

Anthropologically introduced biases in natural history collections, with a case study on the invertebrate paleontology collections from the middle Cambrian Spence Shale Lagerstätte

Anna F. Whitaker and Julien Kimmig

ABSTRACT

Natural history collections are critical for modern scientific investigations, which are greatly expanding on the potential data applications of historic specimens. However, using these specimens outside their original intent introduces biases and potential misinterpretations. Anthropogenic biases can be introduced at any point during the life of museum specimens, from collection, preparation, and accession, to digitization. These biases can cause significant effects when the user is unaware of the collection context, as specific collection biases are often known anecdotally, but not ubiquitously. In this case study, the University of Kansas collection of Spence Shale Lagerstätte material was examined for anthropogenic biases using a collections inventory, interviews with stakeholders, and a literature review. Biases were found related to collector interest, locality preference, and researcher interest and specialization. These biases create a distorted view on the diversity and ecology of the Spence Shale, and need to be considered in future research.

Anna F. Whitaker. Department of Chemical and Physical Sciences, University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, ON L5L 1C6, Canada. annawhitaker01@gmail.com
Julien Kimmig. Earth and Mineral Sciences Museum & Art Gallery, Pennsylvania State University, University Park, Pennsylvania, 16802, USA. jkimmig@psu.edu

Keywords: Natural history collections; invertebrate paleontology; collections bias; soft-bodied preservation; Cambrian

Submission: 25 June 2020. Acceptance: 20 November 2020.

Whitaker, Anna F. and Kimmig, Julien. 2020. Anthropologically introduced biases in natural history collections, with a case study on the invertebrate paleontology collections from the middle Cambrian Spence Shale Lagerstätte. *Palaeontologia Electronica*, 23(3):a58. <https://doi.org/10.26879/1106>
palaeo-electronica.org/content/2020/3238-collections-biases

Copyright: December 2020 Paleontological Society.

This is an open access article distributed under the terms of Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0), which permits users to copy and redistribute the material in any medium or format, provided it is not used for commercial purposes and the original author and source are credited, with indications if any changes are made. creativecommons.org/licenses/by-nc-sa/4.0/

INTRODUCTION

Natural history collections (NHCs) are undergoing a transformative role within science during the twenty-first century. More so than ever, the specimens they contain are being applied to “big data” interdisciplinary problems such as disease control, the global biodiversity and ecological crises, and mitigation and measurement of the effects of global warming (Suarez and Tsutsui, 2004; Winker, 2004; Baird, 2010; Pyke and Ehrlich, 2010; Bakker et al., 2020; Hedrick et al., 2020). New technological applications and methodologies, such as genomics, computational statistics, CT scanning, and machine learning are bringing renewed interest and application to historic collections, decades or even centuries after they have been established (Wandeler et al., 2007; Nelson and Ellis, 2018). Natural history collections span from the seventeenth century on, making them invaluable representations of a record of anthropogenic change on a global scale (Nelson and Ellis, 2018). The value of these historic specimens lies in their ability to be used in time-series data, representing multiple collecting events of a locality or generations of a species over time (Lister and Climate Change Research Group, 2011; Habel et al., 2014; Holmes et al., 2016). Despite the nearly unlimited potential of natural history collections, they are under threat from lack of funding, cuts to stewardship positions (curators, collections managers, researchers), and lack of space. However, there are several reasons natural history collections must be maintained: e.g., historic collections are impossible to replicate on the same scale, they represent taxa and localities that no longer exist, and they are cost-effective to maintain in comparison to trying to accumulate new data sets (Allmon and Poulton, 2000; Suarez and Tsutsui, 2004; Lister and Climate Change Research Group, 2011).

While the application of collections-based data grows, the potential distance between the contextual knowledge of the data and end users is also increasing. Database aggregators and digitization efforts are mobilizing data on exponential scales, allowing researchers access to millions of specimen records across thousands of institutions (Nelson and Ellis, 2018; Hedrick et al., 2020). While this increases the potential scope of the analyses, it means the researcher is often removed from personal familiarity of each data point. Museum collections data have inherent vices that researchers must be made aware of, lest they misinterpret the resulting signals and patterns. Natural

history collections are nonrandom and represent a ‘presence only’ dataset, and must be treated as such (Wehi et al., 2012; Lipps, 2018; Hedrick et al., 2020). Examining sources of anthropogenic bias can help determine whether absences are true absences, or false signals due to introduced conditions.

This paper provides an overview on anthropologically introduced biases in natural history sciences that create discrepancies between the published literature, museum collections, and the reality of the natural world. As a case study, anthropogenic biases in the University of Kansas collection of the Spence Shale Lagerstätte material are identified through a collections inventory, types of material in the collection, comparison with the published literature, and identification of the collection context by discussion with the collectors, collection managers, and researchers who have used it through time. By identifying these biases, this work will serve as a guide for future researchers using the Spence Shale material, as well as an example for others concerned about collection bias in their studies.

INTRODUCTION OF BIASES IN NATURAL HISTORY COLLECTIONS

The challenge of accounting for collections bias is due to the extreme variability in types of bias. Natural history collections can be biased in infinite combinations relating to specimen taxon, preservation type, when the collection was made, the field collector, the collecting institution, etc. To discuss these issues, this paper is organized according to when a source of bias enters a collection (Figure 1), with examples from across biological/paleontological collections.

Field Bias

Field collection is the moment a living organism or natural object is removed from its environment and becomes a specimen of study. There are numerous biases introduced at this stage, often compounding as the distance between collection event and use in study occurs. To begin, the specimen’s physical characteristics influence selection. Overly large or bulky specimens may be left in the field due to the difficulty of transporting them (Nekola et al., 2019). Paleontological material that is too large, or too remote to remove without specialized mechanical equipment (e.g., cranes or helicopters) must be left in situ, or be accounted for in the budget of the collecting trip. Amateur collectors, or collectors from smaller institutions, are less

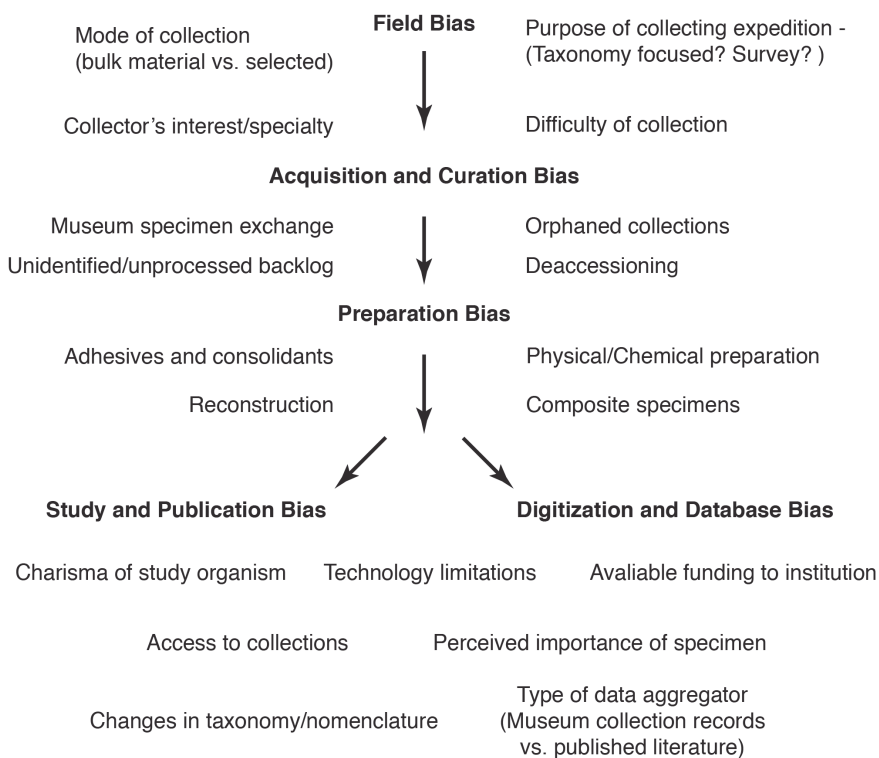


FIGURE 1. Examples of anthropological actions that create bias in natural history collections and when they occur.

likely to be able to fund such coordinated efforts with external contractors. In modern and fossil collections, smaller specimens requiring specialized equipment (microscopes, sieves) to locate may be ignored due to practicality. In contrast, coring or drilling samples may exclude large specimens and favor smaller specimens. In paleontological taxonomic collection, complete or articulated specimens may be favored over incomplete/disarticulated specimens, leading to a completeness bias. While completeness of specimens is crucial, or at least beneficial, to most types of research (e.g., systematics, biostratigraphy), there are other research questions that can still be investigated despite incomplete preservation, e.g., macroevolution, paleoecology, stratigraphic/geographic distributions (Lieberman and Kaesler, 2000; Allmon, 2005), and sometimes an incomplete specimen might be the only representative of a species at a locality. Poor preservation is also critical for understanding the limitations of preservation processes, and provides a vital comparison to high-quality or 'ideal' preservation (Kimmig and Pratt, 2016). Taxonomically unidentifiable material can also be used for destructive sampling, e.g., for stable isotope analyses or for testing effects of preparation and preservation before use on other

specimens, but should be cataloged and marked as such. Thus, even for taxonomically-oriented research, poorly-preserved or unidentifiable material should be considered for collection.

Specimens collected live (and to a certain extent fossil) in the field can be biased based on the stage of their lifecycle (Panchen et al., 2019), as well as towards the most common times of fieldwork: during the off-time of the academic calendar, holidays, etc. (Nekola et al., 2019). Additionally, the geopolitical landscape affects collection opportunities; this is especially the case in times of high geopolitical unrest, leading towards variable collection over time (Panchen et al., 2019).

The purpose of the collecting event often defines what information is associated with the specimen. A major bias results from the fact that fieldwork is often taxonomically-oriented, meaning some groups of organisms are favored over others. This undermines the future use of the data for occurrence-based ecological and diversity work, as some taxonomic groups will have artificially inflated specimen records (Lieberman and Kaesler, 2000; Allmon, 2005; Nekola et al., 2019). Taxonomic collection events are more favorable than stratigraphic, systematic, or bulk sampling events due to their inherent selective nature: excess material can

be discarded in the field, it is more difficult to transport bulk material, systematic sampling takes more time, and comprehensive sampling may require multiple localities.

Other biases are created through risk-versus-reward conditions of the collection event. Localities with known high productivity are often revisited, rather than attempting new localities with potentially lower productivity (Nekola et al., 2019). Additionally, there is a diametrically opposing influence of local vs. destination collection. Localities closer to home or near access points are easier and less costly to visit (Panchen et al., 2019), while iconic localities are touted (Nekola et al., 2019). These biases are often based on the research area of the scientist, as well as funding available.

Rarity of a species influences their collection: rare specimens are often over-represented compared to common specimens (McGowan and Dyke, 2009; Nekola et al., 2019). Perceived rarity can be caused by several factors; such as the charismatic nature of comparatively rare specimens, e.g., theropod dinosaurs are rarer ecologically than sauropod dinosaurs, but are over-collected (this pattern has been dubbed ‘Stromer’s Riddle’) (McGowan and Dyke, 2009). This can create disproportionate reconstructions of ecology and biodiversity (e.g., Davis and Pyenson, 2007).

Many natural history collections exhibit a Pareto distribution, where a majority of the samples are accrued by few collectors (Isaac and Pocock, 2015; Panchen et al., 2019). Therefore, minute eccentricities, errors, or preferences of the collectors can be amplified in a collection. It can be difficult to identify a collector-induced bias, unless it is seen firsthand or a comparison can be made of material from a different collector.

Another collection bias rooted in field selection is sex bias. The sex of specimens can affect research conclusions in taxonomy, genomics, comparative anatomy, stable isotope ecology, and morphological evolution (Cooper et al., 2019). A bias towards male specimens has been observed in several taxonomic groups, including modern birds and mammals, and is especially pronounced in species with males with ornate or showy characters, such as birds of paradise, and less pronounced in species with reversed characteristics (Cooper et al., 2019). However, this does not explain the bias in other mammals, like small rodents, which suggests the field selection procedures were somehow biased towards capturing/selecting males (Cooper et al., 2019). In order to address population sex ratios in the fossil record,

and compare them to modern population sex ratios, other sources of bias, including collecting bias, need to be ruled out in both collections. For example: in a study of fossil moa, a highly skewed sex ratio (determined by genetic sampling) was observed in two localities (Allentoft et al., 2009). While the analyzed material represents a nearly holistic sample of the available museum collection material, the validity of whether the observed patterns is a primary signal is contingent on whether “the museum collections are representative of the original collections”, i.e., a potential curatorial bias, as well as whether the collected material represents an accurate sampling of the composition of the fossil deposit (Allentoft et al., 2009). In fossil specimens that cannot be genetically sexed, analysis of sex must rely upon interpretations of morphological differences. Sexual dimorphism in the fossil record can be represented through discrete characters (presence/absence) or continuous characters, and interpreted behaviors (Kimmig and Spielman, 2011; Mallon, 2017). The same collection biases toward larger, showier, or ornate specimens could impact fossil sexual selection analyses, and considered in future studies.

