Received: 20 August 2019 | Revised: 17 January 2020 | Accepted: 27 March 2020

DOI: 10.1111/gbi.12397

REPORT

gebiology WILEY

Chemical signatures of soft tissues distinguish between vertebrates and invertebrates from the Carboniferous Mazon Creek Lagerstätte of Illinois

Victoria E. McCoy[1,2] | Jasmina Wiemann[3] | James C. Lamsdell[4] | Christopher D. Whalen[3] | Scott Lidgard[5] | Paul Mayer[5] | Holger Petermann[6] | Derek E. G. Briggs[3,7]

[1]Department of Geosciences, University of Wisconsin, Milwaukee, Milwaukee, WI, USA

[2]Institute of Geosciences, University of Bonn, Bonn, Germany

[3]Department of Geology and Geophysics, Yale University, New Haven, CT, USA

[4]Department of Geology and Geography, West Virginia University, Morgantown, WV, USA

[5]Field Museum of Natural History, Chicago, IL, USA

[6]Department of Earth Sciences, Denver Museum of Nature and Science, Colorado Boulevard, Denver, CO, USA

[7]Yale Peabody Museum of Natural History, New Haven, CT, USA

Correspondence
Victoria E. McCoy, Department of Geosciences, University of Wisconsin, Milwaukee, Milwaukee, WI, USA.
Email: mccoyv@uwm.edu

Jasmina Wiemann, Department of Geology and Geophysics, Yale University, 210 Whitney Avenue, New Haven, CT 06511, USA.
Email: jasmina.wiemann@yale.edu

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: RU 665/13-1

Abstract

The chemical composition of fossil soft tissues is a potentially powerful and yet underutilized tool for elucidating the affinity of problematic fossil organisms. In some cases, it has proven difficult to assign a problematic fossil even to the invertebrates or vertebrates (more generally chordates) based on often incompletely preserved morphology alone, and chemical composition may help to resolve such questions. Here, we use in situ Raman microspectroscopy to investigate the chemistry of a diverse array of invertebrate and vertebrate fossils from the Pennsylvanian Mazon Creek Lagerstätte of Illinois, and we generate a ChemoSpace through principal component analysis (PCA) of the in situ Raman spectra. Invertebrate soft tissues characterized by chitin (polysaccharide) fossilization products and vertebrate soft tissues characterized by protein fossilization products plot in completely separate, non-overlapping regions of the ChemoSpace, demonstrating the utility of certain soft tissue molecular signatures as biomarkers for the original soft tissue composition of fossil organisms. The controversial problematicum *Tullimonstrum*, known as the Tully Monster, groups with the vertebrates, providing strong evidence of a vertebrate rather than invertebrate affinity.

KEYWORDS

chitin, chordate, in situ Raman spectroscopy, keratin, protein fossilization products, Tully Monster

1 | INTRODUCTION

Early fossil vertebrates, particularly those without a biomineralized skeleton, can be difficult to interpret based on their preserved soft tissue morphology alone. A lack of diagnostic characters may compromise phylogenetic assignments and even make it difficult to determine whether affinities lie with invertebrates or vertebrates (Parry et al., 2018; Purnell et al., 2018). However, preserved soft tissues may retain compositional information that can be used to address such questions (Alleon et al., 2017; Briggs & Summons, 2014; Wiemann, Fabbri, et al., 2018). The structural tissues of many invertebrates are composed almost entirely of the polysaccharide (=long-chain sugar) chitin, whereas most vertebrate structural tissues are based on a protein scaffold. Polysaccharides and proteins fossilize primarily through oxidative crosslinking (Wiemann, Fabbri, et al., 2018), and differences in the concentrations of polysaccharides

and proteins between tissue samples yield different compositional trends in the resulting fossilization products.

Fossil soft tissue morphologies preserved as carbonaceous films are composed of a variety of compounds including N- and S-heterocyclic polymers as well as their carbonyl (C=O)- and carboxyl (COOH)-rich peroxidation products. N-heterocycles form as oxidative crosslinks through the reaction of polysaccharide-derived reactive carbonyl species with available amine groups (which can derive from glycosaminoglycans for example), so they are most prevalent in the fossilization products of tissues with high polysaccharide concentrations (Wiemann, Fabbri, et al., 2018). S-heterocycles, thiols and thioethers, which require the protein-derived amino acid cysteine for crosslinking, are most common in the fossilization products of tissues with high protein concentrations (Wiemann, Fabbri, et al., 2018). Protein-rich tissues can also degrade via peroxidation processes, resulting in carbonyl (C=O)- and carboxyl (COOH)-rich protein fossilization products (Wiemann, Fabbri, et al., 2018; Wiemann, Yang, & Norell, 2018). These fossilization products can be reliably identified via their characteristic Raman signals, which reflect diagnostic C-N, C-S, and C-O and vibrations (Wiemann, Fabbri, et al., 2018). They provide potential biomarkers for identifying the original composition of fossil soft tissues and therefore are a valuable tool for assigning problematic or incomplete fossils to major animal groups (Wiemann, Crawford, & Briggs, in press).

