

Dear Editor

I reviewed the manuscript entitled “On the morphological variability of *Ichniotherium* tracks and locomotion upgrading in the sistergroup of amniotes” by Michael Buchwitz and Sebastian Voigt. The manuscript represents an interesting quantitative approach to a very well-known ichnotaxon from the Late Carboniferous to Early Permian ichnological record. However, some weaknesses in the methods used may have a greater impact on the interpretations and conclusion of the work. Below some indications and discussions are reported that must be carefully considered before the manuscript can be considered for publication. I'm not a native speaker, but English seems correct and very fluent.

Below, some methodological considerations and other more theoretical ones are reported, which I hope can improve the manuscript and the conclusions obtained and presented.

-In Figure 1C the last occurrence of diadectids is made to correspond to the Kungurian. However, the new diadectid *Alveusdectes fenestralis* has been recently described from the Upper Permian of China (Liu and Bever, 2015). It may be useful to update the scheme using the new cladogram provided by Liu and Bever (2015, fig.2, p. 4), and extending the ‘*Diadectes* and allies’ branch up to the Wuchiapingian.

-Concerning the Principal Component Analysis, from the dataset results, as usual, the presence of several missing entries. It would be useful to specify how the missing values were handled in the analysis. In PAST the default modality to replace missing entries is the ‘mean value imputation’: however it is largely preferable to use the ‘iterative imputation’ algorithm (Hammer 2013). For the iterative imputation the missing values are replaced by their column average in a first stage; a PCA analysis is then run and used to compute regression values for the missing data, and this procedure is simply iterated to reach convergence (see Hammer 2013). If this algorithm was not used, it is strongly recommended to try to redo some analyses to see if significant different results are obtained.

-In the same way, if the raw data were not log-transformed before the analysis, misleading results can be obtained. Transformation into logarithm is strongly recommended to fit linear models and for the correspondence of the log-transform to an isometric null hypothesis (see Chinnery, 2004; Cheng et al., 2009; Romano and Citton, 2015, 2016; Romano, 2017). Again it is recommended then to perform the analysis only after having log-transformed the raw data.

-It would be useful to show the loadings of the single principal components, and describe briefly to what the greatest variance in each component is linked. If, for example, the length of digit IV contributes consistently to the variance (for convention loadings values greater than 0.3 and lower than -0,3) the use of this length to normalize the others results in a both methodological and conceptual error.

It might also be useful to show the scatter plot of components 2 vs 3 (and all other components with significant variance), which very likely are more related to morphometry of the footprints and trackways parameters rather to absolute dimensions. Usually is fine to show only the first two principal components if they describe more than 95% of the detected variance. However, in the submitted work the first two components describe on average only 80% of the variance; thus is there an additional 20% of variance in other components which might be useful to show, and perhaps lead to different discussions and conclusions. In the case of PCA performed on the six trackway parameters, the two first principal components account for only about 63% of the variance. It is therefore necessary to show how the remaining 40% is distributed and if the scatter plots considering the other components provide different result.

An analysis of PCAs using digit lengths in manus and pes of *Ichniotherium* has already been conducted recently, and described in detail in Romano and Citton (2015). It might be worth mentioning briefly this work and comparing the results with the new results.

-One element that can bring to weak results is the choice of the homologue points for digit measurements. In the manuscript, the free digits length is used, a fairly weak character that can change consistently even within the same trackway (pers. obs.), due to factors both directly related to the mode of locomotion and interaction between substrate and autopods and of a purely taphonomic nature (substrate conditions and relative footprint preservational type). It is recommended otherwise to use the phalangeal length of the digit (see, Leonardi, 1987; Romano and Citton, 2015, Fig. 1). In fact, the most useful method to obtain anatomical information about the trackmaker is the measure of the digit phalangeal portion length, concerning ‘...*the measure of the segment that joins the distal extremity of the digit with the corresponding mid-point of the metapodial-phalangeal pad*’ (Leonardi, 1987, p. 47), which is more closely related to the real anatomical length of the digit. If the metatarsal-phalangeal pad is not visible or preserved, it is advisable not to consider such footprints in the analysis. In fact, is better to have fewer footprints but definitely linked to a real osteological pattern and structure, rather than many tracks where digit lengths are extremely variant, and not depending on the trackmaker's anatomy. It is likely that, by

decreasing the number of tracks to the best and anatomically significant footprints, the distance between the three recognized morphotypes in *I. cottaie* reduces or totally disappears.