Acquisition and Curation Bias

The collecting or funding institution controls what components of collections are retained and maintained over time. The institution is guided by its collection policies and by oversight from collection staff (collection managers, curators, division heads, directors). Collection policies must be flexible by nature, and reflect the current direction and priorities of the institution. Collections may be deaccessioned if they no longer align with the museum mission, or reassigned to different purposes (teaching collection, exhibition) if they no longer represent viable sources of data.

Museums routinely engage in specimen exchange, breaking up collections. Without published reference to the original composition of the collection, or in-house documentation of the exchanged material, researchers may be unaware of the fragmented and distributed collection.

Orphaned collections often represent specialized research collections belonging to university faculty, or sections of a larger collection (Golden, 2000). When institutions are considering being potential adopters of orphaned collections, they must assess whether the incoming collection would fit with the mission of the institution and whether it complements the current collection’s strengths or fills a weakness (Golden, 2000). The mission and

collections policies of institutions are by nature flexible and constantly updating to reflect the interests of the current stakeholders (curators, researchers, collection managers, directors, etc.). Therefore, there will be a bias towards accepting 'easily-accessible' collections (i.e., easy to move, easy to integrate into existing storage situations) and collections that complement the research interests of the accepting intuition. Collections that represent a large volume of material (by weight and by number of specimens), such as invertebrate and vertebrate paleontology collections, are not favorable from a logistical standpoint, despite being necessary from a research perspective.

Orphaned collections often contain different biases than collections in their adoptive institution, as they have been housed in a different institution and were often collected by different people than the specimens in the adopting institution. If not well documented when integrated into the new collection, the original bias might not be apparent and can become overwritten once accessioned into the new collection context. Because orphaned collections usually belong to smaller institutions or are a subset of a collection, they can contain numerous biases (e.g., localized or narrow zone of sampling; difficult collections that require maintenance, etc.).

Additional bias is introduced by the curation of the collection and changes of curators and collections managers over time. This is especially the case with large diverse collections. A curator is often specialized in a certain clade of animals or plants, which can lead to bias towards the curation of this clade. This bias is often reflected in improved identification of the clade of interest, whereas other clades are comparatively neglected, leading to poor higher-level taxonomy or potentially unidentified specimens. The more curators and collections manager work in a collection, and the more diverse their background, the smaller this bias. While numbers vary vastly between museum collections, there will always be unidentified or badly identified specimens in a collection.

Limited staff and limited time to deal with new accessions and backlog can also result in unprocessed and uncatalogued specimens. This can lead to researchers working on the collections of the museum not having all specimens of interest available to them, not knowing that certain specimens exist, resulting in potential misinterpretations of the diversity of a locality or missing taxonomically informative specimens. These as well as other curation methodologies and limitations can potentially create biases.

Collections and specimens are processed, curated, identified, and made accessible in non-random orders. These methodologies overlap with preparation, study, and digitization biases, as material can only be studied once it is made accessible, either physically or online.

Preparation Bias

While field bias mainly influences what is brought into the collection, preparation bias influences what information is lost or altered during this transfer.

The intention for preparation influences the procedures. Exhibition preparation is to maximize the presentation of the specimen, i.e., removing a paleontological specimen completely from the matrix, or preparing one side of a specimen (in slab form). This can restrict the future research done on the specimen; removing it from the context of the matrix removes ecological and taphonomic information, such as the articulation of the specimen, associated species, sedimentological information, and small-scale stratigraphy.

Another outcome of preparation is the restoration of specimens for aesthetic purposes, often repairing damage or adding replicas of missing parts to present a "whole" or "perfect" specimen (Le Cabec and Toussaint, 2017; Vidal et al., 2020). While this may be appropriate under certain museum circumstances (e.g., improving the understanding of a complex organism on exhibit for visitors), it can interfere with the scientific interpretation of the specimen. Display specimens may also be composites, created from multiple incomplete specimens. Historically, there was a lack of differentiation between research specimens and exhibit specimens, creating modern problems when referenced or type material was prepared in a manner befitting exhibition. For example, in a review of 18 "complete" *Xiphactinus audax* on display or in storage, 13 had restorations, and six were composite specimens (Bardack, 1965). Exhibits were valued for their showmanship, which means scientific interpretations occasionally took a back seat for dramatization and "artistic interpretations".

Restoration is not limited to exhibition specimens. Disarticulated or fragmented specimens are often reconstructed during research preparation. This can lead to a self-fulfilling prophecy: the fossils are constructed based on prior knowledge, influencing decisions and the interpretation of the outcome (Le Cabec and Toussaint, 2017; Vidal et al., 2020). An example of this is the La Chapelle-

aux-Saints Neanderthal skull: at the time of the initial restoration, Neanderthals were considered to be the linear mid-step between ape and human, therefore the skull was likewise reconstructed with features to reflect this inherent bias (Le Cabec and Toussaint, 2017). While it is traditionally the responsibility of the researcher to vet their data for inaccuracies, the abstract use of museum collections data and reliance on published references can impede their efforts. Reconstructions can be especially problematic if there is no record of their alteration, as future researchers can incorporate these into their analyses unknowingly, which emphasizes the importance of documentation and collection management.

Air scribing and air abrasion, two common preparation methods in paleontology, can remove or alter the fossil surface. Delicate structures, such as appendages or thin bone can be completely destroyed by these preparation methods if used heavily-handedly. Even 'softer' methods of preparation, i.e., the combination use of alcohol, water, acids, or acetone and using needles, brushes, or cotton swabs to remove matrix, left scratches and abraded the surface in experimental studies (Fernández-Jalvo and Monfort, 2008).

To compound the problem, historically there have been precedents to remove the inflicted evidence of preparation marks, such as using a treatment of hydrochloric acid (Le Cabec and Toussaint, 2017), which itself may remove valuable information.

Acid is a common preparation method to extract or separate fossils from matrix in situations where physical preparation is not suitable. In carbonate preservation settings, where there are both carbonate and silicified organisms, dissolving the matrix will also remove the carbonate fossils, leaving the silicified material (Pruss et al., 2015). Silicification can be taxonomically biased, therefore acid preparing a partially-silicified fauna could create a bias affecting ecological results (Pruss et al., 2015).

Not only can preparation remove characteristics, it may lead to false interpretations. Dental microwear, the study of the scratches and pits on the surface of teeth, records the diet of the animal on the scale of weeks to months. However, air abrasion preparation can alter the count of the microwear textures, altering the lifestyle interpretations, as extreme as misinterpreting a grazing lifestyle instead of a browsing lifestyle (Johnson, 2018). Additionally, preparation can mimic the preservation of soft tissues, leading to false posi-

tives, in which these soft parts are reported, but are in fact not actually present. For example, preservation of water-vascular systems and gut tracts were reported in some echinoderms from the Fezouata Shale Lagerstätten in Morocco, however, upon further examination the interpreted structures were artifacts of latex casts, and weathering (Saleh et al., 2020b). Exites were identified in Carboniferous insects from the Mazon Creek, but were also artifacts of physical preparation (Béthoux and Briggs, 2008). Appearances of specific soft-bodied tissues and organ structures are critical for establishing the timing of evolutionary lineages and characteristics, therefore false positives need to be minimized.

Preparation and preservation can also influence geochemical analyses. Experiments with modern bone specimens have shown common museum consolidants (Paraloid B-72 and Butvar B-98) can affect stable isotope $\delta^{18}\text{O}$ values (France et al., 2015). Similarly, procedures intended to prevent contamination of consolidates prior to radiocarbon dating (e.g., water, methanol, or acetone washes) tested ineffective against vinyl acetate-derived materials and cellulose nitrate lacquers (Brock et al., 2018). Environmental hydrocarbons from storage materials could be a source of biomarker signal contamination in fossil material (Gold et al., 2016).

Study and Publication Bias

The large discrepancy between the amount of specimen information published in peer-reviewed literature and the amount of specimen information in museum collections is well documented. One estimate of this disparity in paleontology collections found there are 23 times more localities represented in museum collections than published in the literature (Marshall et al., 2018). The current estimates of specimens held in natural history museums is $1.2\text{--}2.1 \times 10^9$, with less than 5% available through data aggregators (Ariño, 2010; Bakker et al., 2020). The publication record of museum collections is a magnitude smaller, and so is only capturing a minute percentage of the capacity of museum collections. The disparity between publication records and museum collections is problematic for diversity estimates that are sensitive to the evenness (how the specimen counts of the species compare to one another) of the data source. Because published records are likely more even than what is represented in museum collections, this can distort these diversity estimates (Davis and Pyenson, 2007).

The publication record is entirely dependent on how researchers approach their field, which is subject to shifts in interest and application over time. Koch (1978) noted “the published fossil record owes much to studies which had as their primary function the solving of biostratigraphic problems. It is reasonable that the published record may be biased in favor of biostratigraphically important taxa because these taxa have been studied by more persons and to greater taxonomic detail than fossils, which are seldom used in solving stratigraphic problems.” As research trends shift in response to funding opportunities, new applications, and industry, the published record and the specimens they reference will also shift, creating an ebb and flow of biases.

Charismatic taxa and taxa evoking a positive emotional response receive more public attention and financial support than non-charismatic or negatively associated taxa. (Amori and Gippoliti, 2000; Frynta et al., 2010). Even when harmless, there is a strong emotional response of disgust towards “ugly” taxa (Frynta et al., 2019). Researchers are not immune to the pull of charismatic taxa and are attracted to study certain taxa for similar emotional responses as the public (Lorimer, 2007). Research interest and financial support can inflate publication records, collection volume, and even taxonomic division, increasing apparent diversity (Nekola et al., 2019; Isaac et al., 2004). Primates are charismatic, well-known to the public, have established sources of funding for conservation efforts, motivation for establishment of species for protection legislature, and therefore attract much attention. This has led to an exponential increase of primate species over the last 20 years (Isaac et al., 2004).

Additionally, the ease of access to collections severely affects studies. The inaccessibility of specimens, be it that specimens are ‘under research’ by another group, or access to the collection is limited, can lead to false absences or misidentifications in datasets. If databases are not available online or if specimens cannot be sent on loan the research potential of the specimens is limited to research by those with the ability to physically travel to the collection. Often, those with the capability to make significant collections visits are senior researchers with a funding source and strong inter-institution influence. Junior faculty may be unable to commit to the same research enterprises. The location of the museum can be subject to external limitations that would discourage visiting researchers, such as travel restrictions, geopolitical unrest, location relative to other institutions,

and inadequate facilities to support the level of inquiry (no microscopes, CT scanning facilities, etc.).

Two of the foundational fields of the natural sciences are taxonomy and systematics, which is the process of defining, categorizing, and establishing the relationships between natural organisms. Systematic and taxonomic studies are contingent on type specimens held in museum collections (Lieberman and Kaesler, 2000; Allmon et al., 2018). Primary type specimens should be held in a public trust in perpetuity (as per Recommendation 16C of the ICZN), but formal requirements for publication of a new species are within the purview of the journal (ICZN, 1999). Many journals state they will not publish papers that include privately-held specimens (e.g., *Papers in Palaeontology*), creating a study and publication bias against material held in non-museum collections. Researchers may also be restricted by their professional societies, as some include restrictions on private specimens in their ethics statements (e.g., *The Society of Vertebrate Paleontology*). The ramifications of such constraints, especially for amateur paleontologists, remain an active discussion (Haug et al., 2020). Currently, the ICZN Code (1999) does not require allotypes (specimen of the opposite sex of the holotype) to be included as a type specimen (Rec. 72A, ICZN, 1999). Studies that use type material as species representation incorporate large scale sex biases - 75% of bird and 61% of mammal types (holotypes, syntypes, lectotypes, and neotypes) are male (Cooper et al., 2019). As discussed previously, sex affects a multitude of research outcomes, and this is amplified by limiting analyses to types (Cooper et al., 2019). Type and figured material is also on average larger in dimension than the means of bulk sampled material (Krause et al., 2007; Zhang et al., 2015). While this does not interfere with broad scale analyses comparing to other type and figured material, it should be kept in mind when the type material is used as a strict representative voucher for that species.

Digitization and Database Bias

Digitization is seen as the future of natural history collections, but there are major challenges associated with digitizing the sheer volume of material represented in museum collections, and therefore the “abstracted record” of the material will not match the physical museum collection record for a long time, if ever (Berents et al., 2010; Allmon et al., 2018). Until that theoretical equality is met, there will be a bias of which material is digitized.

Digitization efforts are often funded by sources external from the museum, i.e., large-scale initiatives (e.g., the National Science Foundation's Advancing the Digitization of Biodiversity Collections program) or smaller efforts that target specific collections or sub-sets of collections (Blagoderov et al., 2012; Nelson and Ellis, 2018). Digitization workflows emphasize prioritizing specimens according to institutional policies, user demand, type status, rarity, fragility, metadata completeness, taxonomic interest, ecosystem importance, and research projects (Berents et al., 2010; Blagoderov et al., 2012; Karim et al., 2016; Selmann et al., 2018; Hedrick et al., 2020; Marcer et al., 2020).