The well-known Pennsylvanian Mazon Creek Lagerstätte (~307-309 Ma) of Illinois has yielded numerous fossiliferous concretions, exposed as a result of strip mining for coal in the last century (Clements, Purnell, & Gabbott, 2019). Concretions are known to provide a favorable environment for biomolecule preservation (Grice, Holman, Plet, & Tripp, 2019). Here, we use high-resolution in situ Raman microspectroscopy following the established protocol for fossils (Wiemann, Fabbri, et al., 2018; Wiemann et al., in press) to explore the fossilization products of a range of morphological features of various invertebrate and vertebrate taxa from this site (n = 32 samples from 20 specimens; see SI for method details). These data allow us to evaluate the potential of N- and S-heterocycle signatures in Raman spectra to provide information about the original composition of fossil soft tissues. Representative tissues include a mollusk radula (originally chitinous), arthropod exoskeletons (originally chitinous), jawless fish "teeth" from *Mayomyzon* and *Gilpichthys* (originally keratinous and/or collagenous), a chondrichthyan egg case (originally keratinous and collagenous), annelid jaws (originally collagenous) and setae (originally chitinous), and soft tissues from the problematic *Tullimonstrum* (Richardson, 1966) (Table S1). All these samples yield a complex mixture of organic compounds, most of which are products of diagenetic processes. In order to interpret the results, we used principal component ChemoSpace analysis (Wiemann, Fabbri, et al., 2018; Wiemann, Yang, et al., 2018), discriminant analyses, and spectral assessments to characterize preserved differences in the fossil soft tissues of major groups. We use these results as evidence of the taxonomic affinities of fossil organisms.

2 | RESULTS

In situ Raman microspectroscopy identified a range of inorganic compounds associated with diagenetic mineral precipitation together with organic compounds associated with morphological soft tissue features (Figure 1). The averaged spectra for known invertebrate and vertebrate taxa differ in the relative band signal intensities assigned to N- and S-heterocycles (Figure 2). Exploiting these characteristic differences, a discriminant analysis including only taxa of known phylogenetic affinity (Figure 3a) revealed that increased amounts of N-heterocycles are a characteristic feature of invertebrate fossil soft tissues, which were originally composed primarily of chitin, whereas increased carbonyl, S-heterocycle, and thiol/thioether signals are characteristic of vertebrate fossil soft tissues, which were originally composed primarily of proteins.

A principal component analysis (PCA), including the problematic *Tullimonstrum* samples (Figure 3b), allowed us to plot a ChemoSpace reflecting the major sources of compositional variation (as represented by the principal component axes or PCs) among the fossil samples. The loadings for the principal components are displayed in the Supporting Information Material. PC1 (43%), which is not shown, is dominated by a taphonomic signal representing crystallographic units and ions associated with secondarily precipitated minerals such as clay minerals, iron oxides, and inorganic sulfides (all positive loadings, with a major influence over a Raman shift from 600 to 1,100 cm^{-1}). Thus, although PC1 reveals information about mineral precipitation during fossilization, it does not reflect original tissue chemistry and is not considered further here.

PC2 (31%) and PC3 (11%), in contrast, represent organic constituents of the fossils. Most of these organic compounds are N-heterocyclic in composition and represent the crosslinked fossilization products of proteins, polysaccharides (e.g., chitin), and lipids (see Raman band at 1,550–1,610 cm^{-1}, compare Wiemann, Fabbri, et al., 2018). PC2 represents the relative signal intensity of N-heterocycles, one of the key factors for discriminating between invertebrates and vertebrates (Figure 3a). PC2 (Figure 3b) separates samples based on taxonomy and tissue chemistry, as predicted by qualitative spectral comparison (Figure 2) and the discriminant analysis (Figure 3a). All samples representing originally proteinaceous tissues, including keratinous/collagenous tissues from the vertebrates *Mayomyzon*, *Gilpichthys*, and a chondrichthyan egg capsule, form a distinct group (Figure 3b, blue convex hull). These fossils are characterized by less abundant N-heterocycles (relative to carbonyls, O-heterocycles, and S-heterocycles/thiols/thioethers), a characteristic feature of protein fossilization products (PFPs) (Figure 3a). All *Tullimonstrum* soft tissue samples (Figure 3b; green convex hull) group with vertebrate soft tissues (Figure 3b; blue convex hull for known fossil vertebrate samples and dotted convex hull for all tissues inferred to contain vertebrate protein fossilization products). The crosslinked collagenous (protein) jaws of the polychaete *Esconites* also plot with protein-derived fossilization products. All samples representing tissues that were originally composed of chitin, including the arthropod exoskeletons and the molluskan radula,

