Biodiversity and Conservation (2022) 31:183–209
https://doi.org/10.1007/s10531-021-02328-y

ORIGINAL PAPER

Fecal steroids as a potential tool for conservation paleobiology in East Africa

Andrew C. Kemp[1] · Christopher H. Vane[2] · Alexander W. Kim[2] · Christopher L. Dutton[3,4] · Amanda L. Subalusky[4] · Stuart K. Kemp[5] · Andrew C. Parnell[6]

Received: 6 April 2021 / Revised: 13 September 2021 / Accepted: 6 November 2021 /
Published online: 18 November 2021

Abstract

Conservation paleobiology seeks to leverage proxy reconstructions of ecological communities and environmental conditions to predict future changes and inform management decisions. Populations of East African megafauna likely changed during the Holocene in response to trends and events in the regional hydroclimate, but reconstructing these populations requires development of new proxies. We examine if fecal steroids are a viable proxy for megafauna populations since they are well preserved in sedimentary archives. We measured eleven fecal steroids in 87 fresh dung samples representing 22 species of megafauna in the Maasai Mara National Reserve (Kenya) and a further seven samples from captive animals. Using this reference library, four distinctive groups are identified, which reflect diet and biochemical modification of these inputs during digestion by the gut microbiome. Carnivore dung is characterized by more than ~ 75% cholesterol and primate dung includes uniquely high proportions of coprostanol. Two groups of herbivore are distinguished by their differing proportions of phytosterols that are consumed by eating plants and 5β-stanols produced during digestion. Under cross validation a random forests statistical model accurately classified 72% of dung samples to the species level using fecal steroids. Variability among individuals and between wild and captive animals suggests that fecal steroids in herbivore dung may reflect diversity and variability in diet, while a lack of variability in carnivore dung indicates that they cannot be identified to the species level in most instances. Our results suggest that fecal steroids may have utility in reconstructing the time-evolving composition megafauna populations in East Africa.

Keywords Kenya · Holocene · Biomarker · Dung · Maasai Mara National Reserve

Communicated by Karen E. Hodges.

✉ Andrew C. Kemp
andrew.kemp@tufts.edu

Extended author information available on the last page of the article

Introduction

Since the late Pleistocene, the number and diversity of terrestrial megafauna (used here informally to differentiate from microfauna such as diatoms that are widely used as proxies in paleoenvironmental research) declined across all continents except Antarctica (Malhi et al. 2016). In many instances an initial decline was triggered by the arrival of early humans (Bartlett et al. 2016; Sandom et al. 2014), but accelerated during the Anthropocene through hunting and habitat loss. Extinction in Africa was modest compared to other continents; Malhi et al. (2016) report that all eleven species of the continent's large Pleistocene carnivores (adults heavier than 21.5 kg) remain today, as do 56 of the 74 species of large herbivores (adults heavier than 45 kg). The remaining megafauna of East Africa are vital for maintaining healthy ecosystem services and are valuable to regional economies. However, recent and projected changes in East African climate (particularly the amount and seasonality of rainfall) are anticipated to impact megafauna populations during the remainder of the twenty-first century and beyond. Throughout the Holocene, trends and events in East Africa's hydroclimate (including abrupt, century-scale changes; Humphries et al. 2019; Tierney and deMenocal 2013; Tierney et al. 2013) likely influenced megafauna populations. These changes may provide analogs to help anticipate how future climate changes will impact the iconic megafauna of East Africa.

Conservation paleobiology uses proxy reconstructions of animal communities (composition, abundance, and geographic range) and paleoenvironments (e.g., frequency and severity of drought) to provide a context for recent and predicted changes that extends beyond the limited duration of historical observations and measurements (Dietl and Flessa 2011; Dietl et al. 2015). These reconstructions can elucidate, for example, how species responded to natural climate variability across a range of timescales and they can capture conditions and trends that are inadequately represented in the short observational record, but may be analogous to future changes. In the specific case of terrestrial megafauna, a major challenge for conservation paleobiology is how to characterize past populations because these animals are relatively few in number, widely distributed and mobile when alive, and poorly preserved in the sedimentary record after death. Consequently, finding the remains of megafauna is not a viable approach to quantitatively describe and analyze community changes, particularly on short (centennial to millennial) timescales. An overarching goal for conservation paleobiology is to develop new proxies that overcome the limitations of using sub-fossil remains to infer changes in megafauna populations. One possibility is to utilize dung preserved in sedimentary archives as a proxy for megafauna community composition since it is likely to be abundant, widespread, and preserved in sediment cores recovered from depositional environments (D'Anjou et al. 2012; Schroeter et al. 2020; Vane et al. 2010), unlike sub-fossil remains of the animals themselves.

Fecal steroids are a group of biomarkers present in dung (Bull et al. 2002) that include sterols and stanols (we use the term fecal steroid for simplicity). In the digestive tract of animals, sterols undergo modification by gut-dwelling microbes in anoxic conditions to produce 5β-stanols that are excreted in dung. Modification of the same sterols in oxic, exogenic environments instead produces 5α-stanols (Gaskell and Eglinton 1975). The make-up of fecal steroids in dung is therefore controlled by dietary intake, metabolic processes that provide the sterol precursors (for example, cholesterol is present even in animals with low-cholesterol diets), and the type and effectiveness of

biochemical alterations that occur in the digestive tract (incomplete alteration results in precursor sterols being present in dung). Therefore, variability in diet and digestive biochemistry among species may produce dung with a diagnostic profile of fecal steroid composition.

Analysis of modern dung demonstrated that fecal steroids can objectively identify the origin of dung in some terrestrial (Harrault et al. 2019; Leeming et al. 1996; Shah et al. 2007) and marine (Leeming et al. 2015) ecosystems, often to the species level. These studies commonly focused on differentiating between natural, agricultural, and human fecal inputs into the environment to establish modern pollution sources (Leeming et al. 1996; Vane et al. 2010), or to recognize human occupation and behavior in archaeological contexts (Bull et al. 2001, 1999; D'Anjou et al. 2012; Schroeter et al. 2020). Consequently, existing data primarily characterize fecal steroids from domesticated livestock, humans, and wild megafauna in a handful of regions (e.g., central and northern Asia; Harrault et al. 2019; Schroeter et al. 2020) other than East Africa. Fecal steroids have low solubility in water and bind readily to particulate matter, which makes them well-preserved in depositional environments such as wetlands and lakes (Bartlett 1987; D'Anjou et al. 2012; Schroeter et al. 2020; Vane et al. 2010) that receive sediment (and likely fecal material) from a wider catchment area. This preservation and catchment-scale integration opens up the possibility that the characteristic fingerprint of fecal steroids from particular species/groups of megafauna can be recognized in dated sediment cores to infer community changes through time.

Our goal is to investigate the potential utility of sediment-hosted fecal steroids to reconstruct megafauna communities in East Africa over the Holocene. Proxy development is a multi-stage process that begins with testing if a potential proxy (e.g., fecal steroids) can reliably recognize or reconstruct the variable of interest (e.g., species of megafauna) in samples of known origin (e.g., fresh dung). In the absence of existing data, we aim to (1) produce a modern training set of fecal steroids from the dung of wild East African megafauna that captures within- and among-species variability; and (2) investigate how reliably fecal steroids identify the origin of dung.

Study area

The Maasai Mara National Reserve and Serengeti National Park (Fig. 1a) are a transnational savanna grassland with a high abundance and diversity of wildlife (Ogutu et al. 2011). Rainfall (amount and seasonal distribution; Fig. 1b) is the primary driver controlling reproduction, survival, and migration of African savanna ungulates (and their predators) within the region (Holdo et al. 2009; Ogutu et al. 2008). The spatio-temporal pattern of precipitation in East Africa reflects interactions between seasonal reorganization of prevailing atmospheric circulation and regional topography (Camberlin 2018; Funk et al. 2016). This complexity drives a corresponding diversity of landscapes and ecosystems and indicates that even modest changes in supra-regional atmospheric and oceanic conditions could drive notable shifts precipitation regimes with cascading impacts on regional ecosystems, including megafauna populations. Broadly, East Africa is an anomalously dry region of the humid tropics that receives less than ~ 2 mm/day of precipitation on average compared to locations at similar latitudes in West Africa (~ 5 mm/day) and Indonesia (up to 12 mm/day; e.g., Fig. 1 of Yang et al. 2015). This anomaly arises from zonal Walker circulation in which low-level convergence in the

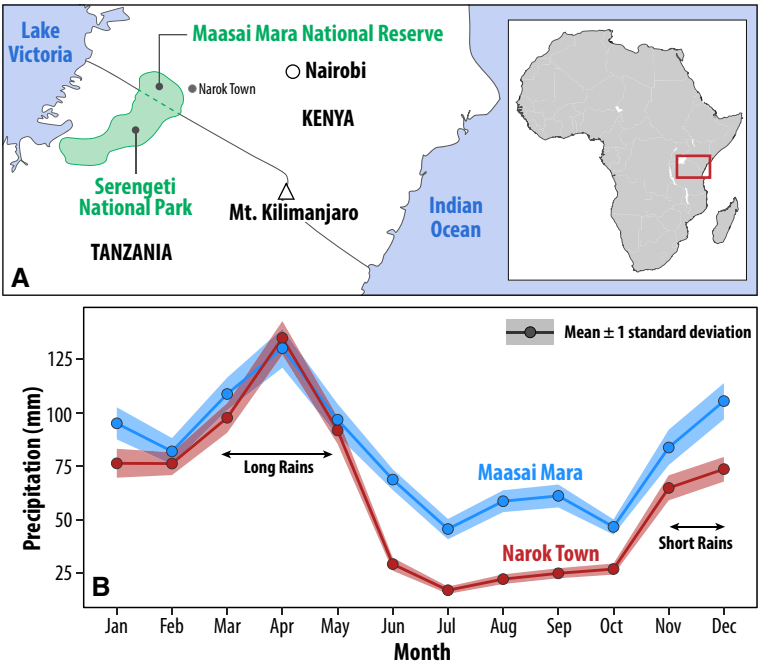

Fig. 1 **A** Location of the Maasai Mara National Reserve in Kenya. **B** Average monthly precipitation in the Maasai Mara National Reserve (1965–2015, compiled from 15 rain gauges) and nearby Narok Town (1913–2015). Data from Bartzke et al. (2018). The approximate timing of the "long rains" and "short rains" are shown for reference

Indo-Pacific warm pool is associated with moist, rising air and low surface pressure, while divergence in East Africa is associated with dry, descending air and high surface pressure.

