A critique of vole clocks

Robert A. Martin

Department of Biological Sciences, Murray State University, Murray, KY 42071, USA

ARTICLE INFO

Article history:
Received 5 February 2014
Received in revised form
7 March 2014
Accepted 9 March 2014
Available online 22 April 2014

Keywords:
Vole clock
Size change
Pleistocene
Spain
Arvicolid

ABSTRACT

Recent attempts to estimate the age of deposition of European fossil localities using mathematical equations derived from size change of the first lower molar in arvicolid rodent lineages as a function of time prompted an assessment of the value of this approach. The accuracy of “vole clocks” depends on accurate dating of a fossil system and establishment of a directional size change pattern through time in a given species from the dated system. Results of this review suggest that vole clocks have limited value for biochronology. In addition to several methodological and statistical problems with published studies, vole clocks in general are untenable because paleontological systems cannot resolve dating to the level of accuracy necessary to construct an accurate equation, size and shape change is never monotonic (constant velocity) in lineages, and size commonly reverses direction in lineages on all time scales.

1. Introduction

The recent paper by Martínez et al. (2013) offers an opportunity to begin a discussion on the value and accuracy of what are colloquially known in the literature as “vole clocks,” the use of size or character change, expressed as linear measurements or ratios in various lineages of arvicolid rodents, to estimate age of deposition of a given fossil locality. Arvicolid rodents are ubiquitous in fossil deposits throughout the Holarctic region from about 5 million years to the present time, and their explosive speciation, accompanied by considerable dental morphological change, makes them ideal subjects for studies of evolutionary patterns, paleoecological reconstruction and biochronology. The famous European MN zones (Mein, 1989) and, in North America, RZ intervals (Martin, 2003), which rely to a great extent on arvicolids, are helpful in ordering depositional sequences and, to a certain extent, in temporal placement of localities. The holy grail of arvicolid studies since I was a graduate student in the 1960s was the derivation of a mathematical equation from the relationship of dental size or morphological change in arvicolids and time that could accurately determine the age of deposition of a given fossil locality; in essence, to derive a function based on structural change with the accuracy of radiocarbon decay. Martínez et al. (2013) claimed to be able to correlate level 10 (EV17) from the Torrent de Vallparadís depositional sequence near Barcelona, Spain, with level TD5 from the Gran Dolina sequence from the Atapuerca hills, Burgos, Spain using an equation calculated from m1 size change in the extinct arvicolid Mimomys savini. They suggested that level 10 at the Vallparadís sequence (there are other outcrops with superposed faunal levels in the Torrent de Vallparadís, such as Cal Guardiola) was deposited between 0.98 and 0.95 Ma. Although an error range was not provided, this estimate implies a level of accuracy normally afforded Milankovich precessional cycling (~0.03 my) and which is nearly impossible to attain with current radiometric techniques other than radiocarbon dating. In this study I will examine the Vallparadís calculations and use a number of examples from the mammalian fossil record to demonstrate that such derived equations have limited chronological value.

2. Chronological issues

Prior to their M. savini calculations, Martínez et al. (2013) summarized the chronology of the Torrent de Vallparadís (entire sequence), Gran Dolina (level TD6) and Fuente Nueva 3 (FN3). The Vallparadís sequence generated the following dates: EVT12-8, 1.07–0.99 Ma; EVT7-EVT3, 0.99–0.78 Ma; Units above EVT3, <0.6 Ma. The date for level TD6 at Gran Dolina was given as 0.88–0.80 Ma, and that for Fuente Nueva 3 was 1.19 ± 0.21 Ma. All dates except those for EVT12-8 were derived from ESR-U series dating of mammal teeth and quartz grains or OSL dating of quartz grains. EVT12-8 dating was determined by identification of the Jaramillo subchron in the lower part of the Vallparadís sequence. Martínez et al. (2013) also cited a date of 1.22 ± 0.16 Ma for the Lower Red
Unit at Sima del Elefante (also known as Trincheria Elefante, Lower Red Unit; TELRU) in the Atapuerca sequence of north-central Spain. Citing a study by Martínez-Navarro et al. (1997), the authors also assumed that the Huéscar 1 site of the Baza Basin was dated to the Jaramillo (Martínez et al., 2013, p.141). Based on these dates and biocultural considerations derived from Spanish Pleistocene rodent replacement patterns (Cuenca-Bescós et al., 2010), Martínez et al. (2013) applied an equation calculated from size change in the m1 in M. savini fossil samples from level 10 (EV7) at Vallparadís, level TD6 from Gran Dolina, and Fuente Nueva 3 and Barranco León D in the Baza Basin of southern Spain.