Errors associated with data input (e.g., transcription errors in locality, collector name, year of collection) can disproportionately represent abundances in collections. Changes in stratigraphic and taxonomic nomenclature can dissociate specimens from their original collections or correct taxonomic group. Duplicate records (i.e., records of the same specimen wrongfully included multiple times, often as a result of a taxonomic change creating records bearing both the old and new assignment) can artificially swell abundances of taxa or localities. Taxa that are subject to more intensive scrutiny (i.e., their taxonomy is being revised and changed more rapidly) will have a higher likelihood of these errors occurring, and therefore this digitization bias originates in a study bias. One of the larger biases that is introduced at this point is the cataloguing of lot specimens, as museums have not found a uniform way of doing this. Some museums will give each specimen in a lot a number, whereas other museums will just give the whole lot a single number. This can lead to problems when a study requires abundance information, as when a lot is assigned a single number, it is impossible to guess how many specimens are in it, whereas when each specimen within a lot has a number it will allow for accurate data. Giving each specimen in a lot a number can be problematic, when specimens are broken, and additionally it is much more time consuming than giving the lot a single number.

Database aggregators can compound the biases present in digitization and publication. Databases that source from museum collection records are only as current as the last time the collections were updated, making them susceptible to the same digitization problems (incorrect taxonomic assignments, stratigraphic nomenclature, etc.). In addition, taxonomic errors can occur if the underlying taxonomic information is not up to date, or if data entry errors occurred at the museum level.

Other biases can come from auto-populated fields in databases where the data is missing or incomplete. Databases that derive from published literature (e.g., Paleobiology Database, which is manually entered by volunteers) perpetuate the study biases that have been created by these publications and might have biases towards certain taxa depending on the people in charge of data entry. While standardization is becoming more prominent, there are still many concerns to be addressed (e.g., Wieczorek et al., 2012; Groom et al., 2019).

Mitigations

Solutions to the introduction of bias for incoming collections are preventive measures and diligent documentation. Many of these solutions are intuitive and apparent, as they represent good collections management and data maintenance procedures. Unfortunately, the methods for dealing with established collections in museums are more akin to coping mechanisms than solutions. Each collection will have its own unique biases, and identifying them constitutes the majority of the effort, and is also the most important aspect. The responsibility for mitigating biases is distributed among all stakeholders: collectors, museum workers, researchers, database users, institutions, etc. Biases, like those created by outdated taxonomy, cannot be fully addressed unless it is at multiple levels (e.g., creating funded, permanent positions for taxonomists, incorporating taxonomic updates as a designated responsibility of collection management positions, and emphasizing the duty of researchers to correct incorrect or outdated identifications when using collections). A list of recommendations for reducing biases at multiple levels in natural history museum collections is provided below. This list is by no means complete, but represents a sampling of actions and considerations museum workers and researchers can implement to reduce biases.

Field Bias

- Holistic sampling (Cook et al., 2017) and "Next-Generation Collections" (Schindel and Cook, 2018) champion the association of multiple integrated data samples from an organism and its surrounding environment. The caveat of this "collecting event cascade" is keeping all the data from one event correctly associated with multiple specimens across multiple databases, disciplines, and institution collections. This collection methodology

would add invaluable context to museum samples, but is restricted by financial and logistical limitations and takes time to successfully implement and integrate.

- Standardized forms for field collection that record not only the presence or actions taken, but additionally absences and inactions (e.g., juveniles specimens not collected, collections above stratigraphic level not made due to safety concerns). Noting taxa not present if known, vs. not collected, could make the difference in preserving the research integrity of a collecting event. Field collection notes should be immediately associated with the specimens from that collecting event through a relationship database.
- When collecting bulk samples, a proportion of the material should be kept unprepared for future comparison (Lister and Climate Change Research Group, 2011; Planavsky et al., 2020).

Acquisition and Curation Bias

- Documentation should be a priority and not limited to the purview of collections staff (Mulkerin, 2013). Loss of institutional memory can be circumvented by establishing exit interviews with staff and faculty leaving their positions. Memoirs, field and lab notes, and journals of researchers and staff should be conserved, transcribed, and digitized. Annual reports of divisions and collections should include methodological changes, events that impacted collections care (natural disasters, leaks, etc.), and large scale procedures (e.g., rehousing and accessioning orphaned collections). These files should be kept in multiple formats and backed up regularly.
- Gaps in collections should be identified and reflected in collections policies to work towards overcoming.

Study and Publication Bias

- Researchers should inquire whether the museum has fieldwork notes or records, and, if the collector is still living, potentially reach out to them to inquire about additional collections and the methodologies and practices they employed.
- Prompt notification of publication and return of specimens to museums for curation by researchers.
- Making sure that the collections numbers are up-to-date and correct, if possible have the

collections manager or curator of the collection check before final publication.

Preparation Bias

- When preparing specimens, documentation of all types of preparation made to the sample should be kept. Samples used for destructive analysis should be cataloged, and noted “Destroyed by sampling”, as well as retaining the resulting data and sampled material (thin sections, prepared slides, etc.). When possible, pictures, casts, or 3-D scans of the specimen should be taken before sampling to preserve morphological features.
- Specimens on exhibit should have clearly delineated restorations or modifications where necessary, that do not interfere with the interpretation of the specimen, and are justifiable based on the aesthetic necessity of the exhibition. Type specimens, or specimens that represent unique or rare individuals, should not be modified at all, as their research integrity should be prioritized.
- Micro-CT scanning and synchrotron micro-CT scanning can reveal past reconstructions and delineate between fossil and introduced material (such as metal, plaster, wood, and consolidants) (Le Cabec and Toussaint, 2017). Interpretation of characteristics and structures can also be supported through SEM, micro-CT, Raman, and other imaging techniques instead of purely optical microscopy (Saleh et al., 2020b). New methods of preparation, such as Laser Induced Breakdown Spectroscopy (LIBS), are capable of discriminating fossil material from matrix, and inflict less damage to fossil surfaces than traditional methods (Roberts et al., 2012).

Digitization and Database Bias

- Digitization bias can be mitigated by improving efficiency of digitization efforts, e.g., automating tasks, crowdsourcing metadata processes, conducting a ‘wall to wall’ approach and not opportunistically sampling (Blagoderov et al., 2012). Digitization efforts can be targeted to reduce spatial, temporal, and taxonomic biases in databases (e.g., Hedrick et al., 2020).
- Databases should follow established standards, e.g., Darwin Core, and constantly update their underlying databases. Researchers who notice taxonomic discrepancies or errors should contact the repository.

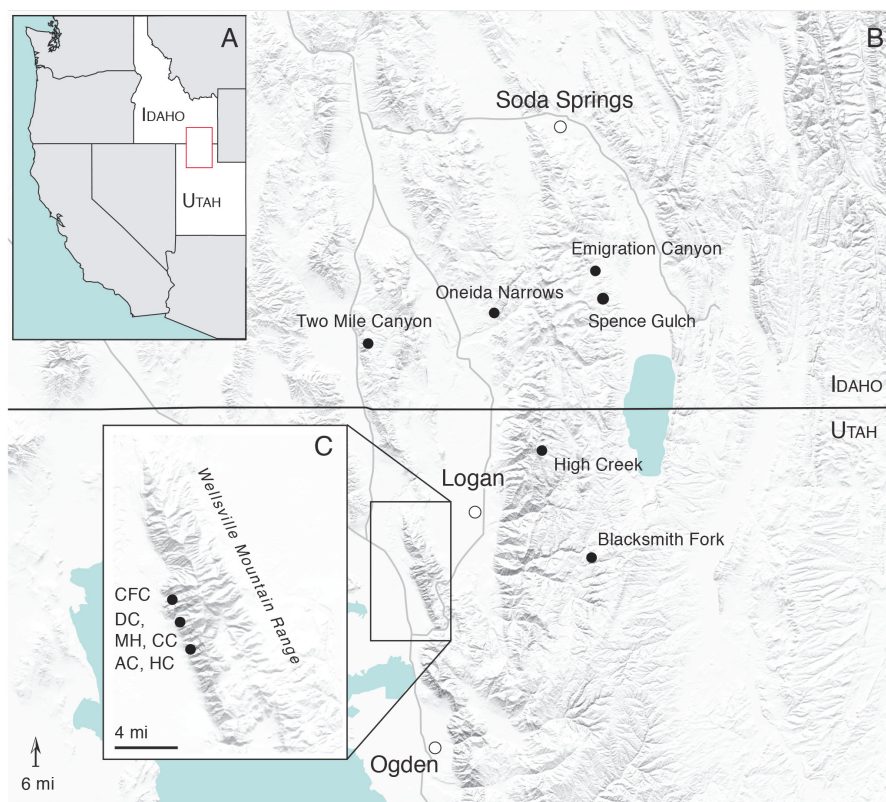


FIGURE 2. Location of the Spence Shale A. Location in the western United States. B. Topographic relief showing the main mountain ranges and collecting localities. The Wasatch Range contains the High Creek and Blacksmith Fork localities, and the Bear River Range contains Emigration Canyon, Oneida and the type locality Spence Gulch. C. Enlargement of the Wellsville Mountain Range (MH=Miners Hollow, CFC=Calls Fort Canyon, DC=Donation Canyon, CC=Cataract Canyon, AC=Antimony Canyon, HC=Hansen Canyon). Modified from Whitaker et al., 2020.

- Potential feature for users to ‘flag’ or annotate taxonomic entries for update, review, or synonymization.

A CASE STUDY IN PALEONTOLOGICAL COLLECTION BIAS: THE SPENCE SHALE LAGERSTÄTTE AT THE UNIVERSITY OF KANSAS NATURAL HISTORY MUSEUM

Introduction

The Spence Shale is the middle member of the Langston Formation, dating to the Cambrian (Miaolingian: Wuliuan). Located in northern Utah and southern Idaho, USA (Figure 2), this fossil Lagerstätte preserves large amounts of trilobites, echinoderms, and brachiopods, as well as lightly-sclerotized and soft tissues (Kimmig et al., 2019). The University of Kansas Biodiversity Institute, Division of Invertebrate Paleontology (KUMIP) holds the largest collection of material from the Spence Shale, totaling approximately 5,700 specimens as of 2020. This material is considered a tax-

onomic collection, as no bulk sampling or systematic stratigraphic collection has occurred (however, specimens retain associated stratigraphic information, if known). In addition, the Harvard Museum of Comparative Zoology, the Natural History Museum of Utah (UMNH), the Smithsonian National Museum of Natural History (USNM), the Yale Peabody Museum of Natural History, the Idaho Museum of Natural History, the Back to the Past Museum (near Cancun, Mexico), and the Ohio State University School of Earth Sciences also hold significant collections of Spence Shale material. These collections are dominated by trilobites and echinoderms, and contain few, if any soft-bodied specimens.

Collection and Publication History

One of the first scientific collections of the Spence Shale was made in 1896 by Robert S. Spence (1844-1916), an attorney practicing in Evanston, Wyoming, who sent the material to Charles Walcott at the USNM (Walcott, 1908;

Resser, 1939). Another local resident, Knud Hansen Fridal, Jr. (1881-1961) who resided in Box Elder County, Utah, also sent Wellsville Mountains material to the USNM (V. Gunther, personal commun.). Walcott described the Spence Shale in 1908, wherein he demarked the type locality, Spence Gulch, named after Mr. Spence, and noted the abundance of fossil material and the potential for excellent preservation (Walcott, 1908; Kimmig et al., 2019). Walcott made extensive collections in the Spence Shale, depositing them in the USNM, but did not publish on the fossil material (Robison et al., 2015). His former assistant, Charles Resser, published a listing of the hard-shelled fauna, as well as the first record of soft-bodied fossils, several specimens of the priapulid worm *Selkirkia spencei*, based on these museum collections (Resser, 1939). The first stratigraphic work was done in the 1920s and 1930s at localities near Blacksmith Fork (Mansfield, 1927; Deiss, 1938). World War II forced a hiatus on research, ending the first half-century of Spence Shale investigation (Robison et al., 2015).

During the second half of the twentieth century, the Spence Shale received renewed interest, mainly due to the influx of fossils collected with soft-tissue preservation by amateur collectors, namely the Gunther family, Phil Reese, and Paul Jamison. Resulting publications focused on identification and taxonomic placement of the Spence Shale fauna (e.g., Robison, 1969; Conway Morris and Robison, 1986; Babcock and Robison, 1988). The paleo-environmental and stratigraphic context of the Lagerstätte is slowly emerging, but still lacking compared to the depth of knowledge of other Cambrian deposits (Maxey, 1958; Liddell et al., 1997; Garson et al., 2012; Kloss et al., 2015). In the early 2000s only a few taxonomic papers were published (Sprinkle and Collins, 2006; Briggs et al., 2008), and the Spence Shale was included in a couple of large-scale datasets of Cambrian Lagerstätten (Hendricks et al., 2008; Gaines et al., 2008). However, in the 2010s, interest in the Spence Shale increased once again, and papers describing new taxa from the Spence Shale (e.g., Robison and Babcock, 2011; Conway Morris et al., 2015a, b; Kimmig et al., 2017; Wen et al., 2019) and revisions of taxonomic interpretations (e.g., Maletz and Steiner, 2015; Legg and Pates, 2017; Whitaker et al., 2020) were published. In addition, increased numbers of papers focusing on macroevolution, paleoecology, taphonomic pathways and preservation of the Spence Shale were published (e.g., Garson et al., 2012; Hendricks, 2013; Broce

and Schiffbauer, 2017; Kimmig and Strotz, 2017; Whitaker et al., 2020). These can be considered as a second generation of publications, where the focus shifts from simply observing characteristics or recording the fauna, to using this data (species counts, faunal compositions, etc.) within a broader application of biodiversity, comparisons to other Lagerstätten, and evolutionary trends.