Shifts in atmospheric circulation during the year result in a strong seasonality of precipitation (Bartzke et al. 2018; Camberlin 2018; Yang et al. 2015; Fig. 1b). For the Maasai Mara National Reserve, the presence of the intertropical convergence zone in the southern hemisphere results in northeasterly winds during January and February. These winds move from the Arabian Sea across Somalia bringing dry weather. In March to May, southeasterly winds deliver moist air from the seasonally-warm Indian Ocean resulting in substantial precipitation (the "long rains"; Fig. 1b). This rain stimulates vegetation growth and draws ~ 1.4 million wildebeest (*Connochaetes taurinus*) along with large numbers of zebra (*Equus burchelli*) and Thomson's gazelles (*Gazella thomsoni*) from the Serengeti to the Maasai Mara National Reserve during their famed migration. Changes to the strength and position of the Somali Jet and the presence of the intertropical convergence zone in the northern hemisphere during June to September reduce onshore moisture transport and bring relatively cool, dry air from the south assisted by seasonally-cool sea surface temperatures. As the availability of food and water decrease in the Maasai Mara National Reserve the wildebeest herd returns southward to the Serengeti. In October to December, weakening of the Indian Monsoon and reversal of Indian Ocean trade winds provides moisture to East Africa (the "short rains"; Fig. 1b).

On multi-annual timescales, Lyon and DeWitt (2012) proposed that precipitation during the long rains declined since ~ 1999 CE because of concurrent changes in sea-surface temperature impacting the transfer of moisture to the continent. On longer timescales, paleoenvironmental reconstructions indicate that drying of East Africa during the twentieth century was unusual in the context of the Common Era and that regionally-coherent phases of wetter and drier conditions lasting centuries to millennia occurred in the past ~ 20,000 years (De Cort et al. 2013; Humphries et al. 2019; Tierney and deMenocal 2013; Tierney et al. 2015). Projections of twenty-first century climate change suggest that East Africa is likely to experience wetter conditions overall, but with reduced long rains and increased short rains in Kenya and Tanzania (Gebrechorkos et al. 2019).

Methods

Sample collection

On January 14–17th 2016, we collected 87 samples of dung in and around the Maasai Mara National Reserve (Fig. 1). These samples represent 19 species of wild animals that are protected within the reserve and a further three domesticated species from small agricultural herds on the periphery of the reserve. In many cases, we observed the dung being deposited directly, but where this was not possible, its origin/age was confirmed by reference to published guides (Stuart and Stuart 2013) and consultation in the field with experienced park rangers and trackers. Most samples were judged to be less than 24 h old, except for some specimens from lions (*Panthera leo*), baboons (*Papio anubis*), hyenas (*Crocuta crocuta*), vervet monkeys (*Chlorocebus pygerythruses*; less than 3 days), and crocodiles (*Crocodylus niloticus*; approximately 6 weeks old based on observed nesting times at the collection site).

To capture a degree of variability among individual animals, we aimed to collect five samples from each species, although this was not possible in all instances (particularly carnivores). No wild species represented by only one individual was analyzed. To ensure that the samples were from different individuals, we sought to collect them from sites located throughout the Maasai Mara National Reserve and on different days. We collected two samples of red oat grass (*Themeda triandra*) to characterize a primary food source of herbivores in the study region. A further seven dung samples (one per species) were provided by Woburn Safari Park and West Midlands Safari Park in the United Kingdom in August 2015 and January 2016 respectively with permission from the British and Irish Association of Zoos and Aquariums. The purpose of collecting these samples was to investigate the potential role of diet in determining the steroid content of dung since captive and wild animals of the same species likely eat different food. The chosen species reflect the availability of captive animals with wild equivalents in Kenya, except for a captive African Wild Dog (*Lycaon pictus,* retained without a wild counterpart to evaluate steroids in canine carnivores). Each sample was sealed in a labeled polyethylene bag and transported in an ice-filled cool box prior to be being frozen at approximately − 18 °C on the day of collection. The frozen samples from Kenya were shipped overnight to the British Geological Survey (Keyworth), where they remained frozen until processing.

Analytical preparation and reproducibility

In the laboratory all samples were freeze dried for 72 h. Approximately 30 g (dry weight) of each dung sample (the entirety of each plant sample) was ground to a coarse powder using a blender (De Longhi KG40) and then further ground into a fine (able to pass through a 63 μm brass sieve), homogenized powder using an agate ball mill (Retsch PM400). Smearing of hyena dung during milling was negated by adding 10 g solvent-cleaned sand.

A 0.2 g aliquot of powdered sub-sample was placed on a watch-glass and spiked with deuterated surrogate standards cholesterol-d6 and 5α-androstanol at 100 μg/g and 50 μg/g respectively in toluene:pyridine (20:1). Samples were then extracted with methanol/dichloromethane (1:1 v/v) using an accelerated solvent extraction system (Dionex ASE 200) operated at 100 °C and 1500 psi. Activated copper powder (2 g) was added to remove elemental sulfur. The extract was reduced to dryness using a Turbovap. Following Tyagi et al. (2007), samples were then saponified by adding 5 mL of 2 N sodium hydroxide in 90% methanol to the total lipid extract and heating at 60 °C for 1 h. Analytes were back-extracted in hexane (3 times 10 mL), and the solvent was evaporated to dryness under a gentle stream of nitrogen. Internal standards, 5α-cholestane-d6 and 5α-androstanol were added at 100 μg/g and 50 μg/g respectively in toluene:pyridine (20:1) and made up to 1 mL with pyridine. A 20 μL aliquot was transferred into a 200 μL glass vial insert containing 140 μL of pyridine and derivatised with 40 μL N, O-bis(trimethylsilyl)trifluoroacetamide with 1% trimethylchlorosilane (Sigma Chemical Co.) at 60 °C for 1 h and left at room temperature for 12 h prior to gas chromatography and mass spectrometry analysis.

In most instances a single, representative sub-sample of the homogenized material was prepared for analysis. However, to investigate analytical reproducibility we repeated the extraction process on multiple sub-samples of the homogenized dung. We chose nine species to investigate the analytical and instrumental reproducibility of fecal steroid measurements. The species were selected to ensure that each steroid was adequately represented. From the homogenized dung sample of one individual per species, we extracted steroids from three representative sub samples. In addition, we performed an extraction on six sub-samples from the homogenized dung of a second elephant (sample MMF–1E). Up to nine replicate measurements (in different batches to ensure they were separated by time and other samples) were made on some of these extractions to investigate instrumental reproducibility. Since the original sample was homogenized ("Gas chromatography and mass spectrometry" section), variability among extractions represents a test of analytical reproducibility and replicate measurements are a test of instrumental reproducibility.

Gas chromatography and mass spectrometry

Fecal steroids were analyzed using a Varian CP3800 series gas chromatograph coupled with a Varian 1200L triple Quadrupole mass spectrometer. Sample injection (1.0 μL) was in splitless mode. Compounds were separated using an Agilent DB-5MS column (30 m length × 0.25 mm inside diameter × 0.1 μm film thickness). The oven was programmed to warm from 60 °C (1-min isothermal) to 250 °C at 20 °C/minute, then to 310 °C at 4 °C/min before being held isothermally at 310 °C for 10 min. The mass spectrometer was operated at 70 eV with a mass range of m/z 30–550 (beam current 150 μA, source temperature 150 °C) with helium as carrier gas at a flow rate of 1 mL/min. Data acquisition was carried out using a Varian MS workstation v6.5. Identification of compounds was made by

retention time and mass spectra comparison with those of available standards or data in the literature (Vane et al. 2010). Specific fragment ions used for quantification and qualification/confirmation are presented in Supplementary Table 1. Quantification was achieved with 5-point internal calibration curves of available standards with relevant fragments, and comparison with a constant internal standard concentration added prior to analysis (Supplementary Table 1, Supplementary Fig. 2). The limit of quantification for individual compounds ranged from 0.01 to 0.04 μg/g (dry weight.); procedural blanks and reagent blanks contained no discernible steroids.

Fecal biomarker nomenclature

Common compound names are used throughout this work to facilitate comparison with previous studies. The eleven fecal steroids measured were cholestane (5α-cholestane), coprostanol (5β-cholestan-3β-ol), epicoprostanol (5β-cholestan-3α-ol), cholesterol (cholest-5-en-3β-ol), 5α-cholestanol (5α-cholestan-3β-ol), coprostan-3-one (5β-cholestan-3-one), campesterol (24α-methyl-5-cholesten-3β-ol), stigmasterol (3β-hydroxy-24-ethyl-5,22-cholestadiene), fucosterol ((3β,24E)-stigmasta-5,24(28)-dien-3-ol), β-sitosterol (24-ethylcholest-5-en-3β-ol) and 5α-stigmastanol (24α-ethyl-5α-cholestan-3β-ol). Chemical structures are presented in Supplemental Fig. 2. We did not determine the concentration of other fecal sterols (such as 5α-campestanol) because suitable standards were unavailable at the time of analysis (Prost et al. 2017).

Statistical analysis

Statistical analysis was performed after transforming measured steroid concentrations (units of μg/g, reported in Supporting Table 2) to percentages because of large differences among samples in absolute concentration. Analysis included all measured steroids. We used random forests (Breiman 2001; Cutler et al. 2007) to test if a dung sample could be accurately attributed to a species using the relative abundance of fecal steroids. A classification tree sequentially partitions a multi-variate dataset into subgroups. Each branch of the tree divides at a node resulting in more subgroups and splitting continues until the subgroups cease to become more homogenous. Random forests builds many classification trees (the forest) in which subsets of the original variables (i.e., types of fecal steroid) and samples are excluded. The importance of each steroid for determining the origin of a dung sample is quantified for the dataset as a whole and for each species (incorporated as a priori groups).