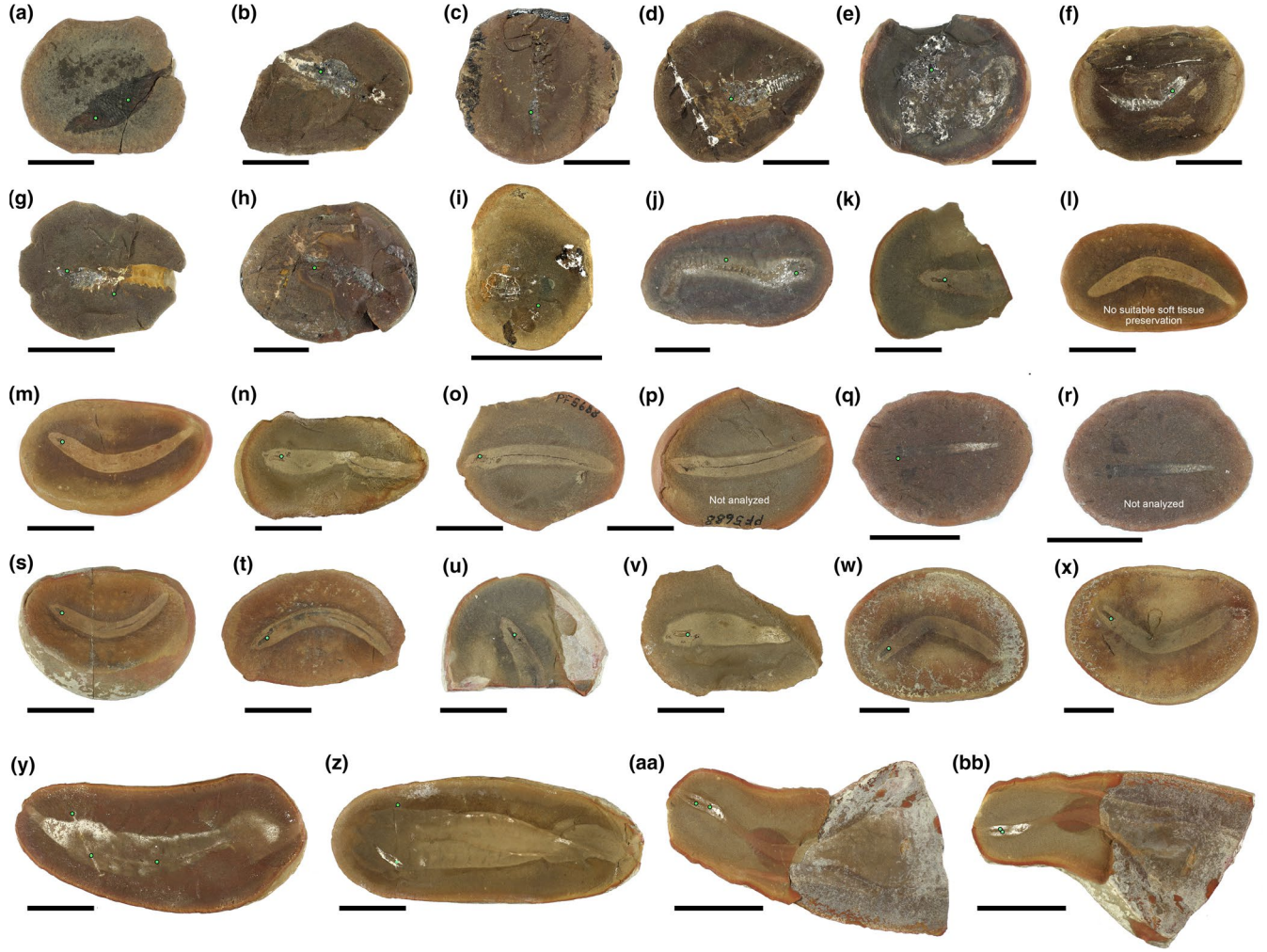

FIGURE 1 Specimen photographs. The green dots indicate areas chosen for point spectra analyses. Numbers in brackets indicate the sample identity in Figure 3. (a) YPM (Temp.) 7764 *Palaeoxyris* chondrichthyan egg case [26, 28], (b) YPM 19754 *Paleocaris typus* [3], (c) YPM 18891 *Acanthotelson stimpsoni* [5], (d) YPM 53990 *Acanthotelson stimpsoni* [2], (e) YPM 50693 *Euproops danae* [4], (f) YPM 52348 *Acanthotelson stimpsoni* [9], (g) YPM 53988 *Acanthotelson stimpsoni* [6], (h) YPM 18910 *Acanthotelson stimpsoni* [8], (i) FMNH PE 39344 *Paleocadmus herdinae* [7], (j) YPM 539746 *Esconites zelus* [1, 32], (k) FMNH PE 8904 *Gilpichthys* sp. [14], (l) FMNH PF 8449 *Gilpichthys* sp. [imaged to illustrate the lack of suitable soft tissues]; (m) FMNH PF 8449 *Gilpichthys* sp. [29], (n) FMNH PF 8348 *Gilpichthys* sp. [17], (o) FMNH PF 5688 *Mayomyzon* sp. part [11], (p) FMNH PF 5688 *Mayomyzon* sp. counterpart, not analyzed, (q) FMNH PF 8167 *Mayomyzon* sp. part [27], (r) FMNH PF 8167 *Mayomyzon* sp. counterpart [not analyzed], (s) FMNH PF 8488 *Gilpichthys* sp. [13], (t) FMNH PF 8488 *Gilpichthys* sp. [31], (u) FMNH PF 8904 *Gilpichthys* sp. [30], (v) FMNH PF 8348 *Gilpichthys* sp. [19], (w) FMNH PF 15354 *Gilpichthys* sp. [20], (x) FMNH PF 15354 *Gilpichthys* sp. [21], (y) FMNH PE 40113 *Tullimonstrum gregarium* [12, 23, 25], (z) FMNH PE 7063 *Tullimonstrum gregarium* [18, 24], (aa, bb) FMNH PE 22056 *Tullimonstrum gregarium* part and counterpart [10, 15, 16, 22]. Additional details for each specimen are listed in Table S1. The scale bars equals 1 cm

form a separate, non-overlapping group characterized by more abundant N-heterocycles (Figure 3b, orange convex hull), which is characteristic of chitin fossilization products. Despite some evidence for compositional variation within individual fossils (e.g., FMNH PE 22056, Figure 3b), and between part and counterpart, the distribution of samples in the ChemoSpace reflects their original tissue composition. The grouping of *Tullimonstrum* with vertebrate soft tissues (Figure 3b, overlapping green and blue convex hulls) reflects an originally proteinaceous composition for the eyes, eye bar, teeth, buccal apparatus, integument, and notochord; in many invertebrate phyla, particularly mollusks, arthropods, or annelids, many of these tissues would be expected to have a chitinous composition.

3 | DISCUSSION

Our analyses clearly show that the Mazon Creek fossils retain a compositional signature of proteinaceous or chitinous tissues that can be utilized to separate vertebrates from some major invertebrate phyla (Figure 2; Figure 3a,b). The relative abundance and composition of N-, S-, and O-heterocycles in the soft tissues of Mazon Creek fossils retain tissue-specific signals (Figure 3a,b). Oxidative crosslinks form through the early diagenetic reaction of particular amino acid residues in proteins (including amine- and thiol-bearing amino acid residues) with reactive carbonyl species (RCS) (Vistoli et al., 2013; Wiemann, Fabbri, et al., 2018). Such RCS can be derived from either