The overwhelming majority of specimen collecting from the Spence Shale has been performed by amateur paleontologists (over 95%), pulling this Lagerstätte into the ongoing complex discussion regarding the legality, ethics, and stewardship of geologic and paleontological heritage in the United States (e.g., Lieberman and Kimmig, 2018). A number of the Spence Shale Lagerstätte localities are on public land, namely the Forest Service. The Forest Service implemented new policies under the 2013 Paleontological Resource Preservation Act (PRPA) to limit collection by amateurs by placing limits on the amount and type of material allowed to collect, as well as prohibiting the use of powered and large hand tools (Gunther, 2016). Because of the amount of work and material needed to extract the fossils of the Spence Shale, Glade Gunther considered this legislature highly discouraging to the future of amateur collecting and research collaboration in the Spence Shale (Gunther, 2016).

The Gunther family moved to the Brigham City area in 1965, and between the years 1965-2015 collected at least 10,000 specimens from the Spence Shale and other Cambrian deposits in Utah and surrounding states, contributing to 75% of the museum collections of Spence Shale material (Gunther, 2016; Kimmig et al., 2019; Kimmig, personal observ.). The Gunther family collectors (Lloyd Gunther, Metta Gunther, Val Gunther, and Glade Gunther) first contacted Stewart Williams at the Utah State University (V. Gunther, personal commun.). Williams recounted the collection history in the Wellsville Mountains, but cast doubt on any hope of finding new material. He mentioned Knud Fridal's name, and Lloyd Gunther managed to locate Knud's sister (May Hansen Fridal, 1885-1974) in the phone book. Unfortunately, Knud Fridal had passed, but his fossil collection and fieldwork notes were kept intact by his sister (V. Gunther, personal commun.). Consulting his notes, the Gunther's were able to rediscover localities that had been lost since the early 1900s, including Calls Fort/Miners Hollow, Cataract Canyon, and Hansen Canyon. In the summer of 1967, they relocated the Spence Gulch type locality (V. Gunther, personal commun.). Val Gunther discovered the Antimony

Canyon locality in the summer of 1974, discovering an abundance of *Gogia* fossils (V. Gunther, personal commun.). The Paleontological Society bestowed the family with the first Strimple Award in 1984 for “outstanding achievement in paleontology by amateurs” based on their work in the Utah formations.

The University of Kansas became a major repository for Spence Shale material in the late 1970s, when Richard A. Robison became an affiliated professor in the Department of Geology. Previously positioned at the University of Utah, he began studying the Spence Shale and accruing collections (part of his research collection was placed in the University of Utah invertebrate paleontology collection, which was later acquired by the UMNH) (Robison, personal commun.). During this time, Robison was approached by Lloyd Gunther, who invited Robison to view his personal collection, beginning the close relationship with the multi-generational fossil collecting family. When Robison transferred to the University of Kansas, he brought along a component of his research collection, and donations from the Gunthers’ and others continued (Robison, personal commun.). This material is the focus of the case study.

Bulk sampling was not conducted by the amateur collectors and none of the major museum collections of the Spence Shale include bulk collections. Two dissertations by Campbell (1974) and Wright (1999), used bulk samples to examine the paleoecology and fauna, but neither study was published, and either the specimens were not retained (Liddell, personal commun.), or information was not available. One bulk sample collection is retained at the Idaho Museum of Natural History, from fieldwork conducted at the Spence Gulch locality during the 2019 field season by J. Kimmig, R. LaVine, and L.J. Krumenacker.

Aim of Study

- Identify gaps in the KUMIP Spence Shale collection to direct future collection efforts
- Identify and anticipate potential sources of anthropogenic bias
- Identify discrepancies between the museum collection and external database records, to guide efforts to resolve the inconsistencies
- Establish a reference framework for future work in the Spence Shale

Because no large-scale, comprehensive stratigraphic collection of the Spence Shale has been made, a direct comparison of the “true” composition of the Spence Shale to the museum col-

lection cannot be made at this time. When such a study is made, this paper will represent a comparison point to identify biases and effects on research in this Lagerstätte. An interdisciplinary approach was applied in order to understand the bias in the Spence Shale material held by the KUMIP.

Methods

First, an inventory was taken by physically locating each specimen present in the collection assigned to the Spence Shale formation in the KUMIP Specify database. At the time of the survey, there was no uncatalogued material from the Spence Shale. Specimens on loan/not present in the collection, and without a databased image, were counted but not given a preservation score. Trace fossils were not counted and not scored for preservation. Each specimen was scored based on the amount of preservation (Appendix 1), isolated fragments were assigned as 0, articulated specimens under 50% complete were assigned as 1, specimens articulated and over 50% complete were assigned as 2, and specimens that were incomplete due to breakage of the rock were assigned as 3. Second, the collection at the KUMIP was compared to the available data from two external databases: The Paleobiology Database (PBDB), which is user-input information from published sources; and iDigBio, which pulls specimen data from museum collection databases. Species occurrence information from the PBDB was gathered from the geologic names “Spence Shale”, “Spence Shale Member”, “Spence Tongue”, and “Lead Bell Shale” to account for the changes in stratigraphic assignment. The purpose of this comparison was to understand the accuracy of the databases as compared to an updated genera list, the KUMIP Specify database, and the specimens in the collection. While certain fields of the KUMIP Specify database are accessible by the public, the iDigBio results illustrate what conclusions would be drawn if the information from the KUMIP database was incomplete (i.e., no specimen pictures), or without personal familiarity and access to the specimens. Third, the anthropologic history of the Spence Shale was compiled by comparing the patterns and timing of publications from the Spence Shale to socio-political events, as well as placing them in the context of the affiliations and research affinities of the authors. Collectors, researchers, and collection managers who interacted with the Spence Shale material were interviewed about their methodologies, motivations, and contemporary views on the collection. Understanding how

TABLE 1. Composition of the KUMIP Spence Shale collection by specimen count.

Taxa Group	Specimen Count	Percentage of total
Echinoderms	2141	38.30%
Trilobites & agnostids	1848	33.06%
Hyaloliths	671	12.00%
Brachiopods	342	6.12%
Problematica (<i>Eldonia</i> , <i>Banffia</i> , <i>Siphusauctum</i> , <i>Armillimax</i>)	166	2.97%
Soft-bodied arthropods	115	2.06%
Hemichordata	82	1.47%
Indeterminate	53	0.95%
Scalidophorans	47	0.84%
Lobopodians	37	0.66%
Algae & cyanobacteria	37	0.66%
Molluscs (<i>Latouchella</i> , <i>Scenella</i> , and <i>Wiwaxia</i>)	32	0.57%
Porifera	19	0.34%
Total	5590	100%

the collection was formed will identify potential gaps in the collection, as well as biases that can be quantitatively tested by future work.

Results

Taxonomic diversity. The University of Kansas Spence Shale collection contains 5590 databased specimens. Trace fossils were not included in the study. By specimen count, the collection is dominated by 38.30% echinoderms, 33.10% trilobites and agnostids, with the remaining 28.6% representing a diverse assemblage (Complete dataset in Table 1; Figure 3).

Specimen distribution by locality. The following localities represent the majority of specimens: 42.68% from Antimony Canyon, 15.17% from the Wellsville Mountains (not specified further), and 12.02% from Miners Hollow (Complete dataset in Table 2; Figure 4).

Completeness/tissue analysis. 5,413 specimens present in the Spence Shale collection were inventoried. Specimens that were on loan and did not have a database image were excluded from this completeness/tissue analysis (177 specimens).

Over the entire collection, 83.84% of the represented specimens are complete (articulated and over 50% complete), 7.65% are incomplete (articulated but less than 50% complete), 4.23% are isolated fragments, and 3.10% are partial due to breakage. Fifty-nine taxonomically-unidentifiable specimens (1.18%) that are included in the inven-

tory have not been assigned a completeness score. The high completeness pattern is driven by biomineralized taxa (e.g., trilobites, echinoderms). Of the biomineralized fossils, 88.42% are complete, 6.36% are incomplete, 1.85% are isolated fragments, and 3.37% are partial due to breakage.

Non-biomineralizing taxa represent 43.48% of the specimens that are complete, 19.02% are incomplete, 25.18% are isolated fragments, and 0.72% are partial due to breakage. Of the non-biomineralizing specimens, 11.59% represent indeterminate taxa that have not received a completeness score.

Of the type and figured specimens, 45.1% of the fossils are complete, 19.6% of are incomplete, 34.6% are isolated fragments, and none are partial due to breakage. The higher proportion of isolated fragments is largely driven by the soft-shelled arthropods (mouthparts, feeding appendages, carapaces, and valves are counted as isolated fragments).

Comparison to data aggregators. Presently, there are 74 genera published from the Spence Shale (Robison et al., 2015; Kimmig et al., 2019; Wen et al., 2019; Kimmig and Selden, 2020; Whitaker et al., 2020) (Figure 5). The genus *Morania* is included, but its affinities as a cyanobacteria are disputed (Robison et al., 2015). The PBDB contains records of 67 unique genera from the Spence Shale from 12 publications (Appendix 2, Figure 5) (Walcott, 1908; Resser, 1939; Maxey, 1958; Sprinkle, 1973; Campbell, 1974; Willoughby and Robison, 1979; Rigby, 1980; Briggs and Robison, 1984; Sumrall and Sprinkle, 1999; Sprinkle and Collins, 2006; Briggs et al., 2008; Wen et al., 2019). Fifty-two of these 67 genera are accurate, i.e., they are present in the most up-to-date faunal record of the Spence Shale. The PBDB is currently sampling 24.49% of the publication record on the Spence Shale. The generic faunal composition based on the PBDB is dominated by trilobites (28 genera, 41.8%), and soft-bodied arthropods (11 genera, 16.4%), with brachiopods (6 genera, 8.9%), echinoderms (5 genera, 7.5%), and a diverse fauna (17 genera, 25.4%) composing the rest.

The iDigBio database records of the KUMIP Spence Shale collection returns 74 scientifically-valid genera (excluding ichnofossil genera). Of those genera, 71 are fossil genera, and three are modern insect genera mistakenly assigned to specimens (Appendix 2). Of the 71 valid fossil genera, 57 are accurate, i.e., they are present in the most up-to-date faunal record of the Spence Shale.

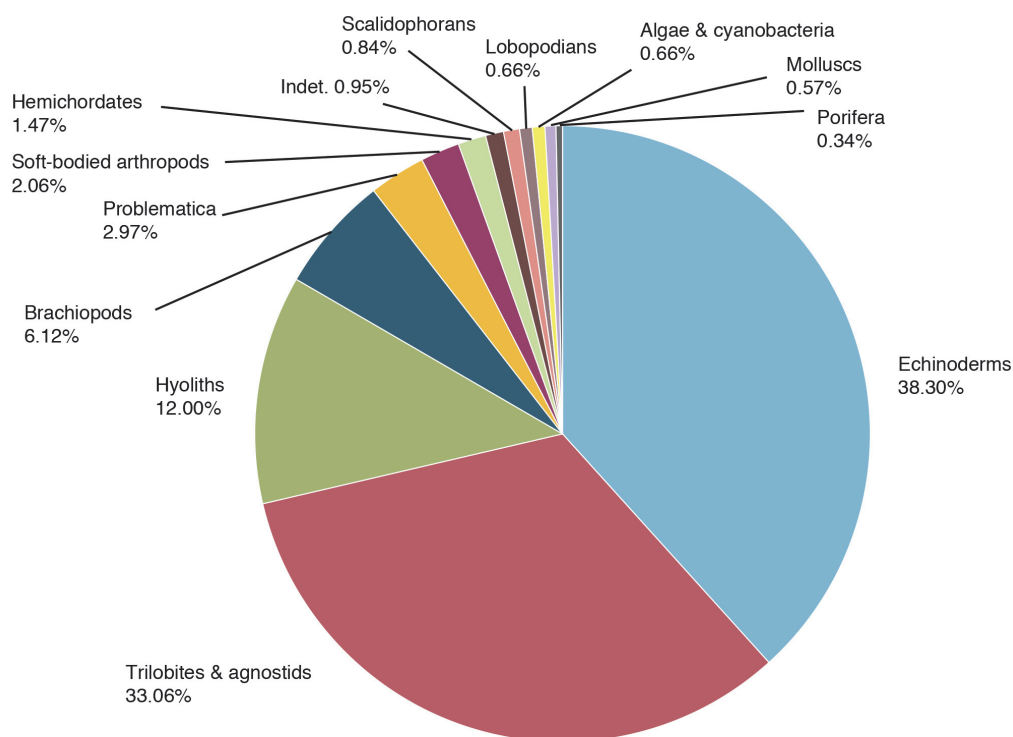


FIGURE 3. Percentage of taxonomic groups in the KUMIP Spence Shale collection by specimen counts.