In an evaluation of a reported pre-Jaramillo date for Barranco León 5 (= Barranco León D [BLD] (Toro-Moyano et al., 2013), Muttoni et al. (2013) showed that the original authors’ own error range on their electron spin resonance data precluded an assumption of a pre-Jaramillo age for the locality. Muttoni et al. (2013) also demonstrated that the error range associated with the cosmogenic nuclide date of 1.22 Ma from level 9 (TELRU) at Sima del Elefante was so large that at 2σ (two standard deviations) it encompassed a range of 0.90–1.54, and that the cosmogenic nuclide date from the lower level 7, at 1.13 Ma, generated an error range of 0.77–1.49 Ma at 2σ. A recent summary of ESR-U dates for Gran Dolina (Parés et al., 2013) provided a range of <0.5–>1.0 Ma at 2σ for three dates from level TD6, averaging about 736 Ka. An average of two TL dates from the next higher stratum, TD7, is 960 ± 120 Ka. Based on new paleomagnetic (pmag) data, Parés et al. (2013) chose to reject the ESR dates and accept results from the combined TL and pmag work, giving their final estimate for the top of TD6 (their “Aurora Stratum”) at 0.956 Ma. Their arguments for this conclusion were torturous and, while doing the best they could with the data, do not resolve the dating of Gran Dolina to a level of accuracy implied by the Martínez et al. (2013) M. savini m1 size trend model.

Use of the 1.19 Ma date for Fuente Nueva 3 is also problematic. In the Duval et al. (2012) paper where this date was published, the authors were forced to reject 7 of 9 teeth analyzed by U-Th/ESR methods (including all from Barranco León D), and then chose one tooth from FN3 that produced the 1.19 Ma date, proposing that this date was valid because it was substantiated by mammal biochronology. This amounts to circular reasoning and does not allow confidence in the 1.19 Ma date. A recent paper evaluating the geology, paleomagnetism and biostratigraphy of the Orce region referred Huéscar 1 to a post-Jaramillo age (Gilbert et al., 2007). The Jaramillo has not been identified in the Orce area and Iberomys huescaensis has not been recovered from sediments in the Baza basin during or prior to the Jaramillo. On a more trivial taxonomic note, “Allophaiomys” chalinei was recently allocated to the new genus Victoriamys (Martin, 2012).

The Torrent de Vallparadís section is the only superposed set of Pleistocene sediments in Spain within which the Jaramillo has been identified. The geology is complicated, the exposures are limited and physically separated, and there is a considerable normal overprint to the paleomagnetic data (Madurell-Malapeira et al., 2010). Following Cuenca-Bescós et al. (2010), Martínez et al. (2013) partly substantiated their interpretation of the relatively ancient age for the lower levels at Torrent de Vallparadís based on the combined presence of M. savini and Microtus (Allophaiomys) lavocati in the lower levels of the Cal Guardiola section, referring in particular to the cosmogetic nuclide dates from TELRU. As noted above, the dates for FN3, BLD and TELRU cannot be considered determined and are in any case associated with considerable error. This error envelope alone precludes the use of the Martínez et al. (2013) calculations from generating accurate age estimates. Further, confirmation of a Jaramillo age for the lower levels of Cal Guardiola and Vallparadís based on the presence of M. (Allophaiomys) lavocati is premature, and the combined presence of M. savini and M. (Allophaiomys) lavocati in Spain prior to the Jaramillo remains to be substantiated. These chronological considerations are not trivial, but the more contentious issues revolve around statistical methodology.

