**Supplementary Information 4: Summary of results for each factor.**

**This document contains a summary of the results found for each factor and an extended discussion of the implications of these.**

*Species*

Differences between *species* are discussed in the main text.

*Specimen*

The use of *specimen* was largely to encapsulate inter-individual variation. The within-species datasets demonstrated this as a significant effect in 23 of 32 comparisons across the four DMTA variables. However this effect was not significant when these eight species were combined into a single dataset. This is likely due to the sheer number of comparisons involved with the 353 specimens being compared. Any inter-individual variation for such a large dataset would be swamped by greater variation between species. This also demonstrates how comparisons of interspecific factors across multi-species datasets can be misleading.

Inter-individual variation in *specimen* was thought to perhaps encapsulate other factors such as *wear* and *ecoregion*. However both those factors were utilised in final models developed by the GLMM process. This suggests that inter-individual variation is due to other factors, or combinations of factors, including some not considered here, such as year and season that the specimen was collected particularly given the well-known ‘last supper phenomenon’ regarding dental microwear formation (Calandra and Merceron 2016). This finding also supports the use of *specimen* as a random factor in the GLMM models, as it likely represents a number of factors that in concert do not fit an easily modelled distribution (Zuur *et al.* 2009). This approach may also provide some of the utility behind *specimen,* as it allows for natural inter-individual variation as a random factor, while controlled inter-individual variation is predictively modelled (e.g. for *ecoregion*).

*Microscope*

Subsampling indicated that separating data by microscope enabled taxa to be distinguished with lower sample sizes for *Asfc* but not *epLsar,* and showed little to no improvement in distinguishing taxa for *Sdr* or *Vvv.* The two taxa which could be compared for *microscope* showed significant differences between microscopes in six of eight comparisons, and all four across the entire dataset. Microscope featured in the final GLMM models for *Asfc, Sdr* and *Vvv.* The effects of this cannot be visualised as the factor has only two levels and modelling requires one of these to be the dummy variable (Zuur *et al.* 2009).

Microscope was included in this analysis as it is known to differ between instruments, however it was hoped that the filter of Arman *et al.* (2016) would minimise any systematic differences. While this appears to be largely the case for *epLsar,* it seems that differences are still apparent in the three other SSFA variables considered here*.* Unless further improvements can be made to the Arman *et al.* (2016) method, or an alternate fix is found, these differences can only be managed by incorporating such factors into analysis. Doing so however should not present any serious difficulty, as such systematic differences are likely to be largely uniform and so easier to model than other factors.

*Element*

Subsampling was unable to determine if separating samples by element assisted in differentiating between species, as species sampled across upper and lower molars showed good separation of taxa prior to controlling for element. Of the two species where element was considered in isolation neither showed any difference for any variable. There were differences seen across the entire dataset, but confounding effects are likely. Element featured only in the final GLMM model for complexity (*Asfc*).

That element was seen to differ at all was somewhat surprising as the biomechanics of occlusion should suggest similar forces being found on the opposing wear facets, however experimental studies have shown that human teeth with slight irregularities, and even some without, show marked asymmetry in forces between upper and lower teeth (Osborn 1961). Indeed, a number of other factors; that the dentary is mobile while the maxilla static, that the movement of foods are effected by other factors (e.g. the tongue), and that upper and lower molars possess quite different morphologies, may all be aspects which could lead to differences in microwear based on *element*. Indeed, differences have been noted previously, with Schulz *et al.* (2010) analysing upper and lower teeth separately due to differences found. It then may be likely that differences between upper and lower teeth reflect complexities in the occlusal cycle, particularly considering *element* featured alongside *tooth* and *facet* in the final complexity GLMM model.

*Tooth*

Subsampling revealed little or no improvement in differentiating between *M. robustus* and *W.bicolor* for any variable when controlling for tooth. In the within-species analyses, *tooth* differed only in five of the 32 comparisons undertaken. No significance for tooth was seen when considered across the whole dataset. Tooth features in the final GLMM models for *Asfc* and *Sdr*, and the modelled effect of *tooth* can be seen in figure S4.1. This shows little difference between molars 1-3, but that *Asfc* and *Sdr* are slightly elevated for M4. Results for MX and MY are of little relevance as these were a small number of specimens for which tooth position was unknown.