Discussion

Konservat-Lagerstätten are sites where extraordinary depositional environments have been conducive to preserving tissues and features that would otherwise not be preserved (Seilacher, 1970; Allison, 1988; Butterfield, 2003). These deposits are often referred to as “snapshots in time,” capturing instances of diversity and ecology without the overprinting of time-averaging. Therefore, understanding introduced biases to these Lagerstätten collections are critical for correctly interpreting them. Studies often interpret the museum collections as one-for-one duplicates of the physical fossil record, as well as assuming collection efforts across multiple formations are identical (Saleh et al., 2020a). The Spence Shale has a markedly different collection history from many early Paleozoic Lagerstätten, and this must be taken into account. The major biases of the Spence Shale in the KUMIP collection are caused by collector interest/selection, accessibility of localities, and researcher interest/study bias.

Field bias. The crucial influencers of the collection composition were the amateur collectors. While researchers occasionally provided specific targets for collection efforts, the majority of the collecting was opportunistic and self-motivated. The land sur-

rounding the Wellsville Mountain localities is privately owned, and permission is needed to access the localities. The local amateur collectors and researchers who built a rapport with the landowners had more success receiving permission and being able to collect frequently (Robison, personal commun.). Robison devoted great effort familiarizing himself with the local amateur paleontologist and geology collector scene, viewing their collections and visiting their collecting localities (Robison, personal commun.). In turn, the collectors felt comfortable donating their material to Robison, first at the University of Utah, and later the University of Kansas. These factors created a preference towards local study and publication, and in the choice of collections repository.

The method and focus of collection in the Spence Shale shifted over time. Initially, the Gunthers collected fossils they could readily identify from float material, focusing on large and “nice” complete specimens of trilobites (V. Gunther, personal commun.). After approximately four field seasons at Miners Hollow, the productivity of the surface material was declining. They turned to quarrying the fossil-bearing strata, first with hand tools and later with larger equipment (V. Gunther, personal commun.). The later collection efforts

TABLE 2. Breakdown of Spence Shale collection by locality.

Locality	Number of specimens	Percentage (%) of total specimens per locality	Number of biomineralized specimens	Number of soft-bodied specimens	Percentage (%) of biomineralized specimens per locality
Antimony Canyon	2386	42.68	2314	72	96.98
Wellsville Mountains	848	15.17	790	58	93.16
Miners Hollow	672	12.02	277	395	41.22
Oneida Narrows	635	11.36	629	6	99.06
High Creek	494	8.84	482	12	97.57
Spence Gulch	416	7.44	411	5	98.80
Emigration Canyon	57	1.02	57	0	100.00
Cataract Canyon	38	0.68	20	18	52.63
Hansen Canyon	19	0.34	12	7	63.16
Utah	6	0.11	1	5	16.67
Two Mile Canyon	6	0.11	6	0	100.00
Calls Fort	3	0.05	1	2	33.33
Promontory Mountains	3	0.05	3	0	100.00
Donation Canyon	2	0.04	0	2	0.00
Unrecorded	5	0.09	4	1	80.00
Total	5590		5007	583	

focused on Miners Hollow. Through collaboration with Dr. W.D. Liddell, Professor of Geology at Utah State University in Logan, and understanding of the stratigraphy of the outcrop, the Gunthers were able to hone in on intervals containing soft-bodied fossils, which they, and other collectors, worked extensively (V. Gunther, personal commun.). This is clearly evident in the proportion of soft-bodied material compared to biomineralizing material from Miners Hollow, where soft-bodied fossils represent 58.78% of the material collected (Figure 4).

Other amateur collectors followed similar patterns, focusing on the Wellsville Mountain localities and selecting well-preserved specimens and “rare” specimens (Jamison, personal commun.). Paul Jamison clearly states he doesn’t “collect poorly preserved or partial specimens of common taxa... Once I have a few good specimens of a common taxon, I generally don’t collect more unless they are exceptional in some way. For example, I will only collect a hyolith if it has an operculum and at least one helen, or if it has soft tissue.” Of the inventoried hyoliths, only 6.12% of the specimens represent fragments (i.e., an isolated operculum) or specimens with less than 50% of the shell.

The Gunther family and Paul Jamison both collect out of private interest and provide interesting specimens to Science. Currently there is only one commercial collector, Jacob Skabelund, active

in the Spence Shale, and he also provides scientifically interesting specimens to researchers.

A locality bias is present due to the difficulty of access and preservation issues at some of the sites. At Cataract Canyon, there are well-preserved fossils present, but the shale fissures into diagonal fractures, making collection difficult (V. Gunther and Jamison, personal commun.). Hansen Canyon contains remarkable specimens of *Glossopleura gigantea*, but the hike to the locality was so strenuous that the Gunthers chose to concentrate their efforts on other localities, namely Miners Hollow (V. Gunther, personal commun.). Cataract Canyon and Hansen Canyon have the lowest number of specimens sampled of the major Wellsville Mountain localities. Miners Hollow is universally preferred due to the relative ease of access, faunal diversity, and preservation quality (V. Gunther, personal commun.; Jamison, personal commun.).

When researchers visited the field themselves, it was their personal specialization that drove the collections bias. Robison recounted taking James T. Sprinkle (UT Austin) to a Spence Shale locality, despite warning him that the particular echinoderms he was interested in were difficult to find there. Upon reaching the field site, Sprinkle immediately pinpointed dozens of the echinoderms, proving especially capable of locating the taxa he was familiar with (Robison, personal commun.). The Gunthers were occasionally directed to

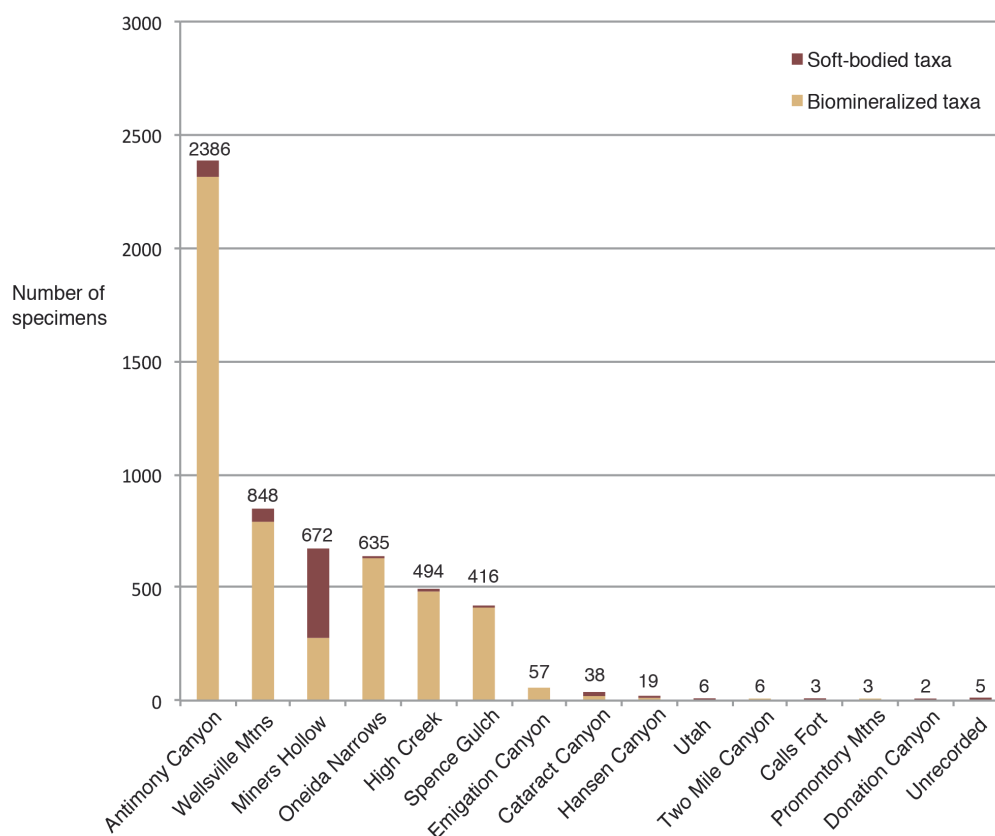


FIGURE 4. Specimens in the KUMIP collections by locality, breaking down number of specimens of biomineralizing and soft-bodied taxa from each locality.

collect large numbers of a taxa, and compensated financially for their work by the requesting researcher (Robison, personal commun.). Robison would also facilitate contact between collectors and researchers with a taxonomic specialty if he believed they would be interested in a certain specimen, e.g., Dr. James Sprinkle and echinoderm specimens (Robison, personal commun.; V. Gunther personal commun.). This created research collections focused on certain taxonomic groups, meaning those university collections are not indicative of the true diversity of the Spence Shale, only disproportionally enriched with a fraction of it.

The changes in collection focus over time have led to less material being collected by the amateur collectors in the last couple of decades, and to concentrate their efforts on soft-bodied fossils, seeing increasing donation of soft-bodied fossils. The Gunther family has also moved away from the Spence Shale and is mostly collecting in the Green River Formation nowadays, which has left Paul Jamison as the only major amateur collector active in the Spence Shale. Additionally, collecting efforts over time have shifted to Miners Hollow for

all collectors, because of the easier access and the quality of preservation, leading to increased numbers of Miners Hollow specimens in the KUMIP collection.

It is difficult to establish when many of the Gunther and Jamison specimens have actually been collected, as donations were often done in bulk, especially in 2017 and 2018, and represented years of collecting trips. Sometimes actual collecting years have been recorded, but often the specimens just include location information.

Preparation bias. The traditional method of recovering specimens in the Spence is fracturing blocks, in the hopes of splitting the rock along a surface containing fossil material. It was a longstanding belief that the matrix of the Spence Shale was too hard to completely prepare out a specimen (Jamison, personal commun.). However, recent collections by Jamison and Skabelund have challenged this belief, recovering three-dimensional trilobite material showing details of preservation, such as axial spines, not revealed by the traditional splitting method (Jamison, personal commun.). While this method requires an expert preparator, it can be

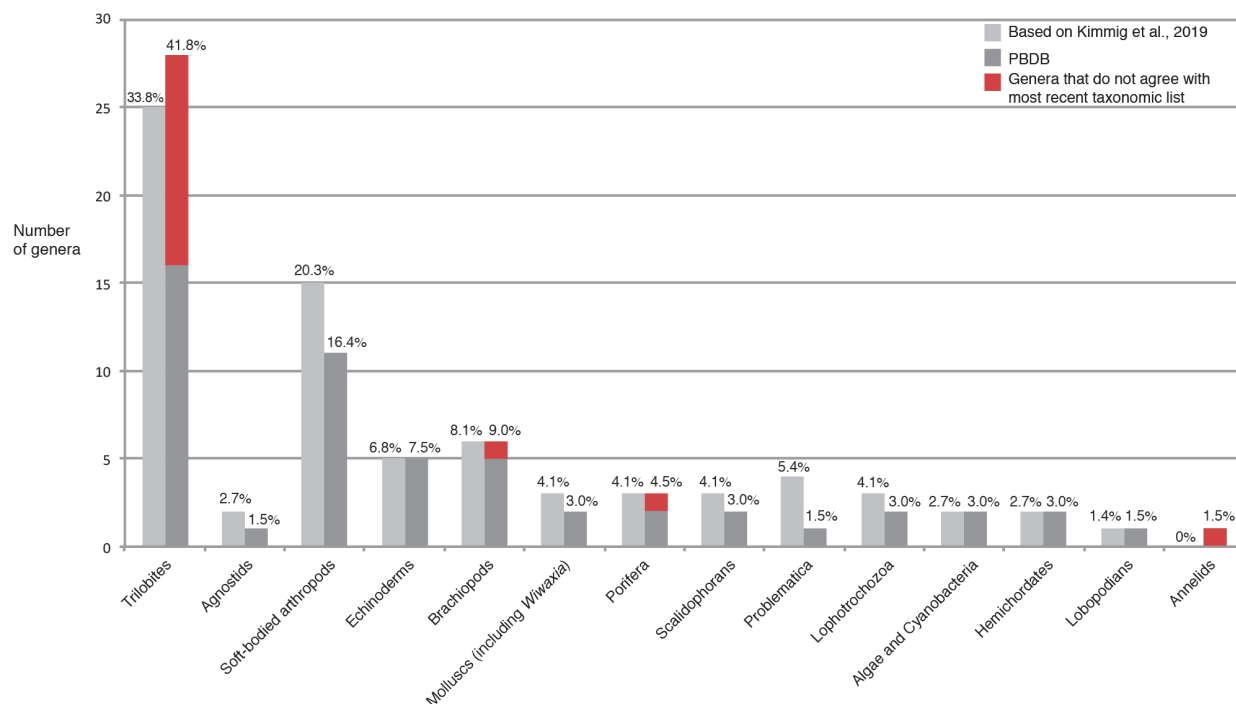


FIGURE 5. Composition of the Spence Shale fauna by number of genera, comparing the most recent faunal list based on Kimmig et al., 2019; and results from the PBDB. Genera in the PBDB not present on the most recent list are colored in red. The proportion of genera from each taxonomic group is expressed as a percentage above each count.

applied to previously collected split specimens, as parts and counterparts can be prepared out and rejoined (Jamison, personal commun.). Preparing specimens after collection also changes the collection method: instead of uncovering a productive layer to split, Jamison focuses on looking for “bumps” that indicate a fossil is encased in situ (Jamison, personal commun.). While the soft-bodied compression fossils are largely unaffected, the splitting collection method could lead to historically-overlooked diagnostic characters in the biomineralized taxa.