We analyzed a dataset comprising 87 observations (individual wild animals) of 11 fecal steroids using the *randomForests* package (Liaw and Wiener 2018) for R (ntree = 1000, with replacement, mtry = 3). Where multiple extractions and measurements were made on the same dung sample they were averaged prior to analysis. Model performance is quantified using the rate of mis-classification (in permutations of out-of-bag data).

The importance of each steroid is quantified using (1) the decrease in accuracy caused by excluding it from analysis during cross validation and (2) the decrease of node Gini impurity from splitting on it. A large decrease in accuracy indicates that the excluded variable was important for differentiating among species. Gini impurity is the probability that an observation is misclassified if grouped randomly. Therefore, it is a measure of heterogeneity in a set of items, where a value of zero indicates homogeneity and larger values (up to almost 1) indicate increasing heterogeneity. Since the goal

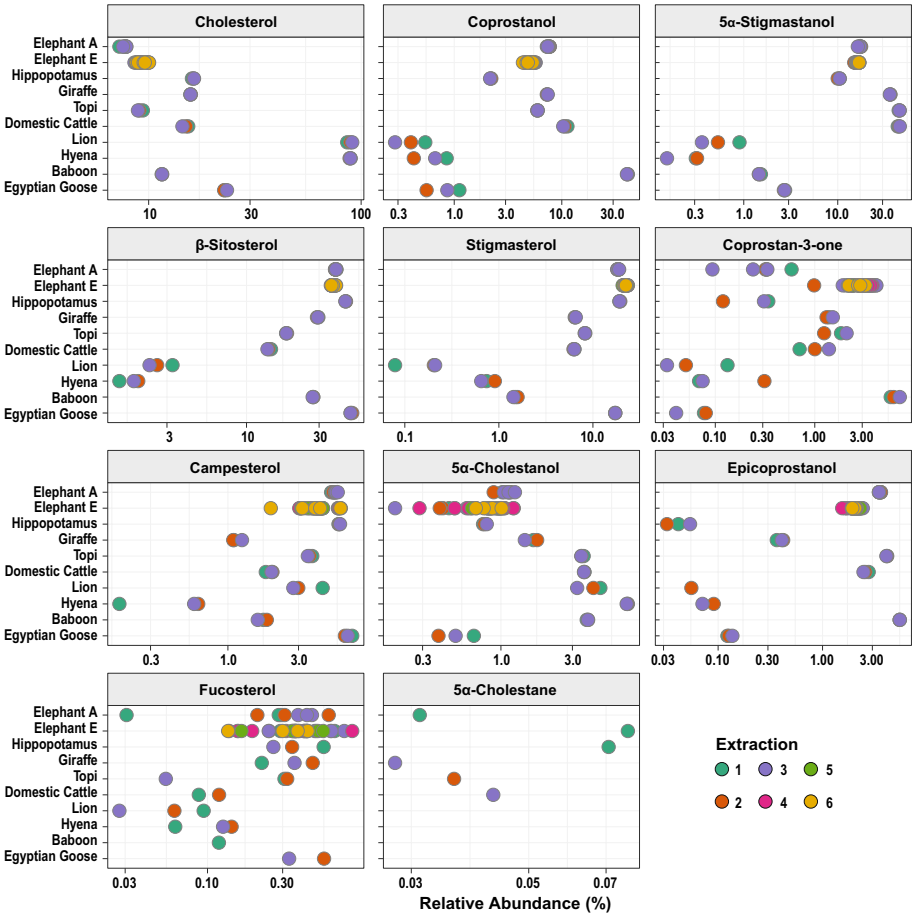

Fig. 2 Analytical and instrumental reproducibility of fecal steroid measurements performed on animal dung collected in the Maasai Mara National Reserve, Kenya. The abundance of each steroid (individual panels) is expressed as a percentage of total measured steroids; note that (log) scale varies among panels. Species are organized approximately by type (herbivores, carnivores, primates, birds). For each species, one dung sample from a single individual was homogenized and extractions were performed on three sub-samples (symbol fill). Up to three replicate measurements were made on each extraction. The exception to this sampling regime is elephant, where two individuals (A and E) were analyzed. From the homogenized dung sample of the individual elephant E there were six extractions and up to eight replicate measurements

of classification is to provide homogenous groups, lower Gini impurity is desirable. Reported values are the decrease in node impurities from splitting on the variable, such that large values are desirable because they indicate more homogenous (purer) groups. For both measures, reported values are the mean across the decision trees divided by their standard deviation and are therefore dimensionless. Mean decrease in accuracy and Gini impurity are calculated for the entire dataset (i.e., across all species) and for individual species. Variable importance is more intuitively examined through relative ranking of variables within a dataset, rather than absolute values.

Results

Analytical and instrumental reproducibility

The five most abundant steroids in the subset of samples used to investigate analytical and instrumental reproducibility were cholesterol (7.3–91%), coprostanol (0.3– 41%), 5α-stigmastanol (0.2–46%), β-sitosterol (1.5–50%), and stigmasterol (0.1–24%; Fig. 2). For individual animals, we calculated the difference between the smallest and greatest abundance of each sterol across all extractions and replicate measurements. This range is used to quantify variability rather than the standard deviation since there are relatively few observations. Across all species, the mean range was 1.05% (< 0.1–4.8%) for cholesterol, 0.52% (< 0.1–1.4%) for coprostanol, 0.86% (< 0.1–2.2%) for 5α-stigmastanol, 0.9% (0.3–3.0%) for β-sitosterol, and 0.52% (< 0.1–2.8%) for stigmasterol. There is no indication that measurements of these five fecal steroids are more or less reproducible for any species. We conclude that our approach (sub sampling of homogenized dung samples from which fecal steroids are then extracted and measured using the methods described in "Methods" section) generates reproducible results for the most common fecal steroids and therefore differences among species or individuals are unlikely to be attributable to analytical or instrumental bias/error.

The remaining six measured fecal steroids had a maximum relative abundance of less than 7% across all species and samples (Fig. 2), of which two had a maximum abundance of less than 1%. For individual animals, the mean range of measured abundance was 0.74% (< 0.1–7.2%) for coprostan.3.one,

0.8% (0.2–7.0%) for campesterol, 0.4% (0.2–7.0%) for 5α-cholestanol, 0.2% (< 0.1–5.5%) for epicoprostanol, 0.2% (< 0.1–0.8%) for fucosterol, and < 0.1% (all < 0.1%) for 5α-cholestane. Therefore, the absolute reproducibility of these measurements among sub-samples, extractions, and replicates is similar to the more common fecal steroids, but in some cases it is large compared to the underlying relative abundance. We conclude that subtle changes in the least common fecal steroids among individuals and species may not exceed variability that arises from sample processing in a laboratory and instrumental measurement.

Variability among individuals

The most abundant fecal steroid in carnivorous mammal dung is cholesterol (more than 85% for all individual lions (*Panthera leo*), hyenas, leopards (*Panthera pardus*), cheetahs (*Acinonyx jubatus*), and African wild dogs) and our analysis demonstrates that the abundance of this steroid in dung from captive animals (lion and cheetah) lies within the range of measured values for wild animals in Kenya (Fig. 3). For example, dung from wild lions included 88.6–96.4% cholesterol, compared to 89.7% in a captive individual. The observed differences between wild and captive carnivores are small for other fecal steroids except for 5α-cholestanol, which was more common in captive animals than in their wild counterparts. For example, three wild lions produced dung with 1.1–4.0% 5α-cholestanol compared to 7.6% for a captive lion. 5α-cholestanol is a thermodynamically-stable stanol produced by microbial reduction of cholesterol outside of the gut (Leeming et al. 1996). Therefore, differences among wild and captive lions/cheetahs in 5α-cholestanol are unlikely to be due to the influence of diet.

