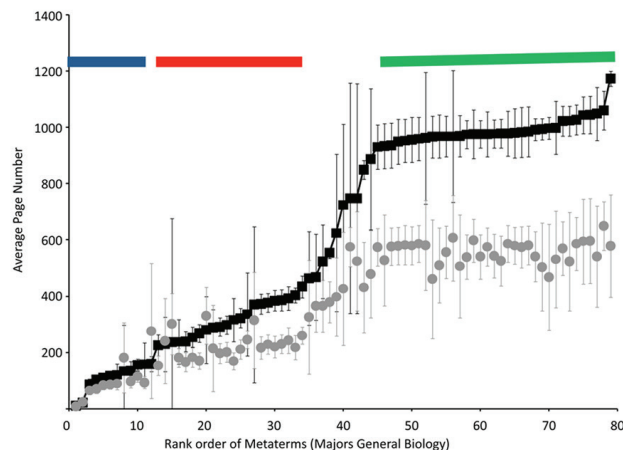


FIGURE 4. Page positions of metaterms in general biology textbooks for majors and nonmajors. The average and standard deviation of page numbers for metaterms in MGB and NMGB books were plotted against the rank order position of these terms in MGB books. The red bar indicating the position of cluster B metaterms and the green bar indicating the position of cluster C metaterms each cover 90% of the metaterms associated with those clusters (Fig. 1).

included a MGB and NMGB textbook. A very high degree of correlation was observed ($R^2 = 0.88$, $p < 0.01$) between these books, with most of the variance representing a repositioning of a small number of metaterms (Fig. 5B). In contrast, the second set of books by the same author included a HB and NMGB textbook. No correlation was observed ($R^2 = 0.004$, $p = \text{not significant}$) between these books, indicating that the order of metaterms had been rearranged (Fig. 5A).

Textbooks in introductory biology courses

To assess whether textbooks in our analysis are representative of those used in undergraduate instruction, we purchased the BOWKER'S PUBTRACK database describing the textbooks purchased for introductory biology and human biology courses in the US. A total of 77 different textbooks were identified in this database. These books were required in 61,000 course sections serving ~2.7 million students. The 25 "unique books" in this study, including all editions of the texts examined, were used in 29,000 course sections enrolling 1.4 million students, representing 47% of all course sections and 50% of all students in BOWKER'S PUBTRACK database. Significantly, no other texts in the BOWKER'S PUBTRACK database were used in more than 1% of course sections or by 2% of students.

DISCUSSION

This work is predicated on educational research suggesting that the ordered progression of concepts plays an important role in the construction of student learning. Studies emphasize

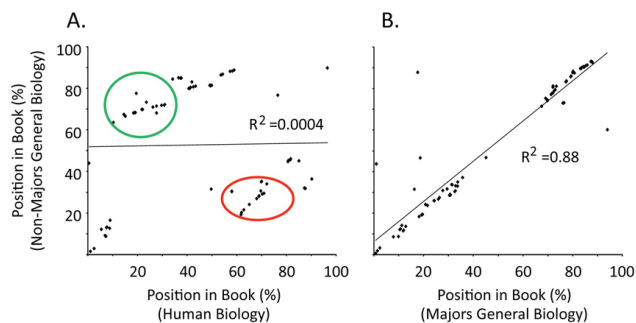


FIGURE 5. Correlation of metaterm positions in pairs of books by the same author for different courses. A) Regression analysis comparing positioning of metaterms in HB and NMGB books by the same author. B) Regression analysis comparing positioning of metaterms in MGB and NMGB books by the same author. Two highlighted nonlinear clusters of metaterms (panel A) correspond to metaterms associated with cluster B (red circle) and metaterms associated with cluster C (green circle).

that learning scientific concepts is enhanced by connecting content in a way that enables students to construct an understanding of these concepts from their knowledge of core material. This not only refocuses learning on critical thinking, as opposed to rote memorization, but also represents science in a way that is congruent with the process of scientific understanding and inquiry. Moreover, this approach provides a functional foundation for future learning in both formal and informal settings. The maps created by AAAS Project 2061 in the Atlas of Science Literacy, the literature on the use of concept maps (27, 23), and the literature on learning progressions (7, 12, 28, 34) are important examples of this perspective and how it can link to student development, instruction, assessment, and the process of science itself.