Study and Publication Bias

Researcher interest and specialization has driven the publication history and published diversity of the Spence Shale (Figure 6). A majority of the published literature (50 out of 81 publications) is primarily taxonomic or systematic in nature (Appendix 3). The publication record and taxonomic diversity of the soft-bodied material are greatly disproportionate to the number of specimens, especially the soft-bodied arthropods. Despite comprising only 2.06% of the specimens, soft-bodied arthropods represent 21% of the generic diversity, and are the subject of 18% of the taxonomic publications on the Spence Shale. Other

groups, such as the brachiopods, are under-represented due to lack of research interest. Six species are currently known, but have not been subject of detailed interest or revision since Resser (1939). They are one of the taxa collectors consider “common,” and their collection is undesirable (Jamison, personal commun.) Therefore, it is highly likely that the true diversity of brachiopods is much higher than what is currently represented in the collection (Kimmig et al., 2019; Robison personal commun.).

Digitization and database bias. The publication record of the Spence Shale is under-represented and distorted by the PBDB. The references providing an overall faunal list are older (Walcott 1908; Resser, 1939; Maxey, 1958; Campbell, 1974), and the newer references are focused on specific taxonomic groups, e.g., echinoderms and soft-bodied arthropods (Sumrall and Sprinkle, 1999; Sprinkle and Collins, 2006; Briggs et al., 2008). While the composition of the fauna by number of genera of the PBDB is similar to the latest review paper (Figure 5), the shared genera between the PBDB and Kimmig et al., (2019) is only 77%. That is to say, despite there being similar numbers of genera, they are not recording the same genera. The PBDB also misrepresents the relative diversity of some taxonomic groups: it overstates the number

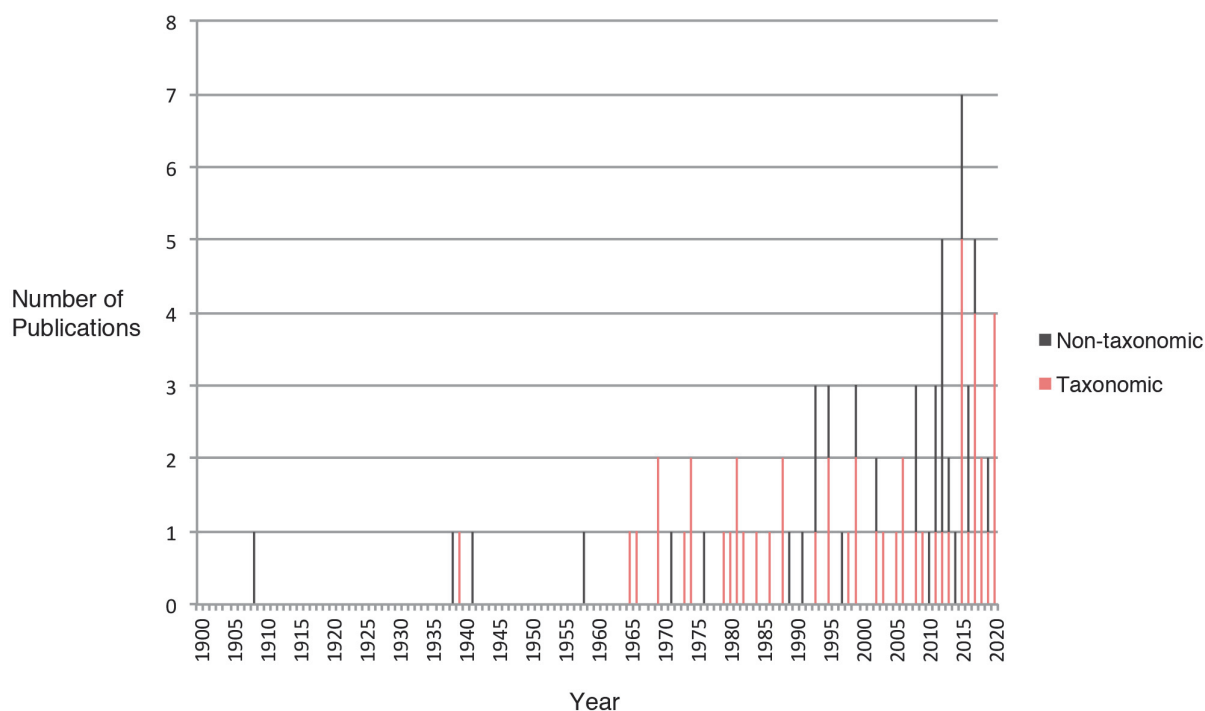


FIGURE 6. Publications over time that refer specifically to Spence Shale specimens or use quantitative data from the Spence Shale in an analysis.

of trilobite genera (28) than what is currently known (25) due to taxonomic inflation and redundancy, and under-represents the soft-bodied arthropods (11 genera recorded in the PBDB, compared to 15 currently published).

The iDigBio aggregator of the KUMIP Spence Shale collection records accurately reflects the number of specimens present in the collection, but introduces error in the transfer of information from the Specify database to the iDigBio aggregator. In cases where specimens are not identified to a species or genus level, a modern genus is occasionally substituted (e.g., KUMIP 107343, identified to the class level as *Trilobita* in the Specify database, is incorrectly identified to the modern insect genus *Trilobia* in iDigBio). This is a common problem uniting taxonomic systems across database systems, especially when combining multiple types of biological collections. While this type of error may be more difficult to catch than simple misspellings (because the taxonomic status is flagged as valid), data cleaning is an integral responsibility of the researcher. Eighty percent of the genera are accurate compared to the most up-to-date faunal list, but several genera known to be in the collection (i.e., *Utahscolex*, *Armillimax*) are excluded, due to

the lag in updating the collections records after publication.

Mitigations

Without large-scale stratigraphic or bulk sampling, quantitative comparisons cannot be made between the museum collection and field abundance. As yet, the direct effect of the anthropogenic biases cannot be established. However, based on the history of collection methods in the Spence, it raises significant concern when compared to other BST Lagerstätten. A recent Burgess Shale ecological study used abundance (number of specimens), and species richness (number of species) as a biodiversity metrics because the authors established that the mode of collecting (centimetric stratigraphic sampling) and the sampling effort (number of specimens collected) represents sufficient data for diversity comparisons (Nanglu et al., 2020). Given its collection history, these same assumptions cannot be made about the Spence Shale.

Comparisons of the ecology between localities cannot be attempted, as the vastly different sampling intensities bias the faunal make-up, and the apparent taphonomic conditions. Due to the consolidated collection effort of soft tissue at Min-

ers Hollow, the proportion of soft-tissue fossils compared to biomineralized fossils is 20 times larger than the proportion of soft-tissue collected from Antimony Canyon, the locality with the second highest number of soft-tissue specimens (Figure 4; Table 2). From the outside, without knowledge of the bias, it would appear that Miners Hollow represents ideal taphonomic conditions, and a unique paleo-depositional environment among the Wells-ville Mountain sites. Establishing the causes of exceptional preservation in the Spence Shale means accounting for the collecting bias, not confusing it as a true signal.

The responsibility for mitigating measuring and resolving these biases is distributed among the collection stakeholders. Collection biases can be overcome by future fieldwork: conducting large-scale sampling, targeting previously under-studied taxa, and keeping bulk samples intact and associated. Preparation bias can be mitigated by incorporating new techniques, collecting in situ samples, and revisiting previously-collected material. These objectives are best achieved through a partnership of amateurs and academic collectors to inform best practices, and maintaining local knowledge of the formation. Digitization biases regarding the publication record can be mitigated by targeting this collection for data entry in the PBDB and updating with future publications. Taxonomic revisions in the KUMIP Specify database is the responsibility of the collection manager, and difficulties with taxonomic identities will be improved with implementing database standardizations, e.g., Darwin Core. Faunal composition biases created from a lack of study can be overcome by targeted interest and taxonomic revision efforts.

CONCLUSION

The Spence Shale collection at the University of Kansas is expanding beyond its original systematics and taxonomic origins and becoming increasingly important for larger discussions of macroevolution, taphonomy, and paleoecology. This mirrors the trajectory of natural history collections at large, which means those who use the collection need to be cognizant of its history and context. The Spence Shale collection does not accurately reflect the ecological and faunal composition present in the raw fossil record, as is the case with many similar collections (e.g., the Wheeler and Marjum Lagerstätten) (Kimmig personal observ.). This is due to collector bias, locality bias, researcher specialty bias, and positive and

negative study bias. The collectors donated fossils, which were ideal, or near ideal specimens, and those they thought would be scientifically important (i.e., soft-bodied material). Some localities are easier to access, transport, and remove material from, creating an unequal distribution of sampling efforts. Specialized interests of academic researchers drove publication history and taxonomic identification, creating well-established diversity in some taxonomic groups (i.e., the soft-bodied arthropods), and underrepresentation in others (i.e., the brachiopods). These biases do not completely undermine the usefulness and importance of the Spence Shale, but they should direct future efforts and scientific investigations. This work, outlining the collection context, history, and current status of the publication record and databases, sets up testable hypotheses to quantitatively analyze the effects of these biases. The Spence Shale fossil collection is a remarkable record of both a mid-Cambrian ecosystem and a devoted contribution to paleontology by a group of amateur collectors. Collection biases need to move beyond being considered anecdotal trivia, towards being quantitatively measured and recorded.

ACKNOWLEDGMENTS

We thank V. Gunther, P. Jamison, and R.A. Robison for the invaluable discussions on their experiences in the Spence Shale. We are also grateful to S. Olsen and P. Welsh for her helpful comments on an early draft of this work. The authors gratefully acknowledge the contributions of the reviewers, whose comments resulted in improvements and clarifications to the manuscript. This work was supported by an Association of Earth Science Clubs of Greater Kansas City Research Grant, a University of Kansas Biodiversity Institute Panorama Grant, a Geological Society of America Graduate Student Research Grant, and a Paleontological Society Kenneth E. and Annie Caster Student Research Award to AW. JK's work in the Spence Shale has been supported by a Paleontological Society Arthur James Boucot Grant, an Association of Earth Science Clubs of Greater Kansas City Research Grant and a Western Interior Paleontological Society Karl Hirsch Memorial Grant. JK thanks the USDA Forest Service for permits. The University of Kansas Biodiversity Institute is thanked for providing funds to pick up large parts of the Gunther and Jamison collections discussed herein.

REFERENCES

- Allentoft, M.E., Bunce, M., Scofield, R.P., Hale, M.L., and Holdaway, R.N. 2010. Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna. *Quaternary Science Reviews*, 29:753-762.
<https://doi.org/10.1016/j.quascirev.2009.11.022>
- Allison, P.A. 1988. Konservat-Lagerstätten: cause and classification. *Paleobiology*, 14:331-344.
- Allmon, W.A., Dietl, G.P., Hendricks, J.R., and Ross, R.M. 2018. Bridging the two fossil records: Paleontology's "big data" future resides in museum collections, p. 35-44. In Rosenberg, G.D. and Clary, R.M. (eds.), *Museums at the Forefront of the History and Philosophy of Geology: History Made, History in the Making*. Geological Society of America, USA.
[https://doi.org/10.1130/2018.2535\(03\)](https://doi.org/10.1130/2018.2535(03))
- Allmon, W.D. 2005. The importance of museum collections in paleobiology. *Paleobiology*, 31:1-5.
- Allmon, W.D. and Poulton, T.P. 2000. The value of fossil collections. *The Paleontological Society Special Publications*, 10:5-24. <https://doi.org/10.1017/S2475262200008923>
- Amori, G. and Gippoliti, S. 2000. What do mammalogists want to save? Ten years of mammalian conservation biology. *Biodiversity & Conservation*, 9:785-793.
<https://doi.org/10.1023/A:1008971823774>
- Ariño, A.H. 2010. Approaches to estimating the universe of natural history collections data. *Biodiversity Informatics*, 7:81-92. <https://doi.org/10.17161/bi.v7i2.3991>
- Babcock, L.E. and Robison, R.A. 1988. Taxonomy and paleobiology of some Middle Cambrian *Scenella* (Cnidaria) and hyolithids (Mollusca) from western North America. *The University of Kansas Paleontological Contributions*, 121:1-22.
- Baird, R.C. 2010. Leveraging the fullest potential of scientific collections through digitization. *Biodiversity Informatics*, 7:130-136. <https://doi.org/10.17161/bi.v7i2.3987>
- Bakker, F.T., Antonelli, A., Clarke, J.A., Cook, J.A., Edwards, S.V., Ericson, P.G.P., Faurby, S., Ferrand, N., Gelang, M., Gillespie, R., Irestedt, M., Lundin, K., Larsson, E., Matos-Maraví, P., Müller, J., von Proschwitz, T., Roderick, G., Schliep, A., Wahlberg, N., and Källersjö, M. 2020. The Global Museum: natural history collections and the future of evolutionary science and public education. *PeerJ*, 8:e8225. <https://doi.org/10.7717/peerj.8225>
- Bardack, D. 1965. Anatomy and evolution of chirocentrid fishes. *The University of Kansas Paleontological Contributions*, 10:1-86.
- Berents, P., Hamer, M., and Chavan, V. 2010. Towards demand driven publishing: approaches to the prioritization of digitization of natural history collections data. *Biodiversity Informatics*, 7:113-119. <https://doi.org/10.17161/bi.v7i2.3990>
- Bethoux, O. and Briggs, D.E. 2008. How *Gerarus* lost its head: stem?group Orthoptera and Paraneoptera revisited. *Systematic Entomology*, 33:529-547.
<https://doi.org/10.1111/j.1365-3113.2008.00419.x>
- Blagoderov, V., Kitching, I.J., Livermore, L., Simonsen, T.J., and Smith, V.S. 2012. No specimen left behind: industrial scale digitization of natural history collections. *ZooKeys*, 209:133-146.
<https://doi.org/10.3897/zookeys.209.3178>
- Briggs, D.E.G. and Robison, R.A. 1984. Exceptionally preserved nontrilobite arthropods and *Anomalocaris* from the Middle Cambrian of Utah. *The University of Kansas Paleontological Contributions*, 111:1-24.
- Briggs, D.E., Lieberman, B.S., Hendricks, J.R., Halgedahl, S.L., and Jarrard, R.D. 2008. Middle Cambrian arthropods from Utah. *Journal of Paleontology*, 82:238-254.
<https://doi.org/10.1666/06-086.1>
- Broce, J.S. and Schiffbauer, J.D. 2017. Taphonomic analysis of Cambrian vermiform fossils of Utah and Nevada, and implications for the chemistry of Burgess shale-type preservation of vermiform fossils. *Palaios*, 32:600-619. <https://doi.org/10.2110/palo.2017.011>
- Brock, F., Dee, M., Hughes, A., Snoeck, C., Staff, R., and Ramsey, C.B. 2018. Testing the effectiveness of protocols for removal of common conservation treatments for radiocarbon dating. *Radiocarbon*, 60:35-50. <https://doi.org/10.1017/RDC.2017.68>
- Butterfield, N.J. 2003. Exceptional fossil preservation and the Cambrian explosion. *Integrative and Comparative Biology*, 43:166-177. <https://doi.org/10.1093/icb/43.1.166>