3. Sampling and statistical issues

In addition to the Martínez et al. (2013) study, Lozano-Fernández et al. (2013a,b) used data from Spanish M. savini to develop equations for chronologival purposes. Some of the m1 data, particularly from Gran Dolina, overlap in the two studies, and a comparison is provided in Table 1. The only significant departures are in the breakdown of the sixth excavation level at Trinchería Dolina (TD6). One mean is presented for TD6 in the Martínez et al. (2013) paper, whereas three means are provided for three subzones of TD6 by Lozano-Fernández et al. (2013a,b). Within this finer partitioning of the M. savini record, we can see a size reversal at TD6-3. The reversal was discussed by Lozano-Fernández et al. (2013b), and they chose to remove it before generating their final equation. The logic behind this decision is important, and will be discussed further below. Although m1s of M. savini were available from other levels in the Torrent de Vallparadís system, as they are illustrated from, for example, Cal Guardiola (Madurell-Malapeira et al., 2010), Martínez et al. (2013) presented m1 data for only level 10 (EV7) in the Vallparadís section. This may have been due to low numbers of rodents from other layers, but based on the authors’ claims regarding a directional size trend, a more complete set of M. savini data from the Torrent de Vallparadís sequence would have been both helpful and appropriate.

The primary statistical issue with both the Martínez et al. (2013) and Lozano-Fernández et al. (2013a,b) studies is that in neither case did the authors conduct the basic statistical procedures necessary to establish that a size trend exists in M. savini m1s. The appropriate test is a one-way ANOVA followed by a post-hoc test, such as Tukey’s HSD. The Tukey test is essential, because a significant F-value for the ANOVA would reveal only that at least one sample was statistically different from one other, and not in any particular direction. Based on published values for the m1 means and standard deviations, a good case can be made without the ANOVA that the gradual trend suggested by the authors is absent. This can be seen in Figs. 4 and 5 of the Martínez et al. (2013) paper and Fig. 3 of Lozano-Fernández et al. (2013b), in which the range of m1 measurements of almost any sample of M. savini m1s chosen for Trinchera Dolina overlaps all the other samples at 2s. At 2s there is almost complete overlap between the m1 lengths from the supposed oldest samples BLD and FN3 and the youngest sample, TD6. This is particularly clear when we examine the partitioned TD6 reported by Lozano-Fernández et al. (2013b), where the mean m1 length from level TD6-3 demonstrates a reversal and an observed range showing complete overlap with all other samples at 2s. Without a confirmed size trend, age estimates generated by the equations in both Martínez et al. (2013) and Lozano-Fernández et al. (2013a,b) have little meaning.

4. Size and shape change in the history of arvicolid

Lozano-Fernández et al. (2013b) argued that removing the TD6-3 sample from their calculations was appropriate because the inferred small body size of this sample represented an anomaly that disturbed an otherwise irreversible trend. Monotonic (constant velocity) size change almost never occurs in the history of a lineage or a clade (Gould and MacFadden, 2004), and the reason is fairly straight-forward; body size (or mass) is highly sensitive to both abiotic and biotic influences such as shifts in climate or quality of...
resources, and responds accordingly (Holbrook, 1977; Smith et al., 1995). Environmental flux may also bring in immigrants that intergrade with the native population, thus rapidly shifting the mean size either up or down. Certainly, lineages and clades increase in average size over various time scales, but close scrutiny, as for example in the history of horses (MacFadden, 1994; Gould and MacFadden, 2004), always reveals a complex history full of twists, turns and temporary reversals. Many studies of arvicolid rodents testify that size can and often does reverse direction, sometimes for millions of years.