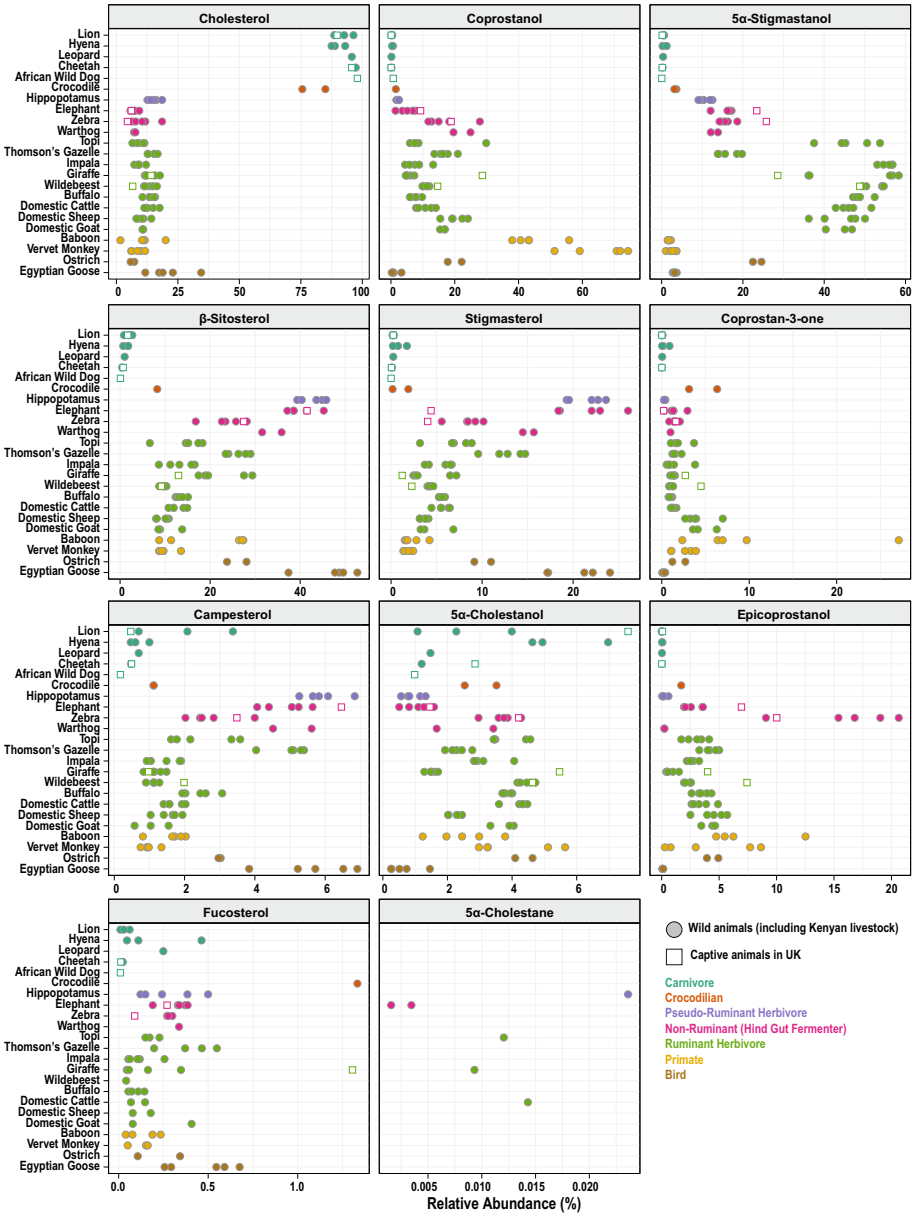

Fig. 3 Steroid composition of animal dung collected from 94 individual animals, representing 23 species including wild examples from the Maasai Mara National Reserve in Kenya (filled circles) and captive individuals in two UK safari parks (open squares). The abundance of each steroid (individual panels) is expressed as a percentage of total measured steroids; note that (log) scale varies among panels. Symbol color denotes biological grouping of species. Individuals are represented by a single data point which (where available) is an average from all sub samples and replicate measurements of a single dung

The other captive animals analyzed were herbivores (giraffe (*Giraffa camelopardalis*), zebra, wildebeest, and elephant (*Loxodonta africana*)) and differences between captive and wild individuals suggests that diet (or captivity in a different climate zone) may influence fecal steroid composition (Fig. 3). Dung from a captive giraffe was relatively enriched in coprostanol (28.7% compared to 5.2–7.3% in five wild individuals) and depleted in 5α-stigmastanol (28.6% compared to 36.1–58.3% in wild individuals), β-sitosterol (13.0% compared to 17.4–29.4 in wild individuals), and stigmasterol (1.2% compared to 2.5–7.2% in wild individuals). Dung from captive elephants and zebras included larger proportions of 5α-stigmastanol than wild individuals (23.4%/25.8% compared to 12.0–17.2%/14.2–18.6% respectively), but relatively less of its phytosterol precursor stigmasterol (4.4%/4.0% compared to 18.4–26.0%/5.6–10.2% respectively). Fecal steroids from a captive wildebeest also contained relatively little cholesterol (6.6% compared to 11.2–16.5%) and relatively more coprostan.3.one, campesterol, and epicoprostanol (7.4% compared to 1.9–2.6%) than five wild individuals. Even for the less common steroids, differences between wild and captive individuals (carnivore and herbivore) are likely larger than uncertainties introduced through sample processing and measurement ("Analytical and instrumental reproducibility" section).

Variability among species

We identify four principle groups of animals (wild individuals only) based on measured fecal steroid content using partitioning around medoids (Kaufman and Rousseeuw 2005;

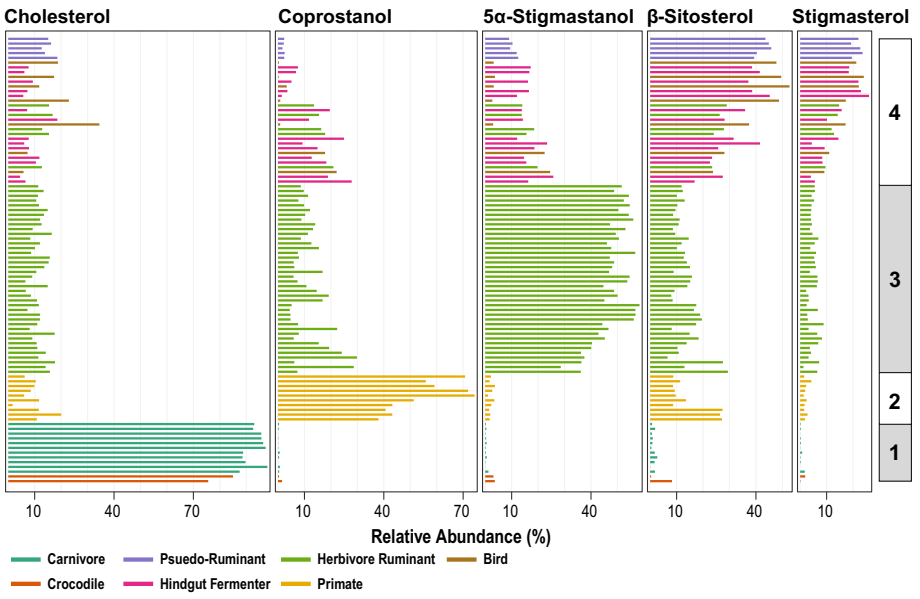

Fig. 4 Steroid composition of animal dung collected from 94 individual animals (bars), representing 23 species including wild examples from the Maasai Mara National Reserve in Kenya and captive individuals in two UK safari parks. For individuals where multiple sub-samples were processed and replicate measurements made, the presented value is an average across all measurements. Using partitioning around medoids, we recognized four distinctive groups. Bar color represents a priori classification of species

Fig. 4). Fecal steroids in dung from carnivores (lion, hyena, leopard, cheetah, and crocodile) are almost exclusively cholesterol (Group 1 in Fig. 4). Across all species and individuals of wild carnivores, the mean abundance of cholesterol was 90.5% (75.6–97.3%). There was no notable difference in cholesterol abundance among species of mammalian carnivores, but crocodiles have lower cholesterol content. In comparison, non-carnivores were characterized by cholesterol content less than 34.5%.

The most common fecal steroid in primates (baboon and vervet monkey) was coprostanol (mean/range of 54.8/38.0–74.3%; Group 2 in Fig. 4). Among non-primates, the maximum occurrence of coprostanol was 29.9% in a dung sample from a topi (*Damaliscus lunatus jimela*). Although there is overlap in measured values, baboon dung included relatively less coprostanol (38.0–55.9%) and 5α-cholestanol (1.23–3.8%) than vervet monkey dung (51.3–74.3% and 3.0–5.6% respectively). Compared to vervet monkeys, baboon dung was enriched in coprostan-3-one (Fig. 3). Four of the five highest measured abundances (> 6%) of coprostan-3-one were in baboon dung (including 27.1% for one individual), while the highest abundance of coprostan-3-one measured in vervet monkey dung was 3.9%.

We identified two groups of herbviores that do not correspond strictly to guilds or functional groups (e.g., browsers vs grazers). For many species of even-toed ungulates that are herbivorous ruminants, the most common fecal steroid was 5α-stigmastanol (Group 3 in Fig. 4). Among topi, impala (*Aepyceros melampus*), giraffe, wildebeest, buffalo (*Syncerus caffer*), zebra, domesticated cattle (*Bos taurus*), sheep (*Ovis aries*), and goats (*Capra hircus*) the mean of 5α-stigmastanol was 47.6% (28.6–58.3%). For these species there is a gradient of 5α-stigmastanol, with impala dung having the highest abundance (mean of 55.4%) and domesticated goats having the lowest (mean of 44.1%). In comparison, the mean 5α-stigmastanol content for species not in this group was 8.8%. Although all species in this group are ruminants, it is notable that not all ruminants were allocated to this group (all five individuals of Thomson's gazelle were placed into group 4).

A fourth group of species is distinguished by elevated concentrations of β-sitosterol (mean/range of 34.3%/17.4–52.7%; Group 4 in Fig. 4). This group is comprised of species with diverse digestive systems including elephants, warthog (*Phacochoerus africanus*), zebra (hindgut fermenters), hippopotamus (*Hippopotamus amphibius*; often described as pseudo-ruminants since they have a three-chambered stomach), Thomson's gazelle (herbivorous ruminant) as well as ostrich (*Struthio camelus*) and Egyptian goose (*Alopochen aegyptiaca*). For species not in this group, the mean concentration of β-sitosterol was 11.2%. Elephant and hippopotami dung yielded high concentrations of β-sitosterol (minimum of 37.2% and mean of 41.5%) and stigmasterol (minimum of 18.4% and mean of 21.5%). Zebra dung is unusually enriched in epicoprostanol (Fig. 3), five wild individuals included a mean of 16.2% (9.1–20.6%), compared to a mean/maximum for all other species of 2.6%/12.5%. Among hindgut fermenters, warthog dung included an elevated concentration of coprostanol (19.6% and 24.9% in the two individuals studied).

Statistical model

We used random forests to determine the likelihood that fecal steroids could reliably identify the origin of a dung sample to the species level. Input for this analysis was fecal steroids measured in dung samples from 87 wild animals (averaged across multiple subsamples, extractions, and measurements where available) spanning 22 species (we include domesticated livestock from Kenya as wild animals and exclude only the captive individuals from the UK). For the whole dataset, 72% of dung samples were correctly classified to