- Campbell, D.P. 1974. Biostratigraphy of the *Albertella* and *Glossopleura* zones (Lower Middle Cambrian) of northern Utah and southern Idaho. Unpublished PhD Thesis, University of Utah, Salt Lake City, Utah, USA.
- Conway Morris, S. and Robison, R.A. 1986. Middle Cambrian priapulids and other soft-bodied fossils from Utah and Spain. *University of Kansas, Paleontological Contributions*, 117:1-22.
- Conway Morris, S., Halgedahl, S.L., Selden, P., and Jarrard, R.D. 2015a. Rare primitive deuterostomes from the Cambrian (Series 3) of Utah. *Journal of Paleontology*, 89:631-636. <https://doi.org/10.1017/jpa.2015.40>
- Conway Morris, S., Selden, P.A., Gunther, G., Jamison, P.G., and Robison, R.A. 2015b. New records of Burgess Shale-type taxa from the middle Cambrian of Utah. *Journal of Paleontology*, 89:411-423. <https://doi.org/10.1017/jpa.2015.26>
- Cook, J.A., Galbreath, K.E., Bell, K.C., Campbell, M.L., Carrière, S., Colella, J.P., Dawson, N.G., Dunnum, J.L., Eckerlin, R.P., Fedorov, V., and Greiman, S.E. 2017. The Beringian Coevolution Project: holistic collections of mammals and associated parasites reveal novel perspectives on evolutionary and environmental change in the North. *Arctic Science*, 3:585-617. <https://doi.org/10.1139/as-2016-0042>
- Cooper, N., Bond, A.L., Davis, J.L., Portela Miguez, R., Tomsett, L., and Helgen, K.M. 2019. Sex biases in bird and mammal natural history collections. *Proceedings of the Royal Society B*, 286:20192025. <https://doi.org/10.1098/rspb.2019.2025>
- Davis, E.B. and Pyenson, N.D. 2007. Diversity biases in terrestrial mammalian assemblages and quantifying the differences between museum collections and published accounts: a case study from the Miocene of Nevada. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 250:139-149. <https://doi.org/10.1016/j.palaeo.2007.03.006>
- Deiss, C. 1938. Cambrian formations and sections in part of Cordilleran Trough. *Bulletin of the Geological Society of America*, 49:1067-1168.
- Fernández-Jalvo, Y. and Monfort, M.D.M. 2008. Experimental taphonomy in museums: preparation protocols for skeletons and fossil vertebrates under the scanning electron microscopy. *Geobios*, 41:157-181. <https://doi.org/10.1016/j.geobios.2006.06.006>
- France, C.A., Giaccari, J.A., and Doney, C.R. 2015. The effects of Paraloid B-72 and Butvar B-98 treatment and organic solvent removal on $\delta^{13}\text{C}$, $\delta^{15}\text{N}$, and $\delta^{18}\text{O}$ values of collagen and hydroxyapatite in a modern bone. *American Journal of Physical Anthropology*, 157:330-338. <https://doi.org/10.1002/ajpa.22697>
- Frynta, D., Marešová, J., Landová, E., Lišková, S., Šimková, O., Tichá, I., Zelenková, M., and Fuchs, R. 2010. Are Animals in Zoos Rather Conspicuous than Endangered? Nova Science Publishers, Hauppauge, New York.
- Frynta, D., Peléšková, Š., Rádlová, S., Janovcová, M., and Landová, E. 2019. Human evaluation of amphibian species: a comparison of disgust and beauty. *The Science of Nature*, 106:41. <https://doi.org/10.1007/s00114-019-1635-8>
- Gaines, R.R., Briggs, D.E.G., and Yuanlong, Z. 2008. Cambrian Burgess Shale-type deposits share a common mode of fossilization. *Geology*, 36:755-758. <https://doi.org/10.1130/G24961A.1>
- Garson, D.E., Gaines, R.R., Droser, M.L., Liddell, W.D., and Sappenfield, A. 2012. Dynamic palaeoredox and exceptional preservation in the Cambrian Spence Shale of Utah. *Lethaia*, 45:164-177. <https://doi.org/10.1111/j.1502-3931.2011.00266.x>
- Gold, D.A., O'Reilly, S.S., Luo, G., Briggs, D.E., and Summons, R.E. 2016. Prospects for sterane preservation in sponge fossils from museum collections and the utility of sponge biomarkers for molecular clocks. *Bulletin of the Peabody Museum of Natural History*, 57:181-189. <https://doi.org/10.3374/014.057.0208>
- Golden, J. 2000. Institutional and individual orphaned collections. *The Paleontological Society Special Publications*, 10:37-42. <https://doi.org/10.1017/S2475262200008947>
- Groom, Q., Dillen, M., Hardy, H., Phillips, S., Willemse, L., and Wu, Z. 2019. Improved standardization of transcribed digital specimen data. *Database*, 2019:baz129. <https://doi.org/10.1093/database/baz129>
- Gunther, G. 2016. Death of discovery. *Journal of Paleontological Science*, 10. <https://www.aaps-journal.org/Death-of-Discovery.html>
- Habel, J.C., Husemann, M., Finger, A., Danley, P.D., and Zachos, F.E. 2014. The relevance of time series in molecular ecology and conservation biology. *Biological Reviews*, 89:484-492. <https://doi.org/10.1111/brv.12068>

- Haug, C., Reumer, J.W., Haug, J.T., Arillo, A., Audo, D., Azar, D., Baranov, V., Beutel, R., Charbonnier, S., Feldmann, R., and Foth, C. 2020. Comment on the letter of the Society of Vertebrate Paleontology (SVP) dated April 21, 2020 regarding "Fossils from conflict zones and reproducibility of fossil-based scientific data": the importance of private collections. *PalZ*, 94:413-429. <https://doi.org/10.1007/s12542-020-00522-x>
- Hedrick, B.P., Heberling, J.M., Meineke, E.K., Turner, K.G., Grassa, C.J., Park, D.S., Kennedy, J., Clarke, J.A., Cook, J.A., Blackburn, D.C., and Edwards, S.V. 2020. Digitization and the future of natural history collections. *BioScience*, 70:243-251. <https://doi.org/10.1093/biosci/biz163>
- Hendricks, J.R. 2013. Global distributional dynamics of Cambrian clades as revealed by Burgess Shale-type deposits, p. 35-43. In Harper, D.A.T. and Servais, T. (eds.), *Early Palaeozoic Biogeography and Palaeogeography*. Geological Society, London. <https://doi.org/10.1144/M38.4>
- Hendricks, J.R., Lieberman, B.S., and Stigall, A.L. 2008. Using GIS to study the paleobiogeography of soft-bodied Cambrian arthropods. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 264:163-175. <https://doi.org/10.1016/j.palaeo.2008.04.014>
- Holmes, M.W., Hammond, T.T., Wogan, G.O., Walsh, R.E., Labarbera, K., Wommack, E.A., Martins, F.M., Crawford, J.C., Mack, K.L., Bloch, L.M., and Nachman, M.W. 2016. Natural history collections as windows on evolutionary processes. *Molecular Ecology*, 25:864-881. <https://doi.org/10.1111/mec.13529>
- ICZN, 1999. *International Code of Zoological Nomenclature*. Fourth Edition. The International Trust for Zoological Nomenclature, London, UK.
- Isaac, N.J., Mallet, J., and Mace, G.M. 2004. Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology & Evolution*, 19:464-469. <https://doi.org/10.1016/j.tree.2004.06.004>
- Isaac, N.J. and Pocock, M.J. 2015. Bias and information in biological records. *Biological Journal of the Linnean Society*, 115:522-531. <https://doi.org/10.1111/bij.12532>
- Johnson, K.A. 2018. Significant effects of fossil preparation using sodium bicarbonate air abrasion on dental microwear. Unpublished PhD Thesis, South Dakota School of Mines and Technology, Rapid City, South Dakota, USA.
- Karim, T.S., Burkhalter, R., Farrell, Ú.C., Molineux, A., Nelson, G., Utrup, J., and Butts, S.H. 2016. Digitization workflows for paleontology collections. *Palaeontologia Electronica*, 19.3.4T:1-14. <https://doi.org/10.26879/566>
- Kimmig, J. and Pratt, B.R. 2016. Taphonomy of the middle Cambrian (Drumian) Ravens Throat River Lagerstätte, Rockslide Formation, Mackenzie Mountains, Northwest Territories, Canada. *Lethaia*, 49:150-169. <https://doi.org/10.1111/let.12135>
- Kimmig, J. and Selden, P.A. 2020. A new shell-bearing organism from the Cambrian Spence Shale of Utah. *Palaeoworld*. <https://doi.org/10.1016/j.palwor.2020.05.003>
- Kimmig, J. and Spielmann, J.A. 2011. Biologic factors influencing phytosaur (Archosauria: Phytosauridae) taxonomy: a prospectus, p. 289-294. In Sullivan, R.M., Lucas, S.G., and Spielmann, J.A. (eds.), *Fossil Record 3*. New Mexico Museum of Natural History and Science, Albuquerque, USA.
- Kimmig, J. and Strotz, L.C. 2017. Coprolites in mid-Cambrian (Series 2–3) Burgess Shale-type deposits of Nevada and Utah and their ecological implications. *Bulletin of Geosciences*, 92:297-309. <https://doi.org/10.3140/bull.geosci.1667>
- Kimmig, J., Strotz, L.C., and Lieberman, B.S. 2017. The stalked filter feeder *Siphusauctum lloydguntheri* n. sp. from the middle Cambrian (Series 3, Stage 5) Spence Shale of Utah: its biological affinities and taphonomy. *Journal of Paleontology*, 91:902-910. <https://doi.org/10.1017/jpa.2017.57>
- Kimmig, J., Strotz, L.C., Kimmig, S.R., Egenhoff, S.O., and Lieberman, B.S. 2019. The Spence Shale Lagerstätte: an important window into Cambrian biodiversity. *Journal of the Geological Society*, 176:609-619. <https://doi.org/10.1144/jgs2018-195>
- Kloss, T.J., Dornbos, S.Q., Chen, J.Y., McHenry, L.J., and Marenco, P.J. 2015. High-resolution geochemical evidence for oxic bottom waters in three Cambrian Burgess Shale-type deposits. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 440:90-95. <https://doi.org/10.1016/j.palaeo.2015.08.048>
- Koch, C.F. 1978. Bias in the published fossil record. *Paleobiology*, 4:367-372.