Our focus, however, is not on the construction or efficacy of learning progressions. Rather, we sought to describe the instructional sequence of concepts in the organization of textbooks for first courses in undergraduate biology and, by inference, the order in which concepts are presented in the first courses of undergraduate biology. For majors in biology and related disciplines, the first college biology course is often an “introductory” course that provides foundational knowledge. However, for students who do not major in biology, this course often represents their final academic exposure to biology (20, 24). For these students, this course has a particularly important role in establishing the foundation for students’ ability to use biological principles in their personal and public lives, and, potentially, for lifelong learning. We were specifically interested in the positioning of genomics in these textbooks and whether the increasingly central role of genomics in biological understanding, biomedical research, and civic discourse (6) was reflected in the order of contemporary undergraduate curricula. Our results suggest that genomics is not positioned as a core concept in undergraduate biology education. All books in

this study began with introductory content focused on the chemistry of life, cell structure and function, and membrane structure and function followed closely by content on energy metabolism. In textbooks for majors and nonmajors general biology courses (MGB and NMGB), this introductory content was followed by genome-related content (Fig. 4). Our analysis shows that in both MGB and NMGB textbooks (Figs. 2B and C), there was relatively little distance between introductory content and genome-related content. This sequence is congruous with the learning progression proposed by Roseman et al. (28) for understanding the role of DNA in determining the characteristics of an organism.

In contrast to the organization of MGB and NMGB books, there was a great distance between introductory and genome-related content in HB books, and most genome-related metaterms were located in the last sections. By locating genome-related content near the end, genomics is not positioned as core content upon which students can construct an understanding of other biological concepts. This is notably different than the positioning of genome-related content near introductory material in general biology textbooks; it also diverges from genome-based paradigms that are increasingly central to biological understanding and inquiry.

This divergent instructional sequence in human biology textbooks could profoundly impact the teaching of biological science. The central concept of genomics is that the information necessary to constitute the structures and functions of a living system resides in the sequence of the genome, and that the complex characteristics of biological systems arise from gene expression and the integrated functions of gene products. In this perspective, genomics is a predicate for biological understanding and inquiry. This progression, not observed in any of the three classes of books studied, is more closely aligned with the sequences observed in general biology books, where genome-related content closely follows content on basic biological structures and functions.

In human biology books, genomic content appears almost as an epiphenomenon to reproduction and development near the end of textbooks. This positioning does not convey the centrality of genomics in emerging biological research or medical advances such as personalized medicine. In fact, the positioning of genomics in the latter sections of lengthy introductory books raises the important question of whether one-semester courses are providing any instruction in genomics at all.

In considering the implications of these data, we note that the books examined in this study were required or recommended texts in more than 60,000 sections of introductory or human biology in 2010, with an annual enrollment of over 2.5 million students. While we are unaware of contemporary data comparing the order of classroom instruction with the content of textbooks, historically, textbooks have had an important role in undergraduate biology education. A 1980 report, still quoted in contemporary publications, states that “textbooks exert an overwhelming

dominance over the science learning experience.” That study found that “90 to 95 percent of 12,000 teachers surveyed indicated they used texts 90 percent of the time” (15). Other authors suggested “college biology textbooks have a great impact on the discipline... Biologists of the future are encouraged into the field in part by their interaction with the introductory textbook” (2). In fact, many studies suggested that textbooks could impact the curriculum (5, 14, 15) and student learning (11, 15). While we cannot rule out the possibility that classroom teachers use this content in a different order, interviews with a number of the senior authors of these textbooks indicate that the outlines and content of their books are explicitly designed to conform to prevailing classroom practice. Several pointed to the practice of having in-service teachers review textbook outlines and chapters as they are written as providing not only peer-review, but also information on classroom practice that is used to ensure that the text aligns with current practice to maximize sales.

Finally, we would note that there is currently no evidence that positioning of genomics late in the instructional sequence of introductory biology negatively affects understanding of the central role of genomics in biology, or that earlier positioning of genomics would promote a deeper understanding of other biological concepts or emerging biological research. Such inferences need to be formally validated through a rigorous assessment of learning outcomes from courses that follow the instructional sequence observed in the present study compared to a reorganized, genome-based curriculum.