**Figure S4.1:** Effects of tooth on interspecific models of DMTA data. Mean and 95% confidence intervals of ecoregion effects of models produced by GLMM (see main text table 5). Area-scale fractal complexity (*Asfc*) left; developed interfacial area ratio (*Sdr*) right. All variables normalised prior to analysis (see SI1), ‘dP3’ tooth used as dummy variable, MX and MY are a small number of specimens for which tooth position was unknown.

*Tooth* was considered *a priori* to be likely to affect microwear as the position of each *tooth* in the dental arcade should alter the size of particles being chewed in different parts of the mouth (Lucas 2004), or that the forces of occlusion may vary with increasing distance from the fulcrum of the temporomandibular joint (Bakke 2006). Inter-tooth differences in other herbivores have been rarely investigated in DMTA, though have been noted in canids along a number of parameters (Ungar *et al.* 2010) as well as *Equus grevyi* (Grevy’s Zebra) in Textural fill volume, an SSFA variable, and *Sal,* an STA variable, but not in *Asfc,* *epLsar*, or *Vvv* which were tested for (Schulz *et al.* 2010).

That there was little effect of tooth position may be in part due to molar progression. Molar progression is the anterior movement of teeth throughout an animal’s life and is best exemplified in macropodids (Lentle *et al.* 1999). This means that later in life posterior molars become effectively anterior molars, which may equalise any inter-tooth effects. With most differences focused between M4 and other teeth may be as this posterior-most molar rarely reaches an anterior position. One possibility to consider for future studies would be only considering microwear on molars 1-3.

*Facet (including reclassifications)*

Subsampling revealed that when controlling for *facet*, differences between *M. robustus* and *W. bicolor* were more evident for *Asfc, Sdr* and *Vvv,* but not *epLsar.* Controlling for the reclassifications of *facet* (*facet side, facet face,* and *facet side + face*)were less able to discriminate the taxa than, when controlling for *facet.* Comparable results were seen however for *facet side + face* in *Asfc.* For the within-species analyses, significant differences were apparent for all taxa except *T. thetis* for most variables. More significant differences were found for *Facet* (15), and *Facet side + face* (17), than *Facet side* (7), or *Facet face* (13). When considered across species, differences were only rarely evident, demonstrating that variation may be hard to detect when considered in multi-species datasets.

As mentioned in the main text, *facet* was the only factor aside from the necessary *species* and *specimen* that featured in each of the models favoured by the GLMM process. The effect of *facet* on the GLMM models is illustrated in figure S4.2. This demonstrates clear inter-facet differences for each of the four variables. It should be noted that that among the most variable facets were those with the lowest sample sizes (e.g., facet 2 *n*=23; facet 7 *n*=34) and the least variable were those with extensive sampling (facet 6 *n*=423; facet 9 *n=*235). This reiterates that if we wish to include broader sampling across facets we must ensure that each facet is adequately sampled to allow for precise modelling of inter-facet differences.



**Figure S4.2:** Effects of tooth on interspecific models of DMTA data. Mean and 95% confidence intervals of ecoregion effects of models produced by GLMM (see main text table 5). Left to right: Area-scale fractal complexity (*Asfc*); exact-proportion Length-scale anisotropy of relief (epLsar); developed interfacial area ratio (*Sdr*), and void volume of the valleys at a material ratio of 80% (*Vvv*). All variables normalised prior to analysis (see SI1), Facet 1 used as dummy variable.

The reclassifications were considered as a simpler categorisation of their functional significance. That they were less successful than the original classifications suggests that despite similarities in apparent function at equivocal parts of each loph, there are still functional differences between each facet which alter microwear.

Overall, facet was one of the most consistently significant variables found, echoing well known concerns of inter-facet variation (Gordon 1984, Krueger *et al.* 2008, Haupt *et al.* 2013). Ensuring consistent documentation of facets scanned then should be a priority for ongoing research. In addition, understanding differences between facets will help understand both DMTA differences and the functional morphology underlying it.

*Wear (including coarse wear)*

The subsampling analysis showed considerable improvement in differentiation between *M. robustus* and *W. bicolor* when the data were separated by *wear* stage for *Asfc,* and *Sdr,* but not *epLsar* or *Vvv.*

In isolation, differences between *wear* stages was significant only for 10 of 32 comparisons (and 8 of 32 for *coarse wear*). The GLMM process did not include *wear* in any of the final models, though *coarse wear* featured in the final model for *epLsar*. This is in contrast to the earlier statements regarding *wear* and anisotropy and may suggest subsampling to be a poor predictor of performance in GLMM models.