In a set of related studies Barnosky (1990, 1993) showed that both size and shape change of Microtus pennsylvanicus M3s differed in the eastern U.S. during the late Pleistocene depending on geographic location of fossil localities. While sampling error cannot be ruled out, the data are best explained by mosaic evolution; different local responses to climate and perhaps biotic influences. These results should serve as a warning that contemporaneous size or shape response of *M. savini* could be different in north-central Spain as opposed to southern Spain, France, or Germany. Gordon (1999) examined size and morphological change in the M3 and m1 of the meadow vole, *M. pennsylvanicus*, from a 7000-year sequence in Yarborough Cave, Georgia, and compared her results with those from other cave systems in the eastern United States. The m1s displayed a significant decrease in length through the sequence, but reversed in size from the highest Pleistocene level (ca 16,000 yr B.P.) to modern time (Fig. 1). M3s showed no significant size change through the entire sequence. Gordon (1999, p. 226) also observed that m1 size change in Georgia mirrored that from Alabama, but was opposite the direction of change in Virginia and Tennessee during the same time period. Martin et al. (2011) also showed that the shift on the Central Great Plains from five- to six-triangle m1s in *M. pennsylvanicus*, which began at a very modest pace during the mid-Pleistocene, accelerated tremendously in the latest Pleistocene, with most of the change taking place in a remarkably brief 1500-year period around 12,500 years ago. Without the latest Pleistocene data, a linear regression model (assuming monotonic change) constructed for the purpose of dating fossil localities with meadow voles would have badly estimated the true age of these sites.

In a study of the extinct archaic vole *Ogmodontomys* in the Meade basin of Kansas, Marcolini and Martin (2008) showed that body size, as estimated by m1 length, both increased and decreased (with at least one additional reversal) for millions of years, demonstrating an overall pattern of stasis (Fig. 2). Because the pattern appears linear from, for example, the period 4.8–3.8 Ma, one could theoretically construct an equation based on these data and then estimate the age of stratigraphically isolated localities during that period if they contained *Ogmodontomys*. There are three problems with this approach: 1) the localities between 4.8 and 3.8 Ma lack radiometric dates and are therefore only roughly placed in time, 2) there may have been numerous reversals of size unrecorded by the available samples and 3) variation (e.g., variance) in the samples precludes their use to accurately estimate age.

Another instructive example is the history of the North American muskrat, *Ondatra zibethicus*. One of the best examples of phylectic (intraspecific) dental change in the fossil record, a number of studies have shown that *O. zibethicus* M1s changed in both size and shape over the past 3.75 million years (Semken, 1966; Martin, 1979, 1996). The most recent analyses of m1 size change were by Mihlbachler et al. (2002) and Martin et al. (2009). Mihlbachler et al. (2002) examined size change over the past 32,000 years in the Aucilla River late Pleistocene assemblages from north Florida. They confirmed an earlier analysis by Martin (1996) that North American muskrats experienced a rapid dwarﬁng event after about 12,000 yr B.P. leading to modern time. Conveniently, Mihlbachler et al. (2002) also provided samples of modern muskrats from the eastern United States, showing that the variation of m1 means (ranging from a low of 7.39 mm in Alabama to a high of 8.02 mm in

---

**Table 1**

Measurements of *Mimomys savini* m1s from Vallparadís (Val), Barranco León D (BL), Fuente Nueva 3 (FN) and Gran Dolina (TD). L = Martínez et al. (2013). O.R. = observed range, s = standard deviation, N = number of specimens.