We are currently assessing a novel course design for nonmajors, which begins with content regarding the structure and function of the genome. This course teaches that information necessary to constitute the structures and functions of living systems reside in the genome of each organism, and that the characteristics of biological systems arise from the patterns of gene expression, the integrated functions of expressed gene products, and their intimate relationship with the environment. Our preliminary observations suggest that the concept of the genome as information and mechanisms for regulating gene expression are challenging for students, whose previous courses have focused on more descriptive and phenomenological instruction.

We believe the time is ripe for more concerted research on how best to integrate the genomic paradigm into the biology curriculum and the perspectives of our students. Current scholarship is less focused on traditional textbooks than on the use of information technologies and the internet to create a more dynamic learning experience. Extensive content has been developed for undergraduate biology instruction ranging from teaching modules from institutions such as Cold Spring Harbor and the Howard Hughes Medical Institute, to the digital content accompanying traditional textbooks and the open courseware available from MIT. Much of this content has been developed by leading investigators and institutions in genomics, and it

might be predicted that these resources may offer a more genome-based perspective on biology. The analytical tools used in this report are well-suited not only for characterizing structured datasets such as textbooks, but also for quantitative analysis of the “big-data” sets represented by these complex resources.

SUPPLEMENTAL MATERIALS

Appendix I: Textbooks included in this study

ACKNOWLEDGMENTS

This work was supported, in part, by the ACTA Foundation and the National Biomedical Research Foundation. Dr. Ledley is a Scholar of the BioScience Education Network (BEN), a program funded by the National Science Foundation (DUE 0532797). The authors thank the authors of some of the books analyzed in this study including Marielle Hoefnagels, Scott Freeman, George Johnson, Judith Goodenough, and anonymous authors who agreed to be interviewed about the process of writing biology textbooks. The authors declare that there are no conflicts of interest.

REFERENCES

1. **American Association for the Advancement of Science.** 2011. Vision and change in undergraduate biology education: a call to action: a summary of recommendations made at a national conference organized by the American Association for the Advancement of Science, July 15–17, 2009. Washington, DC.
2. **Blystone, R. V.** 1987. College introductory biology textbooks: an important communicative tool. *Amer. Biol. Teach.* **49(7)**:418–425.
3. **Bowker.** 2010. [Online.] <http://pubtrackonline.com/products/higher.asp?loc1=education>.
4. **Cakir, M.** 2008. Constructivist approaches to learning in science and their implications for science pedagogy: a literature. *Int. J. Environ. Sci. Educ.* **3(3)**:193–206.
5. **Chiappetta, E. L., G. H. Sethna, and D. A. Fillman.** 1993. Do middle school life science textbooks provide a balance of scientific literacy themes? *J. Res. Sci. Teach.* **30(7)**:787–797.
6. **Collins, F. S., E. D. Green, A. E. Guttmacher, and M. S. Guyer, on behalf of the US National Human Genome Research Institute.** 2003. A vision for the future of genomics research: a blueprint for the genomics era. *Nature* **422**:835–847.
7. **Corcoran, T., F. A. Mosher, and A. Rogat.** 2009. Consortium for Policy Research in Education, Research Report # RR-63, Learning Progressions in Science: An Evidence-based Approach to Reform. University of Pennsylvania. Philadelphia, PA.
8. **Dahabiah, A., J. Puentes, and B. Solaiman.** 2010, February 20–22. Possibilistic Kohonen Maps. Paper presented at the 9th WSEAS International Conference on Artificial