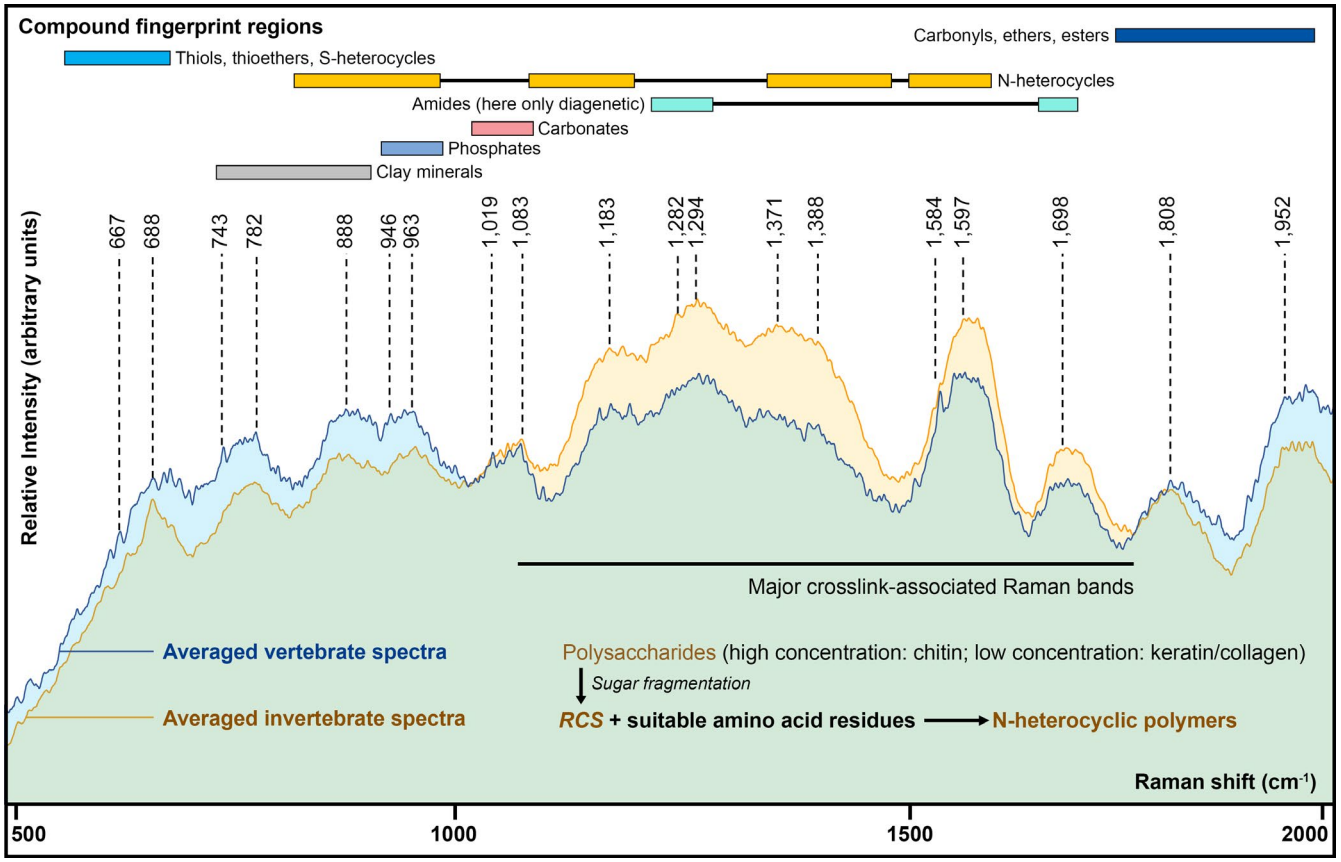

FIGURE 2 In situ Raman microspectroscopy of Mazon Creek fossil soft tissues. The invertebrate spectrum (orange line and shading) is averaged over 9 individual spectra for visibly preserved arthropod cuticle, a mollusk radula, and polychaete bristles. The vertebrate spectrum (blue line and shading) is averaged over four individual spectra for visibly preserved vertebrate soft tissues. Previous method testing (Wiemann et al., in press) revealed that trends in the composition of fossil soft tissues do not significantly change when averaging 4 spectra versus comparing 9 or more spectra. All individual spectra were acquired with a 532 nm laser over a range from 500 to 2,000 cm^{-1} Raman shift, with 5 accumulations, and 10-s exposure time (see "Methods" section in the SI for details)

lipids or polysaccharides in biological tissues (Vistoli et al., 2013; Wiemann, Fabbri, et al., 2018). Depending on the relative abundance of proteins with suitable amino acid residues, and sugars as a source of RCS, heterogeneous polymers with different ratios of N- and S-heterocycles form in fossil soft tissues (Wiemann, Fabbri, et al., 2018). Thus, the molecular composition of fossil soft tissues can provide evidence of the affinities of problematic fossils including, for example, *Tullimonstrum*.