- Krause, Jr, R.A., Stempien, J.A., Kowalewski, M., and Miller, A.I. 2007. Body size estimates from the literature: utility and potential for macroevolutionary studies. *Palaios*, 22:60-73. <https://doi.org/10.2110/palo.2005.p05-122r>
- Le Cabec, A. and Toussaint, M. 2017. Impacts of curatorial and research practices on the preservation of fossil hominid remains. *Journal of Anthropological Sciences*, 95:1-28. <https://doi.org/10.4436/jass.95002>
- Legg, D.A. and Pates, S. 2017. A restudy of *Utahcaris orion* (Euarthropoda) from the Spence Shale (middle Cambrian, Utah, USA). *Geological Magazine*, 154:181-186. <https://doi.org/10.1017/S0016756816000789>
- Liddell, W.D., Wright, S.W., and Brett, C.E. 1997. Sequence stratigraphy and paleoecology of the Middle Cambrian Spence Shale in northern Utah and southern Idaho. *Brigham Young Geology Studies*, 42:59-78.
- Lieberman, B.S. and Kaesler, R.L. 2000. The scientific value of natural history museum collections: the concept of completeness. *The Paleontological Society Special Publications*, 10:109-118. <https://doi.org/10.1017/S2475262200009035>
- Lieberman, B.S. and Kimmig, J. 2018. Museums, paleontology, and a biodiversity science-based approach, p. 335-348. In Rosenberg, G.D. and Clary, R.M. (eds.), *Museums at the Forefront of the History and Philosophy of Geology: History Made, History in the Making*. Geological Society of America, USA. [https://doi.org/10.1130/2018.2535\(22\)](https://doi.org/10.1130/2018.2535(22))
- Lipps, J.H. 2018. Natural history museums: facilitating science literacy across the globe, p. 9-33. In Rosenberg, G.D. and Clary, R.M. (eds.), *Museums at the Forefront of the History and Philosophy of Geology: History Made, History in the Making*. Geological Society of America, USA. [https://doi.org/10.1130/2018.2535\(02\)](https://doi.org/10.1130/2018.2535(02))
- Lister, A.M. and Climate Change Research Group. 2011. Natural history collections as sources of long-term datasets. *Trends in Ecology & Evolution*, 26:153-154. <https://doi.org/10.1016/j.tree.2010.12.009>
- Lorimer, J. 2007. Nonhuman charisma. *Environment and Planning D: Society and Space*, 25:911-932. <https://doi.org/10.1068/d71j>
- Maletz, J. and Steiner, M. 2015. Graptolite (*Hemichordata*, *Pterobranchia*) preservation and identification in the Cambrian Series 3. *Palaontology*, 58:1073-1107. <https://doi.org/10.1111/pala.12200>
- Mallon, J.C. 2017. Recognizing sexual dimorphism in the fossil record: lessons from nonavian dinosaurs. *Paleobiology*, 43:495-507. <https://doi.org/10.1017/pab.2016.51>
- Mansfield, G.R. 1927. Geography, geology, and mineral resources of part of southeastern Idaho (No. 152). US Government Printing Office, Washington, D.C.
- Marcet, A., Haston, E., Groom, Q., Ariño, A., Chapman, A.D., Bakken, T., Braun, P., Dillen, M., Ernst, M., Escobar, A., Fichtmüller, D., Livermore, L., Nicolson, N., Paragamian, K., Paul, D., Petterson, L.B., Phillips, S., Plummer, J., Rainer, H., Rey, I., Robertson, T., Röpert, D., Santos, J., Uribe, F., Waller, J., and Wieczorek, J.R. 2020. Quality issues in georeferencing: from physical collections to digital data repositories for ecological research (Version 1.0). *Quality Issues in Georeferencing in Natural History Collections*, University of Warsaw, Warsaw, Poland. <https://doi.org/10.5281/zenodo.3734848>
- Marshall, C.R., Finnegan, S., Clites, E.C., Holroyd, P.A., Bonuso, N., Cortez, C., Davis, E., Dietl, G.P., Druckenmiller, P.S., Eng, R.C., and Garcia, C. 2018. Quantifying the dark data in museum fossil collections as palaeontology undergoes a second digital revolution. *Biology Letters*, 14:20180431. <https://doi.org/10.1098/rsbl.2018.0431>
- Maxey, G.B. 1958. Lower and middle Cambrian stratigraphy in northern Utah and southeastern Idaho. *Geological Society of America Bulletin*, 69:647-687.
- McGowan, A.J. and Dyke, G.J. 2009. A surfeit of theropods in the Moroccan Late Cretaceous? Comparing diversity estimates from field data and fossil shops. *Geology*, 37:843-846. <https://doi.org/10.1130/G30188A.1>
- Mulkerin, M.S. 2013. Elephants never forget, but museums do: investigating under-documented objects in museum collections. *Collections*, 9:149-166. <https://doi.org/10.1177/155019061300900203>
- Nanglu, K., Caron, J.-B., and Gaines, R.R. 2020. The Burgess Shale paleocommunity with new insights from Marble Canyon, British Columbia. *Paleobiology*, 46:58-81. <https://doi.org/10.1017/pab.2019.42>

- Nekola, J.C., Hutchins, B.T., Schofield, A., Najev, B., and Perez, K.E. 2019. *Caveat consumptor notitia museo*: let the museum data user beware. *Global Ecology and Biogeography*, 28:1722-1734. <https://doi.org/10.1111/geb.12995>
- Nelson, G. and Ellis, S. 2018. The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B*, 374:20170391. <https://doi.org/10.1098/rstb.2017.0391>
- Panchen, Z.A., Doubt, J., Kharouba, H.M., and Johnston, M.O. 2019. Patterns and biases in an Arctic herbarium specimen collection: implications for phenological research. *Applications in Plant Sciences*, 7:e01229. <https://doi.org/10.1002/aps3.1229>
- Planavsky, N., Hood, A., Tarhan, L., Shen, S., and Johnson, K. 2020. Store and share ancient rocks. *Nature*, 581:137-139. <https://www.nature.com/articles/d41586-020-01366-w>
- Ponder, W.F., Carter, G.A., Flemons, P., and Chapman, R.R. 2001. Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology*, 15:648-657. <https://doi.org/10.1046/j.1523-1739.2001.015003648.x>
- Pruss, S.B., Payne, J.L., and Westacott, S. 2015. Taphonomic bias of selective silicification revealed by paired petrographic and insoluble residue analysis. *Palaos*, 30:620-626. <https://doi.org/10.2110/palo.2014.105>
- Pyke, G.H. and Ehrlich, P.R. 2010. Biological collections and ecological/environmental research: a review, some observations and a look to the future. *Biological Reviews*, 85:247-266. <https://doi.org/10.1111/j.1469-185X.2009.00098.x>
- Resser, C.E. 1939. The Spence Shale and its fauna (with six plates). *Smithsonian Miscellaneous Collections*, Washington, D.C.
- Rigby, J.K. 1980. The new Middle Cambrian sponge *Vauxia magna* from the Spence Shale of northern Utah and taxonomic position of the *Vauxiidae*. *Journal of Paleontology*, 54:234-240.
- Roberts, D.E., Du Plessis, A., Steyn, J., Botha, L.R., Pityana, S., and Berger, L.R. 2012. An investigation of Laser Induced Breakdown Spectroscopy for use as a control in the laser removal of rock from fossils found at the Malapa hominin site, South Africa. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 73:48-54. <https://doi.org/10.1016/j.sab.2012.07.019>
- Robison, R.A. 1969. Annelids from the middle Cambrian Spence Shale of Utah. *Journal of Paleontology*, 43:1169-1173.
- Robison, R.A. and Babcock, L.E. 2011. Systematics, paleobiology, and taphonomy of some exceptionally preserved trilobites from Cambrian Lagerstätten of Utah. *Paleontological Contributions*, 5:1-47. <https://doi.org/10.17161/PC.1808.8543>
- Robison, R.A., Babcock L.E., and Gunther, V.G. 2015. Exceptional Cambrian fossils from Utah: a window into the age of trilobites. *Utah Geological Survey*, Salt Lake City, Utah.
- Saleh, F., Antcliffe, J.B., Lefebvre, B., Pittet, B., Laibl, L., Peris, F.P., Lustrì, L., Gueriau, P., and Daley, A.C. 2020a. Taphonomic bias in exceptionally preserved biotas. *Earth and Planetary Science Letters*, 529:115873. <https://doi.org/10.1016/j.epsl.2019.115873>
- Saleh, F., Lefebvre, B., Hunter, A.W., and Nohejlová, M. 2020b. Fossil weathering and preparation mimic soft tissues in Eocrinoid and Somasteroid echinoderms from the Lower Ordovician of Morocco. *Microscopy Today*, 28:24-28. <https://doi.org/10.1017/S1551929519001238>
- Schindel, D.E. and Cook, J.A. 2018. The next generation of natural history collections. *PLoS Biology*, 16:e2006125. <https://doi.org/10.1371/journal.pbio.2006125>
- Seilacher, A. 1970. Begriff und bedeutung der Fossil-Lagerstätten. *Neues Jahrbuch für Geologie und Paläontologie, Abhandlungen*, 1970:34-39.
- Seltmann, K., Lafia, S., Paul, D., James, S., Bloom, D., Rios, N., Ellis, S., Farrell, U., Utrup, J., Yost, M., Davis, E., Emery, R., Motz, G., Kimmig, J., Shirey, V., Sandall, E., Park, D., Tyrrell, C., Thackurdeen, R.S., Collins, M., O'Leary, V., Prestridge, H., Evelyn, C., and Nyberg, B. 2018. Georeferencing for research use (GRU): an integrated geospatial training paradigm for biocollections researchers and data providers. *Research Ideas and Outcomes*, 4:e32449. <https://doi.org/10.3897/rio.4.e32449>
- Sprinkle, J. 1973. Morphology and evolution of blastozoan echinoderms. *Special Publication*, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts.
- Sprinkle, J. and Collins, D. 2006. New eocrinoids from the Burgess Shale, southern British Columbia, Canada, and the Spence Shale, northern Utah, USA. *Canadian Journal of Earth Sciences*, 43:303-322. <https://doi.org/10.1139/e05-107>

- Suarez, A.V. and Tsutsui, N.D. 2004. The value of museum collections for research and society. *BioScience*, 54:66-74.
[https://doi.org/10.1641/0006-3568\(2004\)054\[0066:TVOMCF\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0066:TVOMCF]2.0.CO;2)
- Sumrall, C.D. and Sprinkle, J. 1999. *Ponticulocarpus*, a new cornute-grade stylophoran from the Middle Cambrian Spence Shale of Utah. *Journal of Paleontology*, 73:886-891.
<https://doi.org/10.1017/S0022336000040725>
- Vidal, D., Ortega, F., and Sanz, J.L. 2020. Sauropodomorph skeletal mounts as scientific devices for testing hypotheses. *Journal of Iberian Geology*, 46:177-193.
<https://doi.org/10.1007/s41513-020-00122-3>
- Walcott, C.D. 1908. *Cambrian Geology and Palaeontology*. Smithsonian Museum, Miscellaneous Collections, 53, Washington, D.C.
- Wandeler, P., Hoeck, P.E., and Keller, L.F. 2007. Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution*, 22:634-642.
<https://doi.org/10.1016/j.tree.2007.08.017>
- Wehi, P.M., Whaanga, H., and Trewick, S.A. 2012. Artefacts, biology and bias in museum collection research. *Molecular Ecology*, 21:3103-3109.
<https://doi.org/10.1111/j.1365-294X.2012.05589.x>
- Wen, R., Babcock, L.E., Peng, J., and Robison, R.A. 2019. New edrioasteroid (Echinodermata) from the Spence Shale (Cambrian), Idaho, USA: further evidence of attachment in the early evolutionary history of edrioasteroids. *Bulletin of Geosciences*, 94:115-124.
<https://doi.org/10.3140/bull.geosci.1730>
- Whitaker, A.F., Jamison, P.G., Schiffbauer, J.D., and Kimmig, J. 2020. Re-description of the Spence Shale palaeoscolecid in light of new morphological features with comments on palaeoscolecid taxonomy and taphonomy. *Paläontologische Zeitschrift*, 94:661–674.
<https://doi.org/10.1007/s12542-020-00516-9>
- Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., and Vieglaiss, D. 2012. Darwin Core: an evolving community developed biodiversity data standard. *PLoS ONE*, 7:e29715. <https://doi.org/10.1371/journal.pone.0029715>
- Willoughby, R.H. and Robison, R.A. 1979. Medusoids from the Middle Cambrian of Utah. *Journal of Paleontology*, 53:494-500.
- Winker, K. 2004. Natural history museums in a postbiodiversity era. *BioScience*, 54:455-459.
[https://doi.org/10.1641/0006-3568\(2004\)054\[0455:NHMIAP\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0455:NHMIAP]2.0.CO;2)
- Wright, S.H. 1999. Sequence stratigraphy and paleoecology of the Middle Cambrian Spence Shale Member of the Langston Formation of northeastern Utah and southeastern Idaho. Unpublished MS Thesis, Utah State University, Logan, Utah, USA.
- Zhang, Z., Augustin, M., and Payne, J.L. 2015. Phanerozoic trends in brachiopod body size from synoptic data. *Paleobiology*, 41:491-501. <https://doi.org/10.1017/pab.2015.12>

APPENDIX 1.

Specimens from the Spence Shale Lagerstätte in the KUMIP collection. (Available at <https://palaeo-electronica.org/content/2020/3238-collections-biases>.)

APPENDIX 2.

Genera lists based on Kimmig et al., 2019, and the PBDB and iDigBio. data aggregators. (Available at <https://palaeo-electronica.org/content/2020/3238-collections-biases>.)

APPENDIX 3.

Publication record on the Spence Shale. (Available at <https://palaeo-electronica.org/content/2020/3238-collections-biases>.)