The effect of *coarse wear* in this final GLMM model for anisotropy is demonstrated in figure S4.3. While clearly overlapping, this demonstrates decreasing modelled anisotropy for more worn teeth. One explanation for this may be that as teeth wear the orientation of occlusal surfaces becomes decreased (McArthur and Sanson 1988), so the movement of foods down a facet is lessened, leading to lower anisotropy values. Another possibility is that *coarse wear* correlates with age (Death and Coulson 2016), and demonstrates change in diet as an animal ages.



**Figure S4.3:** Effects of coarse wear on interspecific models of anisotropy data. Mean and 95% confidence intervals of ecoregion effects of exact-proportion Length-scale anisotropy of relief (*epLsar*) models produced by GLMM (see main text table 5). Data normalised prior to analysis (see SI1), Wear stage 1 used as dummy variable.

The reclassified term *coarse wear* was utilised in the anisotropy model over *wear*, suggesting that the reclassified wear stages are in some way more informative than the original classification. This reclassification was undertaken with the intention of increasing sample size for higher wear stages. The success of this suggests that having the larger sample size for fewer wear stages is more informative than lower sample sizes across more wear stages. It also suggests that any differences between these higher wear stages (stages 4, 5 and 6 of McArthur and Sanson (1988) is minimal. This may be due to teeth of wear stages 4–6 are so highly worn that occlusion has been somewhat compromised (McArthur and Sanson 1988). That *coarse wear* was not used in any other models in contrast shows that the success of any reclassification is directly linked to the variable in question.

*Age*

As there are no taxa which feature in both the modern and palaeontological datasets here, *age* cannot be considered by subsampling or the within-species analyses. *Age* was however significant in three of four comparisons across the whole dataset. It should be noted however that lumping the data in this way can be misleading. Note for instance that when lumped, the single species comparisons for *Asfc* found no significant differences for *facet* despite significant differences being noted in *Asfc* for facet in a number of species in isolation as well as the final GLMM model. *Age* did not feature in any final GLMM models, suggesting that taphonomic factors are not of major concern in this dataset.

*Ecoregion*

The subsampling analysis showed some differences between variables when controlling for a single *ecoregion* (tropical grasslands). For example, differences between *M. robustus* and *W. bicolor* for complexity were much greater when controlled for *ecoregion*, while those for *Vvv* were possibly less evident than for the whole dataset. However, *ecoregion* was included in the final models for *Asfc*, *Sdr* and *Vvv.* These seemingly contradictory results for *Vvv* make an important point. It may be that rather than acting universally improving differentiation between species, some levels of factors may in fact indicate circumstances that do not distinguish taxa. For instance it may be that in tropical grasslands taxa cannot be differentiated, but that in temperate forests they can.

The modelled effects of *ecoregion* also indicate the effects of each ecoregion on DMTA data (Figure S4.4). These show similar effects for each variable, with temperate grasslands having a broad range encapsulating the majority of all other data. Tropical grasslands and forests had lower variation, but similar central ranges, while temperate forests and deserts had low and high values respectively. Referring to known characteristics of complexity (*Asfc*), this accords with our understanding. High complexity values for deserts would be consistent with consumption of harder desert grasses (e.g. spinifex) and shrubs in contrast to more succulent vegetation in temperate forests. The implications of central tendencies for temperate grasslands and both tropical ecoregions are unclear.



**Figure S4.4:** Effects of ecoregion on Interspecific models of DMTA data. Mean and 95% confidence intervals of ecoregion effects of models produced by GLMM (see main text table 5) Plots left to right: Area-scale fractal complexity (*Asfc*); developed interfacial area ratio (*Sdr*), and void volume of the valleys at a material ratio of 80% (*Vvv*). All variables normalised prior to analysis (see SI1), ‘N/A’ ecoregion used as dummy variable .