<table>
<thead>
<tr>
<th>Source</th>
<th>L</th>
<th>M</th>
<th>L</th>
<th>M</th>
<th>L</th>
<th>M</th>
<th>L</th>
<th>M</th>
<th>L</th>
<th>M</th>
<th>L</th>
<th>M</th>
<th>L</th>
<th>M</th>
<th>L</th>
<th>M</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locality</td>
<td>Val</td>
<td>BL</td>
<td>BN</td>
<td>FN</td>
<td>BL</td>
<td>FN</td>
<td>TD4B</td>
<td>TD4/3</td>
<td>TD5B</td>
<td>TD5B</td>
<td>TD5A</td>
<td>TD5A</td>
<td>TD6-3</td>
<td>TD6-2</td>
<td>TD6-1</td>
<td>TD6</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.48</td>
<td>3.48</td>
<td>3.23</td>
<td>3.25</td>
<td>3.28</td>
<td>3.43</td>
<td>3.38</td>
<td>3.46</td>
<td>3.46</td>
<td>3.48</td>
<td>3.49</td>
<td>3.43</td>
<td>3.58</td>
<td>3.58</td>
<td>3.58</td>
<td>3.58</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>0.21</td>
<td>0.21</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>22</td>
<td>23</td>
<td>132</td>
<td>53</td>
<td>45</td>
<td>12</td>
<td>5</td>
<td>3</td>
<td>17</td>
<td>50</td>
<td>102</td>
<td>3</td>
<td>9</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Fig. 1.** Change in length of M3 and m1 in the North American meadow vole, *Microtus pennsylvanicus*, from an excavation in the Peccary Room of Yarborough Cave, northern Georgia. Vertical numbers represent sampled levels. Box = 25–75 percentiles; straight line through box = 10–90 percentiles, black dot = mean. Adapted from Gordon (1999).
Pennsylvania) included a number of samples in which the observed ranges did not overlap, though all samples overlapped at 2s. Using the muskrat data for comparison, all of the *M. savini* samples reported by Martínez et al. (2013) and Lozano-Fernández et al. (2013a,b) could have come from a single set of contemporaneous populations.

Martin et al. (2009) focused on the best-dated North American localities with fossil muskrats in order to test previously published hypotheses of size change (Martin, 1979, 1996). Their results supported previous studies that had fit muskrat m1 (and thus body size) change to a complex polynomial (Fig. 3). Although the $r^2$ value for the regression was 0.92, it is clear from the graph and variation around it (of just the means, let alone the variance) that it would be impossible to use the equation to generate anything but the roughest estimates for age of an unknown site based on m1 length. This is true even for the last ~0.65 million years, where the response is basically linear (except for the late Pleistocene reversal). Note the oscillation of mean m1 length around the line, demonstrating that size reverses direction regularly.

Maul et al. (1998) specifically reviewed the use of a variety of ratios of arvicolid dental dimensions for the purpose of geochronology. They were careful to list some of the potential limiting factors in this endeavor, but despite these limitations found their results somewhat encouraging. In particular, they used the A/L ratio of van der Meulen (1973), the enamel differentiation (SDQ) ratio of Heinrich (1978), and the HSD ratio, the latter a modified version of Rabeder’s (1981) HH index. The A/L ratio measures the length of the m1 anteroconid complex as a function of m1 length in *Microtus*. The SDQ index is a composite of ratios of the thickness of enamel on the anterior and posterior edges on five triangles and the posterior loop of a *M. savini* or *Arvicola* m1. The HSD index is a ratio of the height of the hyposinuid (HSD), one of a number of enamel-free areas

**Fig. 2.** Summary chart for changes in select measurements and qualitative characters in the extinct vole *Ogmodontomys* from Meade County, Kansas through more than two million years. MPTS – magnetic polarity time scale, L.f.s – local faunas, Schmelzmuster – microscopic enamel prism banding pattern, T – triangle, R – radial enamel, IT – incipient tangential enamel, T – tangential enamel, DL – discrete lamellar enamel, g – grams (estimated from m1 length), linea sinuosa – wavy enamel contact between crown and root, acd – anteroconid. Disappearance of lamellar enamel in the schmelzmuster represents replacement of a population in the Meade basin (*Ogmodontomys sawrockensis*) with lamellar enamel by another without it (*O. poaphagus*). Note that tangential enamel continues to increase in quality through the sequence. From Marcolini and Martin (2008).

