Supplementary information of the manuscript

Recalibration of insect evolutionary timescale using Mount San Giorgio fossils suggests survival of key lineages through the End-Permian Extinction

Authors

Matteo Montagna[1,*], K. Jun Tong[2], Giulia Magoga[1], Laura Strada[3], Andrea Tintori[3], Simon Y. W. Ho[2], Nathan Lo[2]

Affiliations

[1]Dipartimento di Scienze Agrarie e Ambientali, Università degli Studi di Milano, Via Celoria 2, I-20133 Milano, Italy.

[2]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Sydney, Australia.

[3]Dipartimento di Scienze della Terra "Ardito Desio", Università degli Studi di Milano, Via Mangiagalli 34, I-20133 Milano, Italy.

**Corresponding author*

Email: matteo.montagna@unimi.it

1. Mount San Giorgio fossil collection

Mount San Giorgio fossils used in this study consist of a subset of the 19 specimens (Table S1) collected during fieldwork activities between 1997 and 2003 in the Lower Kalkschieferzone, the uppermost part of the Meride Limestone, at the Val Mara site D near Meride on the Swiss side of the UNESCO World Heritage site of Mount San Giorgio (Italy-Switzerland; Figure S1). All of the specimens have been deposited at Museo Cantonale di Storia Naturale di Lugano (MCSN), Switzerland. Information relating to the MSG fossils used for calibration are available in Tables S1 and S2.

2. The eight MSG fossils used for calibration

The eight MSG fossils used for calibration are: *i*) the completely phosphatized *Gigamachilis triassicus* (Archaeognatha: Machilidae), ascribed to Archaeognatha based upon the presence of large maxillary palps (partially preserved), abdominal coxopodites showing coxopodal vesicles and styli, paired annulated cerci and the basal part of the filum terminale, with stylus-like structures on the second thoracic leg and scales on appendages that prompt its attribution to the extant group Machilidae [16,17]; *ii*) *Tintorina meridensis* (Ephemeroptera: Tintorinidae), with a general habitus and wing venation that prompt its attribution to the infraorder Ephemeroidea [21]; *iii*) *Archetingis ladinica* (Hemiptera: Tingidae) preserving clearly visible areole on the forewings (protruding beyond abdomen margin) and on pronotum, a moderately long rostrum with bucculae and tarsi two-segmented [18]; *iv*) *Praedodromeus sangiorgiensis* (Coleoptera: Trachipachidae) with a general habitus and diagnostic characters such as the presence of a simple sulcate antenna cleaner in the distal part of the tibiae and the metacoxa separating thorax and abdomen that prompt its ascription to subfamily Eodromeinae of Trachypachidae [22]; *v*) a single elytron showing features of the venation that allowed assignment to the genus *Notocupes* (Coleoptera: Cupedidae) [21]; *vi*) specimen MCSN8464 preserved in dorsolateral view assigned to suborder Polyphaga (not Staphyliniformia) in Coleoptera [22,24]; *vii*) a completely phosphatized specimen identified as a stonefly nymph

(Plecoptera) that exceptionally preserves two pairs of wing stubs (confirming the hemimetabolous metamorphosis of the taxon) and part of the brain, and a prognathous head harbouring smooth and symmetrical mandibles and maxillary palps (specimen MCSN8462) [22,24]; and *viii*) a phosphatized male webspinner (Embioptera) of 18.3 mm in length, which preserves: swollen fore basitarsi, three segmented tarsi, forelegs and hindlegs with enlarged femora, rather large head (as can be inferred from the insertion of the partially preserved antennae and maxillary palps), and impression of folded wings and terga X [22,24].

Nodes at which minimum age constraints were informed by the MSG fossils are the following: *i*) node 161, based on *G. triassicus* and corresponding to crown Archaeognata (Machilidae + Meinertellidae); *ii*) node 166, based on *T. meridensis* and corresponding to crown Palaeoptera; *iii*) node 217, based on *A. ladinica* and corresponding to the split between crown Cimicomorpha and Pentatomorpha, represented by *Acanthosoma haemorrhoidale* and *Notostira elongata*; *iv*) node 249, based on *P. sangiorgiensis* and corresponding to crown Adephaga in Coleoptera; *v*) node 248, based on *Notocupes* sp. and corresponding to the split between the beetle suborders Archostemata and Adephaga, here represented by the extant *Carabus granulatus*, *Gyrinus marinus*, and *Priacma serrata*; *vi*) node 245, based on specimen MCSN8464 (Polyphaga in Coleoptera) and corresponding to crown Polyphaga in Coleoptera; *vii*) node 176, based on the stonefly nymph (Plecoptera) fossil, corresponding to the split between Plecoptera and Orthoptera, Dictyoptera, Mantophasmatodea, Grylloblattodea, Embioptera, and Phasmatodea; and *vii*) node 188, based on the webspinner MCSN8457 and corresponding to the split between Embioptera (*Haploembia palaui* and *Aposthonia japonica*) and Phasmatodea (*Timema cristinae*, *Aretaon asperrimus*, and *Peruphasma schultei*).