Since its discovery more than fifty years ago, *Tullimonstrum* has been compared to mollusks, chordates, annelids, and arthropods, although the two most favored hypotheses considered it a mollusk (Foster, 1979; Rogers et al., 2019; Sallan et al., 2017) or chordate (Beall, 1991; Clements et al., 2016; Denton & Goolsby, 2018; McCoy et al., 2016). The radulae of mollusks are chitinous, whereas the keratinous "teeth" of jawless vertebrates have a proteinaceous scaffold. If the affinity of *Tullimonstrum* lay with mollusks or arthropods, both of which have chitinous mouthparts, we would expect an overlap with the invertebrate (i.e., mollusk–arthropod–annelid setae) group. However, no such overlap is observed. Instead, we record a substantial overlap between *Tullimonstrum* tissues and the tissues of the vertebrate group (Figure 3b). The small offset between *Tullimonstrum* fossil tissues and the identified vertebrate samples represents minor compositional differences in the reaction educts (i.e., reactive amino acid residues in different proteins) (Wiemann et al., in press). Such differences in fossil vertebrate tissues have been shown to be phylogenetically informative (Wiemann et al., in press). This, in combination with the absence of an overlap with the chitinous tissues of the invertebrate group, is consistent with a vertebrate-like proteinaceous composition for *Tullimonstrum* teeth and favors a vertebrate affinity for *Tullimonstrum*.

4 | CONCLUSIONS

The phylogenetic position of *Tullimonstrum* has been controversial since it was first described (Richardson, 1966). Recent anatomical results that strongly suggest a vertebrate identity (Clements et al., 2016; McCoy et al., 2016) have been considered equivocal (Rogers et al., 2019; Sallan et al., 2017). The molecular composition of *Tullimonstrum* soft tissues, analyzed here in a comparative framework, supports a vertebrate

