nature research

Corresponding author(s): Mark Norell, Jasmina Wiemann, Matteo Fabbri

Last updated by author(s): Mar 21, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No customised computer code was used to collect data for this study. All Raman data were collected using the LabSpec 5 software, and all histological sections were imaged using Leica LAS Core Software.
Data analysis	Acquired Raman spectra were standard-processed in SpectraGryph 1.2 spectroscopic software (freeware). Spectral data were stored as taxon-character matrices in Microsoft Excel (Office 365). The DA and Cluster Analysis were run in the PAST 3 software (freeware). An ancestral state reconstruction was run in Mesquite 3.40. Compound figures were created in Photoshop CS5.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The authors declare that all relevant Raman spectra and eggshell codings are available within the manuscript and its Supplementary Information. Materials can be made available upon request to the corresponding author.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We present mineralogical, organochemical, and ultrastructural evidence for an originally non-biomineralized, soft-shelled nature of exceptionally preserved ornithischian Protoceratops (n=1, spectra with 3 replicates throughout thickness) and basal sauropodomorph Mussaurus eggs (n=1, spectra with 3 replicates throughout thickness). Statistical evaluation (Discriminant Analysis + Hierarchical Cluster Analysis) of in situ organic phase Raman spectra (10 accumulation = technical replicates for each eggshell spectrum) obtained for a representative set of hard- and soft-shelled, fossil and extant diapsid eggshells (total n=24 + n=1 Protoceratops + n=1 Mussaurus), clusters the originally organic, but secondarily phosphatized Protoceratops and the carbonaceous Mussaurus eggshells with soft eggshells. Histology corroborates the organic composition of these two soft-shelled dinosaur eggs, revealing a stratified arrangement resembling soft turtle eggshell. An ancestral state reconstruction (maximum parsimony & maximum likelihood, total n=112 specimens) of composition and ultrastructure compared eggshells from Protoceratops and Mussaurus to those from other archosaurs, and revealed that the first dinosaur egg was soft-shelled. The calcified dinosaur egg evolved at least three times independently throughout the Mesozoic, explaining the bias towards eggshells of highly derived dinosaurs in the fossil record.
Research sample	We used exceptionally preserved ornithischian Protoceratops (n=1, spectra with 3 replicates throughout thickness) and basal sauropodomorph Mussaurus eggs (n=1, spectra with 3 replicates throughout thickness).This material complemented with a total of n=24 hard- and soft-shelled, fossil and extant diapsid eggshells, and eventually contextualised in an Ancestral State Reconstruction covering a total of n=112 specimens (literature-based).
Sampling strategy	No sample size calculation was performed, and Protoceratops and Mussaurus present the only currently known soft dinosaur eggshells. Other hard- and soft-shelled, fossil and extant diapsid eggshells were selected for a representative coverage of the diapsid phylogeny, and the Ancestral State Reconstruction includes all fossil eggshell types published to our knowledge.
Data collection	Mark Norell collected the Protoceratops eggshell specimen, and Diego Pol collected the Mussaurus eggshell specimen. The Protoceratops thin sections were prepared by David Varricchio and Anita Moore-Nall. The Mussaurus thin sections were prepared by Jasmina Wiemann. Jasmina Wiemann designed the Raman spectrscopy protocol, developed the biomineralisation proxy, collected, processed and analysed all Raman and histological data. Matteo Fabbri collected all literature data for the ancestral state reconstruction, and ran the analysis.
Timing and spatial scale	Specimens are housed at the AMNH New York and the Yale Peabody Museum. All data were collected in 2019, at Yale University in New Haven, CT.
Data exclusions	In the cluster analysis shown of hard- and soft-shelled eggs, the organic phase was used to characterise if fossil samples were biomineralised or not. With the requirement of organic spectral signatures, hard-shelled Alligator and turtle eggshells had to be excluded, as they did not produce a significant organic signal with the Raman surface assessment (these are hypercrystalline eggshells). When mapping out Raman signatures of protein fossilisation products, amides, and S-heterocycles across polished egg sections, one of the two soft-shelled dinosaur eggs, the Protoceratops sample, had to be excluded from Raman mapping due to the crumbly nature of the phosphatized organic eggshell. Texture effects resulting from a crumbly surface in a polished section affect Raman signatures, especially in mapping procedures. Signal quality is a pre-determined criterion when analysing fossil organic matter.
Reproducibility	All spectra were acquired with 10 accumulations = technical replicates, and replication of the biomineralisation proxy has been successful!
Randomization	Randomization cannot be applied to this study, because Raman spectra are characteristic for every sample, and can be taxonomically assigned at a glance based on their unique and distinctive taphonomic and compositional signatures. Also, we do not use any statistics that would require randomization. Samples were grouped together based on their identity (eggshell versus sediment), and their degree of biomineralisation (biomineralised versus non-biomineralised). The Principal Component Analysis and Discriminant Analysis shown in the Figs. 3, 4 suggests that samples group based on their organic phase as eggshell or sediment, and as biomineralised or non-biomineralised.
Blinding	Spectral data were analysed 'blinded' for the hierarchical cluster analysis, and sample identity was only revealed in the result of the analysis. Otherwise, blinding did not apply for our study, since Raman spectra of different eggshells contain characteristic spectral features, and are readily identifyable even without associated specimen information.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☐	☒ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Palaeontology

Specimen provenance	We analyzed a well-preserved ornithischian egg clutch (IGM 100/1021) attributed to Protoceratops from the Ukhaa Tolgod locality (Campanian), Mongolia. Our sample also includes eggs attributed to the basal sauropodomorph Mussaurus (Fig. 1c) from the Laguna Colorada Formation (Late Triassic/Early Jurassic). These were found in the same locality and stratigraphic level as the type material of Mussaurus patagonicus. Several eggs preserve embryonic and juvenile, including elements of the mandible. These elements bear apomorphic features of Mussaurus patagonicus, including a dentary with anterodorsal peg at the symphyseal anterior end
Specimen deposition	The Protoceratops clutch is housed at the AMNH in New York, and the Mussaurus eggshells are housed at the CONICET, Museo Paleontológico Egidio Feruglio, Trelew, Argentina.
Dating methods	No new dates are provided in our study.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.