**Fig. 3.** Graph of m1 length as a function of time in the North American muskrat, *Ondatra zibethicus*. Equations and line generated by Microsoft Excel. Ma – millions of years ago. From Martin et al. (2009).
(sometimes also referred to as dentine tracts) on the labial side of an arvicolid m1, to m1 length. Although ratios are dimensionless quantities, if composed of variables correlated with body size (e.g., tooth length) they are still subject to size-related selection and consequently can reverse direction even if the numerator value in the ratio (such as HSD height) does not. The SDQ index is a different matter, and could be size-independent, though that property needs to be tested with samples taken from geographically separated, accurately dated, quarries. But herein lies a problem. The vast majority of European arvicolid records come from isolated localities within which there is, at best, only a fraction of the palaeomagnetic record, and if beyond the limit of radiocarbon dating and without volcanic rocks they cannot be accurately dated. The chronostatigraphic placement of these localities is often based on the afore-mentioned dental ratios in M. savini or Arvicola, which once again amounts to circular reasoning. Furthermore, as noted earlier, detailed studies of dental patterns in late Pleistocene and extant arvicolids reveal a mosaic pattern, which is virtually impossible to distinguish with fossil materials because, if beyond radiocarbon dating, no two samples can be determined with certainty to have been contemporaneous. Indeed, in a recent paper on patterns of dental evolution in European arvicolids, Maul et al. (2013) recognized the complex, mosaic expression of dental changes through the Pleistocene. Even the SDQ ratio, despite its potential independence from size, displays a mosaic pattern in extant Arvicola, including variation with altitude (Kratochvíl, 1980, 1981, 1983), and Rudd (2011) recently suggested the SDQ ratio in Arvicola may express ecodephytophenotypic variation as a result of climatic shifts affecting food quality. Evaluating Pleistocene Arvicola m1 morphology with outline analysis and SDQ calculations, Escude et al. (2008) rejected the use of Arvicola dental data for biochronology.

5. Conclusions and recommendations
Sampling, chronologic, and statistical issues seriously limit the accuracy and thus practical application of vole clocks generated from fossil arvicolid rodent samples. There is no evidence to support the hypothesis that tooth size directionally “decays” with the predictability of a radioisotope; indeed, the available evidence indicates that tooth size commonly reverses direction on all temporal scales. The expression of mosaic patterns of variation in extant species (including qualitative features such as enamel differentiation) suggests that comparison of fossil samples should be limited to restricted geographic regions unless it can be unequivocally demonstrated that all contemporaneous populations of a given widespread species evolved in the same manner over geologic time. This is usually impossible to determine because the number of fossil samples is limited and contemporaneity is difficult to establish if samples are beyond the limits of radiocarbon dating. The problem is exacerbated when dealing with isolated deposits, such as cervices or fissure-fills, and issues of dating and stratigraphic positional uncertainty may be intractable for these kinds of sedimentary circumstances. More promising are the extensive sedimentary basins of southern Spain and other regions, where stratigraphic relationships can be determined and palaeomagnetic profiles can be established. Nevertheless, to end on a positive note, if we reflect on Fig. 2 for Ogmomedontomys poaphagus from the Meade Basin of Kansas (a set of stratigraphically superposed samples recovered from an area of about 50 km²), it might be possible for an enterprising statistician to develop a multivariate index simultaneously composed of data from tooth size and schmelzmonster, linea sinuosa, and occlusal patterns to delineate a unique morphospace for each arvicolid sample. Perhaps, if enough independently evolving parameters are included, a pattern might emerge among samples that has some value at least for chronologic sequencing.

Acknowledgments
I thank L. Gibson and C. Scott for their encouragement of my arvicolid studies in Spain, and G. Cuenca-Bescós, C. Falgueres, G. Muttoni and I. Lozano-Fernández for providing references and information. Evaluation of the manuscript by C. Fernández-Cahadell, L. Gibson and two anonymous reviewers was very helpful. Ogmomedontomys research in the Meade Basin of Kansas is supported by the National Geographic Society (5963-07, 6547-99) and the National Science Foundation (EAR 0207582).

References


van der Meulen, A.J., 1973. Middle Pleistocene smaller mammals from the Monte Peglia (Orvieto, Italy) with special reference to the phylogeny of *Microtus*. Quaternaria 17, 1–144.