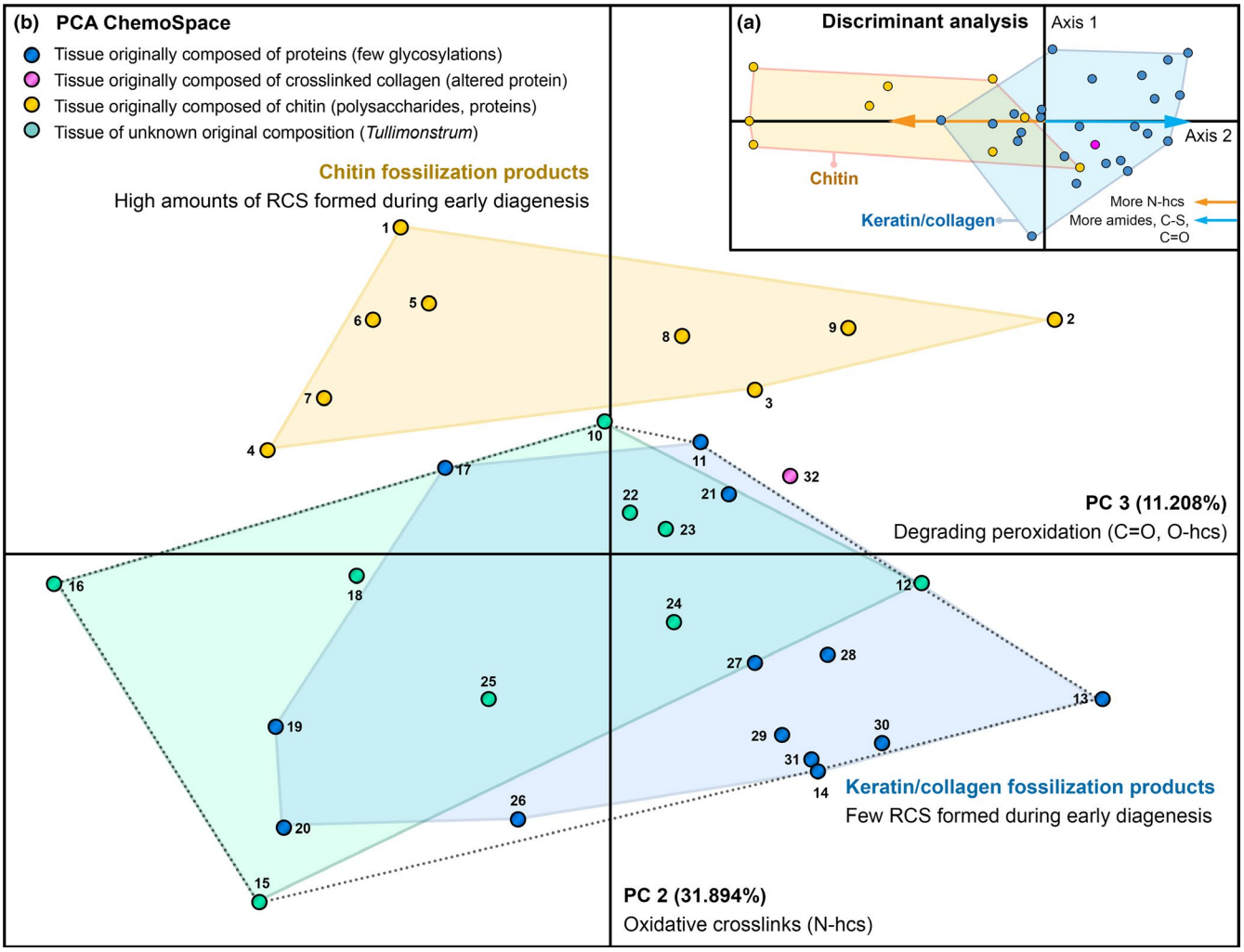

FIGURE 3 Analysis of phylogenetically informative soft tissue composition in Mazon Creek fossils. Blue dots represent whole spectra obtained for fossil soft tissues from vertebrates indicating an originally proteinaceous composition. The pink dot denotes a spectrum for fossil polychaete jaws originally composed of crosslinked collagen. Orange dots represent whole spectra acquired for fossil soft tissues from the known invertebrates sampled, originally chitin. Green dots represent *Tullimonstrum* soft tissues of unknown original composition. The orange convex hull groups tissues of originally chitinous composition, whereas the blue convex hull groups tissues of originally proteinaceous composition (excluding the polychaete jaws originally composed of crosslinked collagen). The green convex hull groups all *Tullimonstrum* samples, and the dotted convex hull combines all inferred vertebrate samples with protein fossilization products in this ChemoSpace. All methods are extensively described in the SI section "Methods." Biplot, loadings, and scree plot can be found in the SI. (a) Discriminant analysis, identifying increased relative amounts of N-heterocycles as a distinct feature of chitin fossilization products, and identifying amides, S-heterocycles, thiols, and thioethers, as well as peroxidation, as characteristic features present in protein fossilization products. Abbreviations and chemical notations: N-hcs = N-heterocycles; O-hcs = O-heterocycles, C-S = organic sulfur (S-heterocycles, thiol, thioethers); C=O = carbonyls. (b) Principal component ChemoSpace PC2 (33%) and PC3 (11%) (for loadings of PC1-3, see Supporting Information; PC1 does not sort samples for organic phase signals). PC2 sorts for the abundance of N-heterocycles (discriminating factor identified in a), and PC3 sorts for the relative abundance of peroxidation products. See Table S1 or number in brackets in the caption of Figure 1 for sample identifications