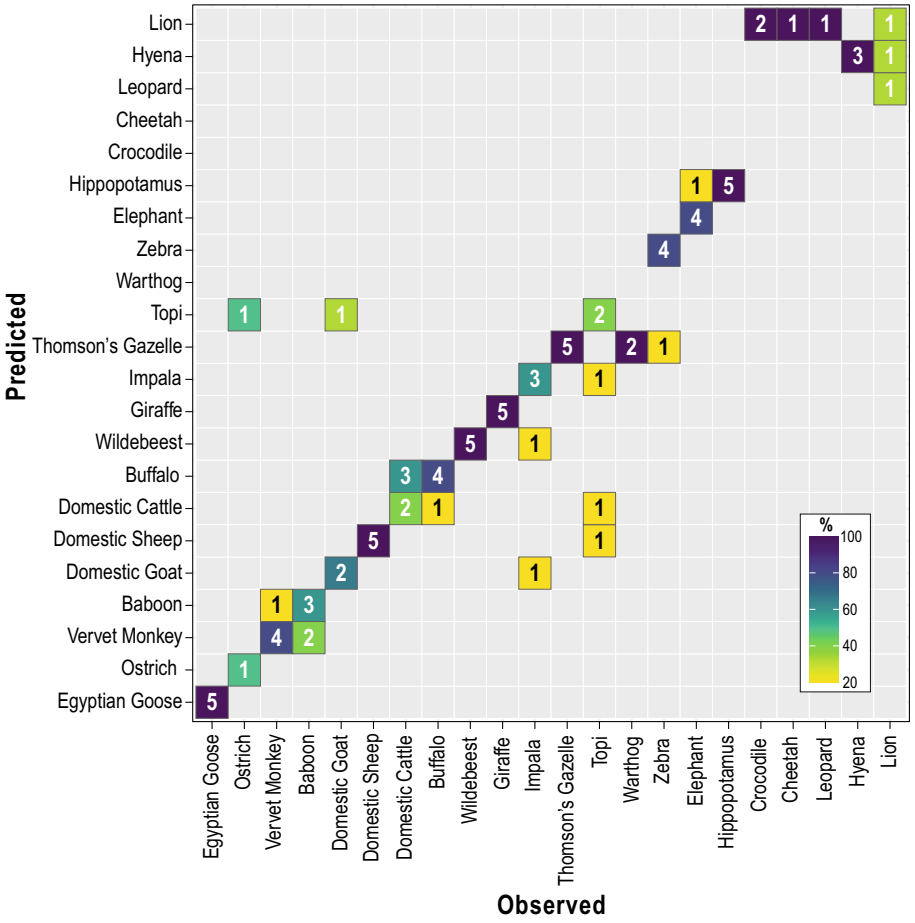

Fig. 5 Classification of dung samples from animals in Kenya (including domesticated livestock) using random forests. Each of the 87 dung samples of known origin (observed, x-axis) was assigned to a species (predicted, y-axis) based on its fecal steroid composition. Since the number of collected dung samples varies among species from two to five, the frequency of sample classification is expressed as a percentage (shading). Labels within cells display actual number of samples. Values equal to 0% are not shaded or labeled

the species level during cross validation (Fig. 5). Random forests quantified the importance of each fecal steroid in classifying dung samples using the Gini impurity index and mean decrease in accuracy. For the entire dataset, the steroid with the greatest predictive power was 5α-stigmastanol (Gini impurity index = 10.7; Fig. 6a), after which six fecal steroids had similar importance as evidenced by a narrow range of Gini impurity index (from 10.0 for coprostanol to 8.4 for campesterol). Despite spanning the widest range of measured abundance (1.6% in a baboon dung to 97.3% in cheetah dung), cholesterol had a relatively low importance of 6.6, likely because of the bipartite division into species with high (carnivores) or low (non-carnivores) cholesterol content in their dung. The least important variables (fucosterol with Gini impurity index = 2.7 and 5α-cholestane with Gini impurity index = 0.16) were also the least abundant. For example, 25.8/93.5% of animals across all species produced dung with no fucosterol/5α-cholestane and these fecal steroids had a

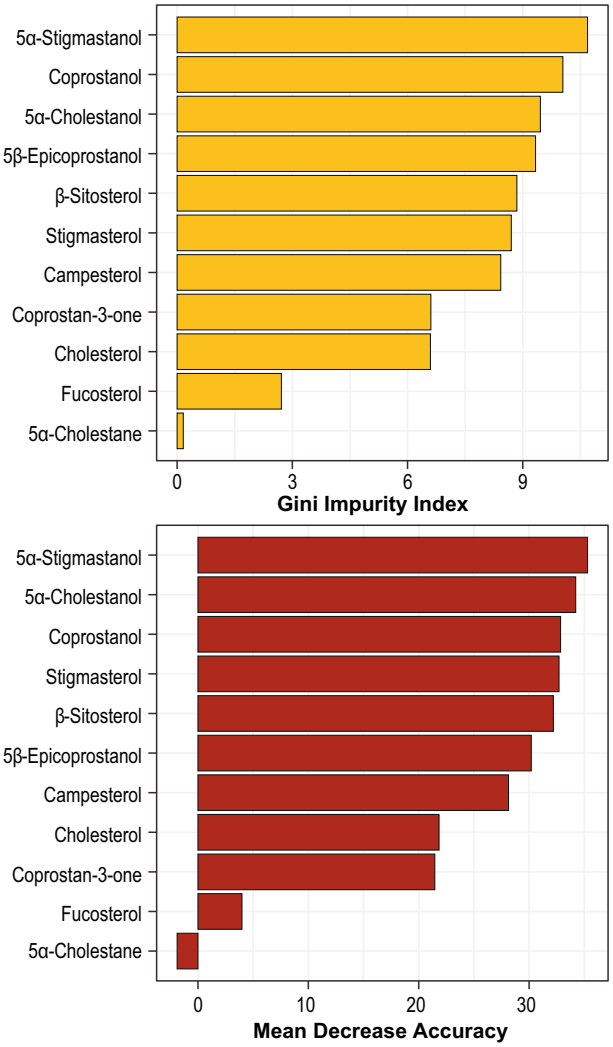

Fig. 6 Predictive power of fecal steroids measured using the Gini Impurity index (top) and mean decrease in accuracy (bottom) to quantify importance for the entire modern training set used in random forests analysis

maximum abundance of 1.3/0.02% (Fig. 3). The relative importance of each fecal steroid was similar when measured using the mean decrease in accuracy (Fig. 6b).

There are notable differences in how frequently some species are (mis)classified compared to the overall success rate of 72% (Fig. 5). Dung samples from Egyptian goose, domestic sheep, wildebeest, giraffe, Thomson's gazelle, hippopotamus, and hyena were correctly classified in all instances. The most important steroids for recognizing several of these species were 5α-stigmastanol, 5α-cholestanol, and 5β-sitosterol (Fig. 7). For Thomson's gazelle, campesterol was unusually important. Conversely, our analysis indicates that carnivores are particularly challenging to differentiate since only 40% of samples

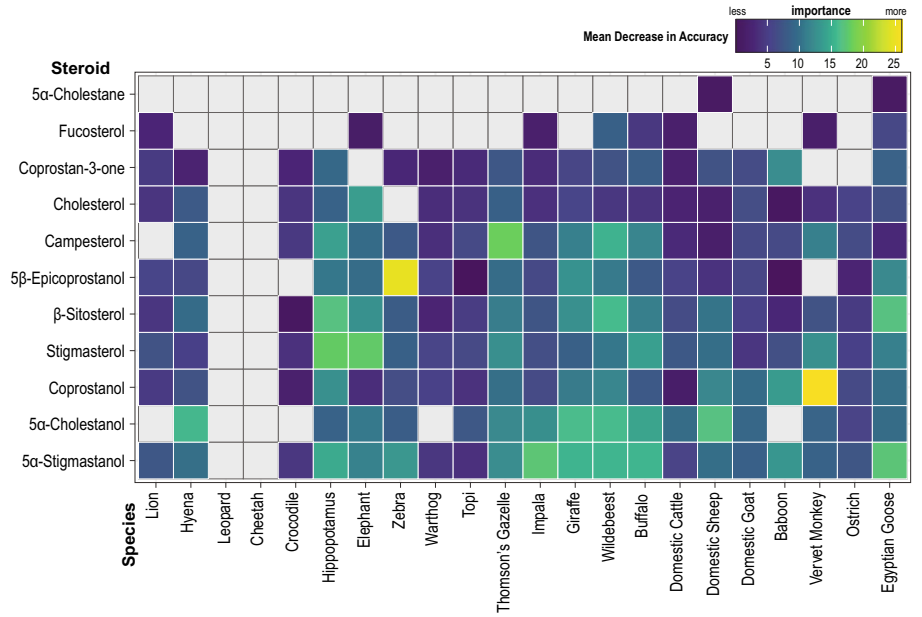

Fig. 7 Importance of fecal steroids for recognizing dung from each species as measured by the mean decrease in accuracy (cell shading) during the random forests analysis. Steroids are ordered from the greatest (bottom; most important) to smallest (top; least important) mean decrease in accuracy for the dataset as a whole. Unshaded cells have no importance value, either because the steroid was not measured in samples from that species or because a species was represented by only one individual (cheetah, leopard) and no value of importance was generated during cross validation

(four from ten) were correctly classified. Of the four correctly classified samples of carnivore dung, three were from hyenas. Carnivore dung is characterized by high abundances of cholesterol (at least 75% of fecal steroids if crocodiles are included and at least 87.5% when crocodiles are excluded), which likely leaves little scope for differentiating among species. The correct classification of hyenas despite high cholesterol content, appears to arise from the importance of 5α-cholestanol (Fig. 7), which is more abundant in dung from hyenas (4.9–7.0%; Fig. 3) than other carnivores (less than 4.0%). Elevated concentrations of 5α-cholestanol in hyena dung could be attributed to aerobic microbial action on cholesterol after the dung was excreted (Leeming et al. 1996). However, two of the hyena dung samples were collected shortly after known individuals were observed defecating. The third sample was visibly paler/drier, but estimated by experienced hyena researchers to be less than 3 days old. Despite a longer exposure to environmental conditions outside of the body, this sample yielded the least 5α-cholestanol, which suggests that the distinctiveness of hyena dung was unlikely caused by post-depositional alteration.

There is bi-directional misclassification between pairs of similar species, but at different rates. Among vervet monkeys and baboons, bi-directional misclassification was modest with one of five vervet monkey samples misclassified as baboon dung, compared to two of five in the opposite direction (Fig. 5). Instances of misclassification likely occur because dung from both species is characterized by high abundances of coprostanol. However, our analysis indicates that baboon and vervet monkey dung can be regularly distinguished from one another because of the importance of elevated coprostan-3-one for recognizing

baboon dung (Figs. 5 and 7). The single most important combination of steroid and species for improving classification of dung samples was coprostanol and vervet monkeys (mean decrease in accuracy of 25.9 when this steroid is not used in classification; Fig. 7). 5α-stigmastanol and stigmasterol are also important for differentiating baboon and vervet monkey dung, despite no clear differences in abundance. Rather there appears to be minimal variability of 5α-stigmastanol/stigmasterol measured in baboon/vervet monkey dung, which likely makes specific abundances of these steroids characteristic of the two species.