3. Image acquisition and figure preparation

Direct observations and measurements were performed using a stereomicroscope (Leica MS5) with an ocular micrometer. The specimens were photographed under two different settings. First, macrophotography was performed with a Canon EOS 450 with 60 mm macro lens, with several image

stacks taken to obtain an entirely sharp high-resolution image. The stacks were subsequently fused and stitched with Zerene and Adobe Photoshop CS. Additionally, multi-layer microphotographs were acquired under transmitted light or fluorescence using a Zeiss Axio Zoom V16 stereomicroscope with the digital camera Zeiss Axiocam 506. Insect drawings included in figures were made using Wacom Intuos Pro graphics tablet and Adobe Illustrator CC. Figures were assembled using Adobe Photoshop CS.

4. Genomic dataset, alignment, selection of meta-partitions and amino-acid substitution model

The amino acid data set was produced by extensive filtering of the original data set of 1,478 orthologue groups (1.3 million sites), which had been assembled from transcriptomic data [12], with alignment-masking techniques using the program Aliscore [80] to remove randomly similar sequences and alignment blocks that were scored as ambiguous [80,81] (see Section 3.3 in Supplementary Text of [12]). In this way, the filtering process maximized phylogenetic signal [80] and ensured that key groups had minimal missing data. The filtering process also removed compositionally heterogeneous sequences, because these violate the assumptions of most phylogenetic models [82].

The amino acid data were divided into meta-partitions using PartitionFinder2 [83]. Each meta-partition contains sequences that PartitionFinder2 judges to share similar evolutionary dynamics and can be analysed as a single unit, according to the corrected Akaike information criterion. Meta-partitions with lengths below 500 amino acids were excluded, leaving a data set consisting of 220,615 amino acid sites. The 'greediest' search algorithm was used in PartitionFinder 2 to split the data set into 85 meta-partitions, with the substitution model for each meta-partition selected using PartitionFinder2 [83]. The LG amino-acid replacement matrix was chosen for 78 meta-partitions [84], the WAG matrix for five meta-partitions [85], and the JTT matrix for two meta-partitions [86].

5. *Description of the five sets of phylogenomic dating analyses performed*

In Analysis 1, we attempted to replicate the dating analysis of Misof et al. [12] by using their 37 fossil calibrations (20 lognormal and 17 uniform priors on node times). In Analysis 2, we used the calibrations from Misof et al. [12] and added the eight MSG calibrations as uniform priors on node times. In Analysis 3, we used the 37 fossil calibrations of Misof et al. [12] but all in the form of uniform priors on node times. This represents a more conservative treatment of fossil evidence because it makes no further assumptions about the relationship between the fossil age and the node time except in providing a minimum age. In Analysis 4, we used the calibrations of Misof et al. [12] and the eight MSG calibrations, all specified in the form of uniform priors on node times. Analyses 1 to 4 all involved the use of an independent-rates relaxed-clock model, whereby the branch rates are independently drawn from a common distribution [87]. In Analysis 5, we matched the calibration scheme of Analysis 4 but instead used an autocorrelated-rates relaxed-clock model.

X. Supplementary references

80. Kück P, Meusemann K, Dambach J, Thormann B, von Reumont BM, Wägele JW, Misof B 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Front. Zool.* **7**, 10.

81. Misof B, Misof K 2009. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: A more objective means of data exclusion. *Syst. Biol.* **58**, 21–34.

82. Jermiin L, Ho SYW, Ababneh F, Robinson J, Larkum AWD 2004 The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Syst. Biol.* **53**, 638–643.

83. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B 2017 PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773.

84. Le SQ, Gascuel O 2008 An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320.

85. Whelan S, Goldman N 2001 A general empirical model of protein evolution derived from multiple protein families using a maximum–likelihood approach. *Mol. Biol. Evol.* **18**, 691–699.

86. Jones DT, Taylor WR, Thornton JM 1992 The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282.

87. Rannala B, Yang Z 2007 Inferring speciation times under an episodic molecular clock. *Syst. Biol.* **56**, 453–466.