affinity. The ancient disparity of soft-bodied jawless vertebrates (e.g., lampreys, hagfish) may have been much greater than presently known.

ACKNOWLEDGMENTS

We thank Tom Testa for donating his collection. Yale University funded the chemical analyses. VEM was funded by a Deutsche Forschungsgemeinschaft grant (RU 665/13-1, project number 396637283). This is contribution no. 11 of the DFG Research Unit 2685 "The Limits of the Fossil Record: Analytical and Experimental Approaches to Fossilization." We also thank three anonymous reviewers for helpful comments.

AUTHOR CONTRIBUTIONS

VEM planned and coordinated the study, carried out measurements, taphonomic assessments, and morphological analyses, and wrote the first draft of the manuscript. JW designed the Raman spectroscopy protocol, performed the chemical analyses, ran the PCA and discriminant analysis, photographed specimens, prepared the

figures, and wrote parts of the manuscript. JCL reviewed previous phylogenetic analyses. CDW investigated the cranial morphology. SL and PM carried out preliminary specimen investigations, selected specimens for chemical analysis, and photographed specimens. HP assisted with Raman analysis. DEGB helped plan and coordinate the study. All authors edited and approved the final manuscript.

ORCID

Victoria E. McCoy https://orcid.org/0000-0002-5655-0381
Jasmina Wiemann https://orcid.org/0000-0003-3581-1711
Derek E. G. Briggs https://orcid.org/0000-0003-0649-6417

REFERENCES

Alleon, J., Bernard, S., Le Guillou, C., Daval, D., Skouri-Panet, F., Kuga, M., & Robert, F. (2017). Organic molecular heterogeneities can withstand diagenesis. *Scientific Reports*, *7*, 1508. https://doi.org/10.1038/s41598-017-01612-8

Beall, B. S. (1991). The Tully Monster and a new approach to analyzing problematica. In A. M. Simonetta, & S. Conway Morris (Eds.), *The early evolution of Metazoa and the significance of problematic taxa* (pp. 271–285). Cambridge, UK: Cambridge University Press.