- Intelligence, Knowledge Engineering and Data Bases (AIKED '10), University of Cambridge, UK.
9. **Deichmann, J. I., A. Eshghi, D. Haughton, A. Woolford, and S. Sayek.** 2007. Measuring the international digital divide: an application of Kohonen self-organising maps. *Int. J. Knowl. Learn.* **3(6)**:552–575.
 10. **Dhar, P. K.** 2007. The next step in biology: a periodic table? *J. Biosci.* **32(5)**:1005–1008.
 11. **Driscoll, M. P., M. Moallem, W. Dick, and E. Kirby.** 1994. How does the textbook contribute to learning in a middle school science class? *Contemp. Educ. Psychol.* **19**:79–100.
 12. **Duncan, R. G., A. Rogat, and A. Yarden.** 2007. Learning Progressions in Genetics. Paper presented at the American Association for the Advancement of Science Project 2061: Knowledge Sharing Institute of the Center for Curriculum Materials in Science.
 13. **Evans, G. A.** 2000. The periodic table of biology. *Microb. Compar. Genomics* **5(4)**:179–180.
 14. **Gottfried, S. S., and W. C. Kyle Jr.** 1992. Textbook use and the biology education desired state. *J. Res. Sci. Teach.* **29(1)**:35–49.
 15. **Harms, N. C.** 1980. Project synthesis: summary and implications for action. ERIC Clearinghouse for Science, Mathematics, and Environmental Education, Columbus, OH; National Science Teachers Association, Washington, DC.
 16. **Hua, G., M. Skaletsky, and K. Westermann.** 2009. Exploratory analysis of CIA factbook data using Kohonen self-organizing maps. *Case Studies for Business, Industry and Government Statistics*, **3(1)**:48–59.
 17. **Kantardzic, M.** 2003. Data mining: concepts, models, methods and algorithms. IEEE Press, Piscataway, NJ.
 18. **Kaski, S., and T. Kohonen.** 1996. Exploratory data analysis by the self-organizing map: structures of welfare and poverty in the world. In A.-P. N. Refenes, Y. Abu-Mostafa, J. Moody, and A. Weigend, (ed.), *Neural Networks in Financial Engineering. Proceedings of the Third International Conference on Neural Networks in the Capital Markets*, London, England, 11–13 October, 1995 (p 498–507). World Scientific, Singapore.
 19. **Kohonen, T.** 2001. *Self-organizing maps* (3rd ed). Springer-Verlag, New York, NY.
 20. **Labov, J. B.** 2004. From the National Academies: the challenges and opportunities for improving undergraduate science education through introductory courses. *Cell Biol. Educ.* **3(4)**:212–214.
 21. **Lander, E. S.** 1996. The new genomics: global views of biology. *Science* **274(5287)**:536–539.
 22. **Larose, D.** 2004. *Discovering knowledge in data: an introduction to data mining.* Wiley.
 23. **Mason, C. L.** 1992. Concept mapping: a tool to develop reflective science instruction. *Sci. Educ.* **76(1)**:51–63.
 24. **Mervis, J.** 2009. Science education. universities begin to rethink first-year biology courses. *Science* **325(5940)**:527.
 25. **National Center for Education Statistics.** 2006. *Digest of education statistics.* Institute of Education Sciences, Washington, DC.
 26. **Nehm, R. H., et al.** 2009. Does the segregation of evolution in biology textbooks and introductory courses reinforce students' faulty mental models of biology and evolution? *Evol. Educ. Outreach* **2(3)**:527–532.
 27. **Novak, J. D., and A. J. Cañas.** 2008. *The Theory Underlying Concept Maps and How to Construct and Use Them.* Technical Report IHMC Cmap Tools 2006-01 Rev 01-2008. Florida Institute for Human and Machine Cognition, Pensacola, FL.
 28. **Roseman, J. E., A. Caldwell, A. Gogos, and L. Kurth.** 2006. Mapping a coherent learning progression for the molecular basis of heredity. Paper presented at the National Association for Research in Science Teaching Annual Meeting.
 29. **Roseman, J. E., L. Stern, and M. Koppal.** 2010. A method for analyzing the coherence of high school biology textbooks. *J. Res. Sci. Teach.* **47(1)**:47–70.
 30. **Skoog, G.** 1979. Topic of evolution in secondary school biology textbooks: 1900–1977. *Sci. Educ.* **63(5)**:621–640.
 31. **Skoog, G.** 1984. The coverage of evolution in high school biology textbooks published in the 1980s. *Sci. Educ.* **68(2)**:117–128.
 32. **Skoog, G.** 2005. The coverage of human evolution in high school biology textbooks in the 20th century and in current state science standards. *Sci. Educ.* **14(3)**:395–422.
 33. **Swarts, F. A., O. Roger Anderson, and F. J. Swetz.** 1994. Evolution in secondary school biology textbooks of the PRC, the USA, and the latter stages of the USSR. *J. Res. Sci. Teach.* **31(5)**:475–505.
 34. **Zalles, D.** 2010. *Evidence-Centered Design to State Large-Scale Science Assessment, Technical Report 10: Using Evidence-Centered Design to Support Assessment, Design, and Validation of Learning Progressions (DRAFT).* SRI International, Menlo Park, CA.