-Another possible problem is that the PCAs were performed not on linear measures, but on ratios between measures (in particular the various lengths normalized for digit IV). First of all, it is strongly advised to not use ratios at all in principal component analyzes (see Hammer and Harper, 2006); ratios are best indicated, differently, for classical cluster analyzes. By dividing all the measures by the length of digit IV (or by other lengths), the necessary independence of the different variables is lost (a fundamental requirement for classical PCA), with the possible obtainment of misleading trends or clusters (especially if the digit chosen as reference is variant or peculiar). In fact, normalizing the same measure can lead to more uniform and separate groups, but nevertheless artificial. The same methodological problem applies to track parameters included as a ratio, and normalized for the length of digit IV (or other lengths). According to Hammer and Harper (2006) “*Compositional data (relative proportions) must be analyzed with a special form of PCA (Aitchison 1986, Reyment & Savazzi 1999), because we need to correct for spurious correlations between the variables*”. The biggest problem is indeed the “spurious correlations”, with detection of artificial clusters due to normalization, and to the non-independence of variables.

-In the text is highlighted “*a surprising difference between the Bromacker sample and older trackways from the Birkheide and Gottlob localities*”. However, two of the three specimens from Gottlob in Figure 10 C fall completely into the morphospace of Bromacker. Moreover, an overlap is also present with the Birkheide specimens in Figure 10 D and again with the ones from Gottlob in Figure 10 E.

-Since in the work a sort of evolution in posture and in trackways parameters is described, with *Ichniotherium praesidentis* as outgroup, followed by *I. sphaerodactylum*, I am surprised that these taxa have not been included in the PCA analyzes and in the other graphs of figures 10 and 12. Such an analysis (by using linear measures, as already indicated, and not ratios) it could show how the three recognized morphotypes for *I. cottaie* result closer to each other than to *I. praesidentis* and *I. sphaerodactylum*. If that is not the case, it means that the length of the fifth digit is not a solid and sufficient character for the division of ichnotaxa. Since the authors speak of phylogeny, it is good to remember how the use a single character is typical of the nineteenth-century pre-cladistic typological philosophy. Starting with Hennig, if one want to talk about phylogeny, is necessary to

consider the congruence tests of a complex set of characters (for example under the principle of parsimony or other principles), and not of individual characters (see Romano and Nicosia, 2015).

-An important element to be discussed is the ichnotaxonomic consequences of the study. According to the authors, within the specimens referred to *Ichnoterium cotta* (on the basis of a relatively short digit V) three distinct morphotypes are recognizable, in particular the 'Birkheide-Gottlob type', the 'Bromacker type' and Hainesi-Willsi type'. Again according to the authors, each of the morphotypes shows a functionality and locomotion substantially different, therefore attributable to three different trackmakers (or evolutionary degrees in the conclusions). A first crucial point on the philosophical level is the problem to include three different recognized biological trackmakers under the same ichnospecies; especially if sufficient characters are available to even distinguish evolutionary degrees in locomotion and functionality from trackways parameters. The question could be as follows: if such a high number of track parameters to distinguish three morphotypes and different functionalities are available, why not set up three separate ichnotaxa? From what has been reported and discussed in the paper, seems that the fundamental feature (the “synapomorphy”) characterizing *Ichnoterium cotta* is simply a relatively short digit V. However, the length of a single digit, even at the level of genetic plasticity, is an extremely variant element in evolution, with loss and addition of phalanges in repeated and paraphyletic ways. It seems that this character on which all the *I. cotta* “box” is founded, results weaker than a large number of trackways parameters, best referable to different trackmakers. If the ichnotaxonomy, in a new illuminated vision, must be based on natural biological taxa (i.e. zoological trackmaker and their unique combination of synapomorphic, plesiomorphic and autapomorphic characters; see Olsen, 1995; Carrano and Wilson, 2001; Romano et al., 2016), it is difficult to imagine that different degrees of evolution in locomotion and functionality (referable to different trackmakers as found in the work) are hierarchically subordinate to the number of phalanges on a single digit, within the framework of the hierarchic enclitic system founded by Willi Hennig.