Briggs, D. E. G., & Summons, R. E. (2014). Ancient biomolecules: Their origins, fossilization, and role in revealing the history of life. *BioEssays*, *36*, 482–490. https://doi.org/10.1002/bies.201400010

Clements, T., Dolocan, A., Martin, P., Purnell, M. A., Vinther, J., & Gabbott, S. E. (2016). The eyes of *Tullimonstrum* reveal a vertebrate affinity. *Nature*, *532*, 500–503. https://doi.org/10.1038/nature17647

Clements, T., Purnell, M. A., & Gabbott, S. E. (2019). The Mazon Creek Lagerstätte: A diverse late Palaeozoic ecosystem entombed within siderite concretions. *Journal of the Geological Society*, *176*, 1–11.

Denton, J. S. S., & Goolsby, E. W. (2018). Measuring inferential importance of taxa using taxon influence indices. *Ecology and Evolution*, *8*, 4484–4494. https://doi.org/10.1002/ece3.3941

Foster, M. W. (1979). A reappraisal of *Tullimonstrum gregarium*. In M. H. Nitecki (Ed.), *Mazon Creek fossils* (pp. 269–301). Academic Press, New York.

Grice, K., Holman, A. I., Plet, C., & Tripp, M. (2019). Fossilised biomolecules and biomarkers in carbonate concretions from Konservat-Lagerstätten. *Minerals*, *9*, 158. https://doi.org/10.3390/min9030158

McCoy, V. E., Saupe, E. E., Lamsdell, J. C., Tarhan, L. G., McMahon, S., Lidgard, S., … Briggs, D. E. G. (2016). The "Tully monster" is a vertebrate. *Nature*, *532*, 496–499. https://doi.org/10.1038/nature16992

Parry, L. A., Smithwick, F., Nordén, K. K., Saitta, E. T., Lozano-Fernandez, J., Tanner, A. R., … Vinther, J. (2018). Soft-bodied fossils are not simply rotten carcasses - Toward a holistic understanding of exceptional fossil preservation. *BioEssays*, *40*, 1700167. https://doi.org/10.1002/bies.201700167

Purnell, M. A., Donoghue, P. J. C., Gabbott, S. E., McNamara, M. E., Murdock, D. J. E., & Sansom, R. S. (2018). Experimental analysis of soft-tissue fossilization: Opening the black box. *Palaeontology*, *61*, 317–323. https://doi.org/10.1111/pala.12360

Richardson, E. S. Jr (1966). Wormlike fossil from the Pennsylvanian of Illinois. *Science*, *151*, 75–76. https://doi.org/10.1126/science.151.3706.75-a

Rogers, C. S., Astrop, T. I., Webb, S. M., Ito, S., Wakamatsu, K., & McNamara, M. E. (2019). Synchrotron X-ray absorption spectroscopy of melanosomes in vertebrates and cephalopods: Implications for the affinity of *Tullimonstrum*. *Proceedings of the Royal Society B*, *286*, 20191649.

Sallan, L., Giles, S., Sansom, R. S., Clarke, J. T., Johanson, Z., Sansom, I. J., & Janvier, P. (2017). The "Tully Monster" is not a vertebrate: Characters, convergence and taphonomy in Palaeozoic problematic animals. *Palaeontology*, *60*, 149–157. https://doi.org/10.1111/pala.12282

Vistoli, G., De Maddis, D., Cipak, A., Zarkovic, N., Carini, M., & Aldini, G. (2013). Advanced glycoxidation and lipoxidation end products (AGEs and ALEs): An overview of their mechanisms of formation. *Free Radical Research*, *47*, 3–27. https://doi.org/10.3109/10715762.2013.815348

Wiemann, J., Crawford, J. M., & Briggs, D. E. G. (in press). Phylogenetic and physiological signals in metazoan fossil biomolecules. *Science Advances*.

Wiemann, J., Fabbri, M., Yang, T.-R., Stein, K., Sander, P. M., Norell, M. A., & Briggs, D. E. G. (2018). Fossilization transforms vertebrate hard tissue proteins into N-heterocyclic polymers. *Nature Communications*, *9*, 4741. https://doi.org/10.1038/s41467-018-07013-3

Wiemann, J., Yang, T.-R., & Norell, M. A. (2018). Dinosaur egg colour had a single evolutionary origin. *Nature*, *563*, 555–558. https://doi.org/10.1038/s41586-018-0646-5

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: McCoy VE, Wiemann J, Lamsdell JC, et al. Chemical signatures of soft tissues distinguish between vertebrates and invertebrates from the Carboniferous Mazon Creek Lagerstätte of Illinois. *Geobiology*. 2020;18:560–565. https://doi.org/10.1111/gbi.12397