Why DMTA data differ between ecoregions is largely due to environmental factors. *Ecoregions* were originally designed as a conservation tool to categorise ecosystems which share similar species compositions (Olson *et al.* 2001). While this overall is unlikely to directly affect microwear, the distributions of these taxa likely reflect underlying parameters such as rainfall and temperature (Olson *et al.* 2001). In addition, the inclusion of plant communities in ecoregion classification (Olson *et al.* 2001) is likely to have a more direct impact on microwear differences. This leads us then to consider what it is about different ecoregions that causes differences to be present. A few possibilities for such could be the plant communities themselves, however it is unclear which factors of the plants present (e.g. toughness of tissues, presence of phytoliths, C3/C4 pathway) differentiate microwear. Alternatively it may be the relative abundance of particular plants or even the presence of other herbivores through niche partitioning which effects the microwear of a species across *ecoregions*. Another possibility which may be reflected in microwear is gross aridity, particularly considering the ongoing discussion on the effect of airborne dust on microwear (e.g. Ryan 1979, Ungar *et al.* 1995, Lucas *et al.* 2013). Ultimately, determining which of these factors are relevant to DMTA may be undertaken by considering data where such factors can be controlled, such in vitro studies, or specimens where gut contents are available for direct comparison.

#### References

##### Arman, S. D., P. S. Ungar, C. A. Brown, L. R. DeSantis, C. Schmidt and G. J. Prideaux (2016). Minimizing inter-microscope variability in dental microwear texture analysis. *Surface Topography: Metrology and Properties* **4,** 024007.

##### Bakke, M. (2006). Bite force and occlusion. *Seminars in orthodontics*, Elsevier.

##### Calandra, I. and G. Merceron (2016). Dental microwear texture analysis in mammalian ecology. *Mammal Review* **46,** 215-228.

##### Death, C. and G. Coulson (2016). A method for age estimation in the swamp wallaby (Wallabia bicolor). *Australian Mammalogy* **38,** 246-248.

##### Gordon, K. D. (1984). Hominoid Dental Microwear: Complications in the Use of Microwear Analysis to Detect Diet. *Journal of Dental Research* **63,** 1043-1046.

##### Haupt, R. J., L. R. DeSantis, J. L. Green and P. S. Ungar (2013). Dental microwear texture as a proxy for diet in xenarthrans. *Journal of Mammalogy* **94,** 856-866.

##### Krueger, K. L., J. R. Scott, R. F. Kay and P. S. Ungar (2008). Technical note: Dental microwear textures of “Phase I” and “Phase II” facets. *American Journal of Physical Anthropology* **137,** 485-490.

##### Lentle, R. G., K. J. Stafford, M. A. Potter, B. P. Springett and S. Haslett (1999). Ingesta particle size, food handling and ingestion in the tammar wallaby (*Macropus eugenii* Desmarest). *Australian Journal of Zoology* **47,** 75-85.

##### Lucas, P. W. (2004). *Dental functional morphology: how teeth work*, Cambridge University Press.

##### Lucas, P. W., R. Omar, K. Al-Fadhalah, A. S. Almusallam, A. G. Henry, S. Michael, L. A. Thai, J. Watzke, D. S. Strait and A. G. Atkins (2013). Mechanisms and causes of wear in tooth enamel: implications for hominin diets. *Journal of The Royal Society Interface* **10,** 20120923.

##### McArthur, C. and G. D. Sanson (1988). Tooth wear in eastern grey kangaroos (*Macropus giganteus*) and western grey kangaroos (*Macropus fuliginosus*), and its potential influence of diet selection, digestion and population parameters. *Journal of the Zoological Society of London* **215,** 491-504.

##### Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. Powell, E. C. Underwood, J. A. D'amico, I. Itoua, H. E. Strand and J. C. Morrison (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* **51,** 933-938.

##### Osborn, J. (1961). An investigation into the interdental forces occurring between the teeth of the same arch during clenching the jaws. *Archives of oral biology* **5,** 202-211.

##### Ryan, A. S. (1979). A Preliminary Scanning Electron Microscope Examination of Wear Striation Direction on Primate Teeth. *Journal of Dental Research* **58,** 525-530.

##### Schulz, E., I. Calandra and T. M. Kaiser (2010). Applying tribology to teeth of hoofed mammals. *Scanning* **32,** 162-182.

##### Ungar, P. S., M. F. Teaford, K. E. Glander and R. F. Pastor (1995). Dust accumulation in the canopy: A potential cause of dental microwear in primates. *American Journal of Physical Anthropology* **97,** 93-99.

##### Ungar, P. S., J. R. Scott, B. W. Schubert and D. D. Stynder (2010). Carnivoran dental microwear textures: comparability of carnassial facets and functional differentiation of postcanine teeth.

##### Zuur, A., E. Ieno, N. Walker, A. Saveliev and G. Smith (2009). Mixed effects models and extensions in ecology with R. New York: Springer. *574 p*.