Bi-directional misclassification of buffalo/domestic cattle is slightly more common. One sample of buffalo dung was misattributed to domestic cattle, while three of five samples were misclassified in the opposite direction. This likely occurs because dung from both species is characterized by high abundances of 5α-stigmastanol and this steroid was the most important for identifying samples originating from buffalo (mean decrease in accuracy of 15.5 when it was excluded; Fig. 7). There are modest, but seemingly systematic difference in the abundance of several steroids between buffalo and domestic cattle dung (Fig. 3) that may explain why the rate of mis-classification between these species is not greater despite their biological similarities. For example, coprostanol is depleted in buffalo (mean of 7.5% and range of 5.9–9.8%) compared to domesticated cattle (mean of 10.8% and range of 7.9–14.1%).

Impala and topi were misclassified at a rate of 40% or greater with dung from these species most commonly being misattributed to domesticated goats/sheep/cattle, or one another. The second most important combination of steroid and species for improving classification of dung samples was epicoprostanol and zebra (mean decrease in accuracy of 24.6; Fig. 7), which reflects the unusually high amount of epicoprostanol in zebra dung (Fig. 3). Despite this, one of the five zebra samples was misclassified as coming from a Thomson's gazelle. Inspection of the random forests results indicates that this sample is characterized by markedly lower epicoprostanol (9.1%) than the other four zebra samples (15.4–20.6%; Fig. 3).

The five species represented by dung from only two individuals (ostrich, warthog, crocodile, leopard, and cheetah) were misclassified at a rate of 87.5% (seven of eight samples) which illustrates the challenge of predicting sample origin for underrepresented species during cross validation. However, removal of these species in a second random forests analysis increased the overall correct rate of classification by less than 1%.

Discussion

Characterization of major groups

Dung from major groups of species (carnivores, primates, ruminant herbivores, and non-ruminant herbivores) is readily distinguished using fecal steroids (Fig. 4). Carnivore dung includes very high abundances of cholesterol, primate dung includes abundant coprostanol, and two groups of (primarily) herbivores are recognized by their high abundances of 5α-stigmastanol (e.g., impala, giraffe), or β-sitosterol (e.g., elephant, hippopotamus). These patterns were recognized in other ecosystems and are attributed to well-understood combinations of diet, digestive biochemistry, and biosynthesis of endogenous steroids (Leeming et al. 1996). In the following sections, it is important to note that percentage values may not be directly comparable among studies because of differences in how samples were

prepared and analyzed and which specific steroids were (or indeed were not) measured (Bull et al. 2002).

Few studies provide fecal steroid data for carnivorous terrestrial megafauna. Leeming et al. (1996) presented measurements from cats and dogs (presumably domestic, or at least domesticated) that yielded an average cholesterol content of ~ 42% and ~ 72% respectively. Similarly, Shah et al. (2007) reported average cholesterol content of ~ 67% for dogs. These values are considerably lower than the values that we measured in carnivore dung from East Africa (Figs. 3 and 4, minimum of 75.6/87.5% including/excluding crocodiles). Although some difference in values may be attributable to the number and type of steroids measured and laboratory methods, they are likely modest given the similarity in reported values for dogs from two different studies. It is more likely that domestic(ated) cats and dogs do not have a fully carnivorous diet owing to their (presumed) consumption of processed pet food. Shah et al. (2007) measured an average cholesterol content of ~ 96% in dingos (*Canis lupus dingo*; a wild canid), while we measured 98% in dung from a captive African Wild Dog, which supports there being a robust difference between fecal steroids of wild and domesticated carnivores. Similarly, Harrault et al. (2019) showed that a dog from a remote Siberian community fed on fish and meat scraps was classified differently to domestic dogs from Scandinavia because their different diets resulted in corresponding differences in fecal steroids. In marine ecosystems, the cholesterol content of marine carnivores in Antarctica was reported as approximately 74–99% (Leeming et al. 2015; Venkatesan and Santiago 1989).

Across a range of ecosystems and species the fecal steroid content of carnivore dung is almost exclusively cholesterol. This result is attributed to a diet that is rich in cholesterol (~ 60–100 mg/100 g of muscle for farmed animals; Chizzolini et al. 1999), coupled with the absence of a metabolic mechanism (biohydrogenation by bacteria in the anaerobic conditions of the digestive tract) for converting cholesterol into 5β-stanols such as coprostanol (as occurs for example in humans), or epicoprostanol (Leeming et al. 1996). The lower abundance of cholesterol in crocodile dung compared to East African carnivorous mammals may reflect a diet that includes a high proportion of fish coupled with its unusual digestive system that facilitates consumption of large, infrequent meals through prolonged processing in a particularly acidic digestive tract (Farmer et al. 2008).

The consistency of cholesterol abundance among a wide variety of carnivores that consume correspondingly varied prey has two implications for reconstructing animal populations. Firstly, the diversity of prey consumed by modern carnivores can be interpreted as a proxy for how the prey consumed by a specific carnivore could alter through time in response to shifting food availability. This species-for-time substitution indicates that even pronounced shifts in diet would be unlikely to cause fecal steroids in carnivore dung to change dramatically. Similarity in fecal steroids from captive and wild individuals of the same species of carnivore (lion and cheetah; "Variability among individuals" section, Fig. 3) supports this proposition. Secondly, it is unlikely that carnivores can be accurately classified to the species level using fecal steroids because of their uniformly high cholesterol content, as evidenced by the high misclassification rate for carnivores (85%/60% if hyenas are excluded/included) in the random forests model (Fig. 5).

A principal goal of environmental research using fecal steroids has been recognizing human waste in modern water bodies (Leeming et al. 2015; Vane et al. 2010), or identifying human settlement/behavior in the context of archaeological investigations (Bull et al. 2001; Prost et al. 2017; Schroeter et al. 2020; Zocatelli et al. 2017). Elevated concentration of coprostanol (in isolation, or a ratio) is taken as evidence for the presence of human fecal matter (Bull et al. 2002; Liebezeit and Wöstmann 2010). Reported values for coprostanol

in fresh human dung are approximately 61–71% (Leeming et al. 1996; Shah et al. 2007; Zocatelli et al. 2017). In our dataset, similarly high concentrations of coprostanol are found in primate dung (Fig. 3). This composition of human and primate dung is attributed to an omnivorous diet, coupled with microbial activity in the anaerobic digestive tract that converts cholesterol into 5β-stanols including coprostanol (Bull et al. 2002). Due to biosynthesis, cholesterol is present and available for microbial mediation in the digestive tract of humans, even if an individual eats a low cholesterol (e.g., vegetarian) diet (Leeming et al. 1996). This is also likely the case for primates in East Africa since their diet is almost exclusively (vervet monkey) or largely (baboons) vegetarian.

The similarity of coprostanol among humans and primates eating highly variable diets indicates that diet plays a secondary role to gut biochemistry (endogenous production and subsequent conversion of cholesterol) in determining the steroid composition of dung. Therefore, changes in primate diet through time are unlikely to inhibit the characterization of dung using a modern training set. Using elevated coprostanol as a marker for pollution by human waste is underpinned by an assumption that primates are not a viable source of fecal steroids in the system being analyzed. This assumption was robust for studies in Antarctica (Leeming et al. 2015), Europe (Bull et al. 1999, 2001; Vane et al. 2010), Australia/New Zealand (Argiriadis et al. 2018; Leeming et al. 1996), and northern/central Asia (Harrault et al. 2019; Schroeter et al. 2020). In East Africa however, the presence of primates could lead to false positives for pollution by human waste. Similarly, paleoecological trends inferred from time-variable inputs of coprostanol and interpreted to be shifts in primate populations could in fact be caused by humans. Adoption of a multi-proxy approach that utilizes (for example) bile acids as well as fecal sterols may help to distinguish between anthropogenic and primate inputs (Linseele et al. 2013).

The fecal steroid composition of herbivore (plus omnivorous warthog) dung is characterized by high abundances of 5β-stanols that are produced from precursors of biosynthesized cholesterol and phytosterols ingested from plants (Harrault et al. 2019; Leeming et al. 1996; Prost et al. 2017; Shah et al. 2007). The mean composition of grass that we measured was 52.9% β-sitosterol, 31.6% stigmasterol, and 9.6% campesterol, indicating that the diet of many herbivorous megafauna in East Africa is likely rich in these phytosterols. Although some of the herbivores in our dataset do not feed primarily on grass and there is considerable species richness of grasses (Kartzinel and Pringle 2020), these phytosterols are common to most plants (e.g., Bot 2019). For example, giraffe feed primarily on leaves rather than grass, but the steroid composition of the Acacia trees that they regularly browse is approximately 46–54% β-sitosterol (Nasri et al. 2012), which is similar to our measurements from grass.

Despite the commonality of key phytosterols to most plants, our (limited) comparison of dung from captive and wild individuals of the same species indicates that diet may influence fecal steroids in herbivore dung ("Analytical and instrumental reproducibility" section). Therefore, dietary changes (e.g., seasonally during annual migrations, or on Holocene timescales due to wet/dry climate phases which effect the relative availability of grasses and woody vegetation) could result in the fecal steroids that characterize a species being variable through time. Kartzinel et al. (2019) examined plant and bacteria DNA in dung samples from Kenyan megafauna and demonstrated that diet and gut microbiomes can vary seasonally, although the degree of variability and co-variance between diet and microbiome is considerable among species. Changes in diet or gut microbiome composition could result in corresponding shifts in fecal steroids. Therefore, an expanded modern dataset that captures seasonal and geographic variability in diet may be necessary for herbivores in particular. Since our sampling took place over the course of 1 week, seasonality

in diet is one possible explanation for the classification of dung from Thomson's gazelles being different to other ruminants (Fig. 4; "Variability among species" section). Alternatively, differences between wild and captive herbivores may arise from environmental factors other than diet (e.g., climate or stress), in which case analysis of fecal steroids in fresh dung may provide a means to assess population health. This approach would be convenient since collection of dung is relatively easy, cost-effective, and does not require capture and release of animals.