Secondly, one may wonder if having such conservative autopod structures in all morphotypes, we are dealing just with a same kind of biological trackmakers in different ontogenetic conditions, and under different types of locomotion. Simply changing the speed, trackways parameters (as experimentally observed) change accordingly, with trackway length that narrows, stride that increases, axes of autopods more parallel to the direction of advancement and so on. If the various track parameters change together (strong covariance) from a morphotype to another (what seems to happen), it would be more parsimonious to imagine a change in the locomotion and speed of the same trackmaker type. After having seen the tremendous difference between a normal sprawling

crocodile and the same individual under a galloping locomotion, I think that the simple trackway parameters must be taken with extreme caution in obtaining greater conclusions in the macro-evolutionary field.

Another element that could be taken into consideration is the differential depth of the different portions of the footprints, as an evidence to infer the trackmaker biomechanics. In fact, as shown in some analyses performed also on *Ichnotherium* material (see Romano et al. 2016), such a study can shed light on the various functional axes active from time to time during stroke progression, and highlight synapomorphic characters of a particular vertebrate clade, mirrored in the preserved tracks. So, it would be interesting to know whether in the three recognized morphotypes within *Ichnotherium cottaie*, the differential depth of various footprint portions is uniform during the three main phases ('touch-down', 'weight-bearing' and 'kick-off' phases), or if substantial changes in the distribution of pressures are observed (and therefore in biomechanics and functionality that can be inferred).

The authors are free to contact me for any points raised on this review

kind regards

Marco Romano

Berlin

11/10/2017

Literature mentioned

Aitchison, J. 1986. *The Statistical Analysis of Compositional Data*. Chapman & Hall, New York.

Carrano, M.T., Wilson, J.A., 2001. Taxon distributions and the tetrapod track record. *Paleobiology*. 27, 564–582.

Cheng, Y.N., Holmes, R., Wu, X.C., Alfonso, N., 2009. Sexual dimorphism and life history of *Keichosaurus hui* (Reptilia: Sauropterygia). *J. Vertebrate Paleontology* 29, 401e408.

Chinnery, B., 2004. Morphometric analysis of evolutionary trends in the ceratopsian postcranial skeleton. *J. Vertebrate Paleontology* 24, 591e609.

Hammer Ø. 2013. *PAST Paleontological Statistics Version 3.0: Reference Manual*. University of Oslo.

Hammer, Ø. & Harper, D.A. T. 2006. *Paleontological Data Analysis*. Oxford: Blackwell Publishing Ltd, 351 pp.

- Liu, J., & Bever, G. S. 2015. The last diadectomorph sheds light on Late Palaeozoic tetrapod biogeography. *Biology letters*, 11(5), 20150100.
- Olsen, P.E. 1995. A new approach for recognizing track makers. *Geol. Soc. Am. Abs. Prog.*, 27, 86.
- Reyment, R.A. & E. Savazzi. 1999. *Aspects of Multivariate Statistical Analysis in Geology*. Elsevier, Amsterdam.
- Romano, M., 2017. Long bone scaling of caseid synapsids: a combined morphometric and cladistic approach. *Lethaia*. <http://dx.doi.org/10.1111/let.12207>.
- Romano, M., & Citton, P. 2015. Reliability of digit length impression as a character of tetrapod ichnotaxobase: considerations from the Carboniferous–Permian ichnogenus *Ichniotherium*. *Geological Journal*, 50(6), 827-838.
- Romano, M., & Nicosia, U. 2015. Cladistic analysis of Caseidae (Caseasauria, Synapsida): using the gap-weighting method to include taxa based on incomplete specimens. *Palaeontology*, 58(6), 1109-1130.
- Romano, M., Citton, P., & Nicosia, U. 2016. Corroborating trackmaker identification through footprint functional analysis: the case study of *Ichniotherium* and *Dimetropus*. *Lethaia*, 49(1), 102-116.