The relative abundance of phytosterols in elephant and hippopotamus dung is less than in the plants they ingest (compare Figs. 3 and 8), which indicates that there is systematic alteration during digestion. Although these species consume diets that are considerably more varied than red oat grass specifically or grasses generally (Kartzinel et al. 2019), the steroid profile we measured on two grass samples is likely broadly representative of their dietary intake of phytosterols. Compared to other herbivores, elephant and hippopotami dung retains high proportions of β-sitosterol, stigmasterol, and campestrol, which indicates a lower degree of digestion. Indeed, elephant and hippopotamus dung include considerable quantities of identifiable un- and partially-digested plant material. The mechanism for this seemingly low conversion of phytosterols into 5β-stanols may differ between the two species.

Larger herbivores typically have a longer gastrointestinal tract than smaller herbivores, which results in a longer retention time and allows them to consume lower quality forage (Clauss et al. 2003). However, forage consumed by elephants has a shorter retention time in the gut (40–50 h) than is predicted for an animal with its body weight (65–80 h; Illius and Gordon 1992). This accelerated passage may explain the relatively high concentration of phytosterols relative to 5β-stanols in elephant dung. Hippopotami are grazing, non-ruminant, foregut-fermenting herbivores with a very simple hindgut (Clauss et al. 2003). In contrast to grazing ruminants, hippopotami do not stratify their gut content, which makes selective particle retention unlikely and may result in a correspondingly reduced rate of phytosterol conversion to 5β-stanols since partially-digested plant material is passed more often than in ruminants. In contrast to elephants and hippopotami, digestion by ruminants

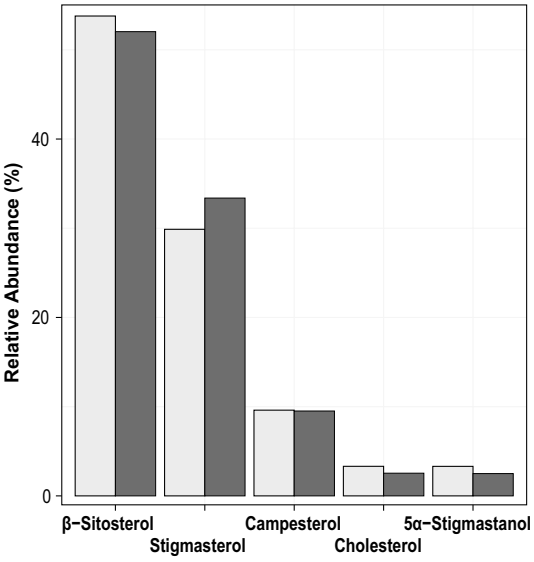

Fig. 8 Steroid composition of red oat grass measured on two samples (colored bars) collected in the Maasai Mara National Reserve

proceeds further in the conversion of phytosterols to 5α-stigmastanol (Fig. 3) because consumed plants are fermented through microbial actions and undergo repeated digestion by rechewing of cud. Furthermore, Kartzinel et al. (2019) demonstrated that ruminant and non-ruminants in East Africa had gut microbiomes with little overlap in composition, which could also influence differences in fecal steroids between these groups.

Zebra dung is unusual because it contains elevated amounts of epicoprostanol (Fig. 3), which is produced by biohydrogenation of cholesterol by anaerobic bacteria (Leeming et al. 1996). Zebras are hindgut fermenters rather than ruminants and compensate for the lower degree of digestion by consuming more and a wider variety of food and processing it quickly (Grubb 1981). Analysis of dung collected in Kenya by Kartzinel and Pringle (2020) demonstrated that zebra almost exclusively consume grass, but within this lineage many species are ingested. Donkeys and buffalo consume a similarly diverse, but exclusively grass-based diet (Kartzinel and Pringle 2020). Despite this similarity, zebra and buffalo dung are readily distinguished using fecal steroids, and horses (likely similar to donkeys) do not produce epicoprostanol to the same extent as zebras. Shah et al. (2007) reported an average epicoprostanol content of 7% for horses, compared to 8–11% in the study of Harrault et al. (2019) and ~ 1% in (Leeming et al. 1996; Prost et al. 2017). Therefore, the unique fecal steroid profile of zebra's is probably attributable to processing of food in the digestive tract rather than an unusual diet.

Dung from two species of bird (ostrich and Egyptian goose) yielded markedly different fecal steroid profiles. Relative to one another, ostrich dung is rich in coprostanol and 5α-stigmastanol, while Egyptian goose dung is rich in β-sitosterol (Fig. 3). Reported fecal steroid measurements from bird dung broadly recognize species that produce dung rich in β-sitosterol (e.g., ducks, swans), cholesterol (e.g., sea gulls, penguins), or a combination of both (e.g., chickens, turkeys; Devane et al. 2015; Leeming et al. 1996; Shah et al. 2007; Venkatesan and Santiago 1989). These divisions may represent a spectrum of diets from almost exclusively herbivorous to omnivorous. However, the presence of coprostanol in ostrich dung suggests that cholesterol provided by its omnivorous diet undergoes conversion to coprostanol during digestion, in a similar fashion to primates and this result appears unusual compared to other birds.

Species-level recognition of dung

The primary application of sediment-hosted fecal steroids is to distinguish dung inputs that are dominated by one group of species or another (often humans vs herbivores). In many cases this is achieved through examining ratios of steroids (or groups of steroids) and applying a threshold value for binary classification (Bull et al. 1999; Prost et al. 2017; Schroeter et al. 2020). This approach is effective for recognizing the impact of humans as recent polluters (Vane et al. 2010) and as agents of environmental change/presence in the archaeological and paleoenvironmental record (Argiriadis et al. 2018; Bull et al. 2001, 1999), particularly in locations lacking primates ("Characterization of major groups" section). Similarly, other (non-steroid) proxies have demonstrated utility in tracking the abundance of some groups of megafauna through time, most notably the concentration of *Sporormiella* in sediment cores is used to infer the density of herbivores through time (Burney et al. 2003; Davis and Shafer 2006).

Efforts to identify the origin of dung to the species level are less common because study goals are adequately met by using ratios, the limited scope of available data (number and scope of samples) to provide a modern training set or reference library, or the number of

different fecal steroids that were measured. Directly combining results from multiple studies to create an expanded reference library remains challenging because of differences in sample preparation and measurement (Bull et al. 2002). However, some studies demonstrated the possibility that fecal steroids can recognize the origin of dung to the species level. For example, Harrault et al. (2019) showed that passive projection of samples into principal components analysis and hierarchical clustering could accurately establish the species origin of samples. They proposed that this success stemmed from the statistical technique used, coupled with increasing the number of measured steroids (11 fecal steroids), a relatively large training set (90 samples representing ten species, although some were removed where a priori evidence indicated the absence of species), and limiting analysis to 5β-stanols to negate the influence of steroids with non-fecal origins.

Analysis using random forests correctly classified the origin of a dung sample to the species level in ~ 72% of cases (62 out of 87 individuals) under cross validation. We used random forests due to their consistently excellent performance in accurately classifying samples in small and large data sets. In particular, random forests is superior to alternative statistical models because it allows non-linear and interacting effects among variables (Cutler et al. 2007, 2012). If carnivore dung (other than hyenas) is excluded from these results the successful classification rate is 76% and exclusion of species represented by two or fewer individuals in the current dataset also improved model performance slightly (to ~ 73%). This suggests two pathways to further refine future efforts. Firstly, it may be necessary to accept that some groups of species (e.g. carnivores, primates/humans, buffalo/domestic cattle) cannot be adequately distinguished using fecal steroids and that these taxa should grouped prior to analysis. Secondly, expansion of the reference library to include additional replicates from new individuals may improve model performance. In some cases (e.g., warthog and ostrich) this would be a trivial task, but it is more challenging for others whose population densities are lower, produce dung infrequently (crocodiles), or are rarely sighted (leopards). Our current reference library does not include all of the species that are encountered regularly in the Maasai Mara National Reserve. Future work should also expand the number of species as well as individuals of each species. This expansion may improve or decrease model performance.

Our results indicate that some steroids are more useful than others for establishing the species origin of dung samples (Figs. 6 and 7) and these results could be used to focus analytical efforts on analyzing more samples or different steroids. However, a known challenge with interpreting the importance of individual steroids in random forests is the influence of correlations with other steroids. If two steroids are strongly correlated in the dataset, it is possible that the importance of those steroids is underestimated during cross validation. For example, in our dataset there are strong, positive correlations between stigmasterol, β-sitosterol, and campesterol since these phytosterols comprise the principal dietary inputs for herbivores (Fig. 9). During cross-validation of the random forests model, exclusion of one phytosterol may have a muted effect since other, strongly-correlated steroids remain in the training set. One potential solution is that of variable selection approaches through which only the most influential variables are retained in the final model (Deng and Runger 2013; Wundervald et al. 2020).

Continued proxy development

Our goal is to develop a sediment-hosted proxy for the distribution and composition of megafaunal populations in East Africa during the Holocene. Proxy development is a

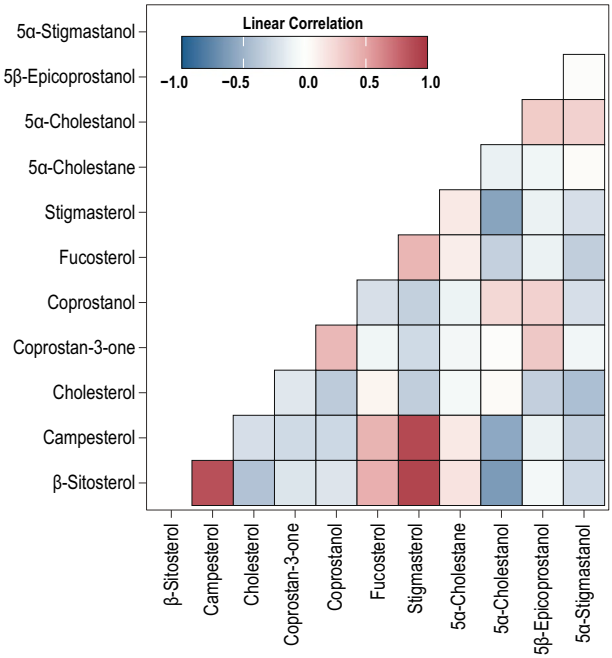

Fig. 9 Linear correlation of fecal steroid pairs measured in the dung of wild animals from the Maasai Mara National Reserve. Cell shading indicates the direction and strength of correlation

multi-stage process that begins with evaluating if fecal steroids (quantified by the relative abundance of specific steroid types) can be used to objectively and accurately recognize dung samples of known origin. Cross validation performed by the random forests model indicates that we could successfully identify the origin of a dung samples at a rate of ~ 72%. Although we consider our approach to have passed the first step of proxy development there are several challenges to consider before sediment-hosted fecal steroids can be reliably employed to reconstruct megafauna populations.

Since dung is almost certain to experience some degree of sub-aerial exposure in terrestrial environments between the time of deposition and burial, it is possible that a modern training set of fresh (pre-alteration) dung would not be appropriate for recognizing dung sources in the sedimentary record. For example, under aerobic conditions cholesterol can be converted to 5α-cholestanol (Leeming et al. 2015), although our dataset indicates that this alteration was not detectable in crocodile or hyena dung on timescales of days to a few weeks. Research is needed to understand and quantify post-depositional modification of fecal sterols across a range of species. Such work could be undertaken in controlled laboratory settings, where sub samples of dung are exposed to tightly controlled environmental conditions. Although most species in the Massai Mara ecosystem and our dataset are terrestrial, it is also important to evaluate possible modification of dung from (semi-)aquatic species such as hippopotami (Subalusky et al. 2018), where alteration may occur in water bodies that span a range redox states (Dutton et al. 2020).

Fecal steroids preserved in sedimentary environments are from mixed sources since depositional settings such as lakes capture sediment and fecal steroids from a catchment area that contains many individual animals and species. Lake sediment may also include inputs of non-fecal steroids through the flux of allochthonous plant material and autochthonous contributions from aquatic plants and algae. For example, our results indicate that fucosterol is a minor component of fecal steroids, but lake sediment may be relatively

enriched in fucosterol because of its elevated concentration in freshwater algae (Gallo et al. 2020; Martin-Creuzburg and Merkel 2016). Measurement of sediment-hosted fecal steroids will inherently reflect the net outcome of this source mixing. Fortunately, similar challenges are encountered when using other proxies such as isotopes in bulk sediment and efforts to produce probabilistic models to disentangle mixed inputs have proven successful and provide a framework for evaluating mixed fecal steroids (Parnell et al. 2013; Phillips et al. 2014). Existing studies demonstrate that fecal steroids are preserved in lake sediment across a wide range of climates and geomorphologies (D'Anjou et al. 2012; Schroeter et al. 2020; Vane et al. 2010). Although reporting conventions for the concentration of fecal steroids in paleo lake sediment vary among studies, values of up to ~ 5000 µg/g are reported in the literature (Schroeter et al. 2020). Concentrations are also likely to vary across space and through time in response to sedimentation rate, climate, and animal density for example.

Fecal steroids preserved in sediment may also reflect several ecological behaviors and physiological traits that should be evaluated. There are considerable differences among species in the amount of dung that is produced which reflects the size of individuals, their frequency of defecation, and population size. Some species display behavior during defecation that may cause their dung to be more/less likely to enter depositional environments. For example, some species deposit dung repeatedly in the same prominent location to mark territory, or to communicate with other individuals through use of communal latrines (e.g., rhinoceros). Other species may conceal their dung (e.g., some felids), while others defecate across a wide area (e.g., grazing herbivores as large herds move), may actively spread dung, or engage actively (e.g., hippopotamus young) or passively in coprophagy. An important next step in the development of sediment-hosted fecal steroids as a proxy for megafauna populations is to measure the abundance of fecal steroids in modern (i.e., surface) sediment from depositional environments that would be targeted for obtaining material for Holocene proxy reconstructions (e.g., lakes and pools). These time and catchment integrated samples will provide insight into the effect of mixing and post-depositional alteration of steroids under field conditions if recovered from locations with supporting datasets of species abundance over recent years to decades.

Conclusions

Conservation paleobiology seeks to leverage paleoenvironmental reconstructions to better anticipate future changes. During the Holocene, East African megafauna populations likely responded to trends and events in regional hydroclimate as they are anticipated to do in the future. However, methods for reconstructing megafauna populations are currently lacking. Sediment-hosted fecal steroids are a potential proxy for megafauna populations. As a first step in proxy development we collected fresh dung from 87 individuals that represent 22 species of wild and domesticated animals in and around the Maasai Mara National Reserve, Kenya. Repeated sub-sampling, laboratory extractions, and instrumental measurements demonstrate that fecal steroids can be reliably quantified. We show that fecal steroid profiles in carnivore dung are dominated by cholesterol, while those from primates are (like humans) characterized by high abundances of coprostanol. Herbivores are readily divided into two groups representing (to the first order) ruminants that more effectively modify phytosterols consumed by eating plants into 5β-sterols, and non-ruminants (particularly elephant and hippopotamus) whose dung contains relatively more/less phytosterols/5β-steroids because of less effective digestion.

To evaluate the possibility that fecal steroids could be used to identify the origin of a dung sample to the species level, we applied a random forests model to the modern training set. Under cross validation, 72% of samples were correctly classified. Misclassification was pronounced for carnivores (except hyena) because all species of carnivore produce dung that is almost exclusively comprised cholesterol. Species represented by two or fewer individuals were frequently misclassified and there was bi-directional misclassification among primates (vervet monkey and baboon) and buffalo/domestic cattle. Our analysis indicates that fecal steroids are a promising proxy for reconstructing megafauna populations in East Africa. Continued proxy development requires multiple steps including (1) expanding the number of species and individuals in the modern training set; (2) undertaking multi-season sampling of fresh dung to capture variability introduced by seasonal diets; (3) quantifying post-depositional modification of fecal sterols under representative conditions; (4) measuring fecal sterols in modern (surface) lake sediment in catchments with well constrained populations to provide analogs for older sediment that likely includes mixed sources of fecal (and non-fecal) steroids.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10531-021-02328-y.

Acknowledgements This worked was funded by National Geographic Society (Waitt Grants program) award W419–20 and the U.S. National Science Foundation (awards DEB–1354053, 1354062, and 1753727). Andrew Parnell's work was supported by: a Science Foundation Ireland Career Development Award (17/CDA/4695); an investigator award (16/IA/4520); a Marine Research Programme funded by the Irish Government, co-financed by the European Regional Development Fund (Grant-Aid Agreement No. PBA/CC/18/01); European Union's Horizon 2020 research and innovation programme under grant agreement No 818144; and SFI Research Centre awards 16/RC/3872 and 12/RC/2289_P2. Dung samples were collected in the field in affiliation with the National Museums of Kenya under research permit number NCST/RCD/12B/012/36 from the Kenya National Commission for Science, Technology and Innovation. We thank Chris Smart (Woburn Safari Park) and Noel Carey (West Midlands Safari Park) for providing dung samples from captive animals with support from the British and Irish Zoos Association (Amy Plowman, August 2015). We are grateful to Brian Heath and the Mara Conservancy for hosting us in the field and providing rangers to help with sample identification and collection. We thank Paul Geemi for assistance with field work. Detailed and insightful comments from two reviewers improved this manuscript during the review process.

References

Argiriadis E, Battistel D, McWethy DB, Vecchiato M, Kirchgeorg T, Kehrwald NM, Whitlock C, Wilmshurst JM, Barbante C (2018) Lake sediment fecal and biomass burning biomarkers provide direct evidence for prehistoric human-lit fires in New Zealand. Sci Rep 8(1):12113

Bartlett PD (1987) Degradation of coprostanol in an experimental system. Mar Pollut Bull 18(1):27–29

Bartlett LJ, Williams DR, Prescott GW, Balmford A, Green RE, Eriksson A, Valdes PJ, Singarayer JS, Manica A (2016) Robustness despite uncertainty: regional climate data reveal the dominant role of humans in explaining global extinctions of Late Quaternary megafauna. Ecography 39(2):152–161

Bartzke GS, Ogutu JO, Mukhopadhyay S, Mtui D, Dublin HT, Piepho H-P (2018) Rainfall trends and variation in the Maasai Mara ecosystem and their implications for animal population and biodiversity dynamics. PLoS ONE 13(9):e0202814

Bot A (2019) Phytosterols. In: Melton L, Shahidi F, Varelis P (eds) Encyclopedia of food chemistry. Academic Press, Oxford, pp 225–228

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Bull ID, Simpson IA, Dockrill SJ, Evershed RP (1999) Organic geochemical evidence for the origin of ancient anthropogenic soil deposits at Tofts Ness, Sanday, Orkney. Organic Geochem 30(7):535–556

Bull ID, Evershed RP, Betancourt PP (2001) An organic geochemical investigation of the practice of manuring at a Minoan site on Pseira Island. Crete Geoarchaeol 16(2):223–242

Bull ID, Lockheart M, Elhmmali M, Roberts D, Evershed R (2002) The origin of faeces by means of biomarker detection. Environ Int 27:647–654
Burney DA, Robinson GS, Burney LP (2003) *Sporormiella* and the late Holocene extinctions in Madagascar. Proc Natl Acad Sci 100(19):10800–10805
Camberlin P (2018) Climate of Eastern Africa. In: Oxford research encyclopedia of climate science. Oxford University Press, Oxford
Chizzolini R, Zanardi E, Dorigoni V, Ghidini S (1999) Calorific value and cholesterol content of normal and low-fat meat and meat products. Trends Food Sci Technol 10(4):119–128
Clauss M, Frey R, Kiefer B, Lechner-Doll M, Loehlein W, Polster C, Rössner GE, Streich WJ (2003) The maximum attainable body size of herbivorous mammals: morphophysiological constraints on foregut, and adaptations of hindgut fermenters. Oecologia 136(1):14–27
Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. Ecology 88(11):2783–2792
Cutler A, Cutler DR, Stevens JR (2012) Random forests. In: Zhang C, Ma Y (eds) Ensemble machine learning. Springer, Berlin, pp 157–175
D'Anjou RM, Bradley RS, Balascio NL, Finkelstein DB (2012) Climate impacts on human settlement and agricultural activities in northern Norway revealed through sediment biogeochemistry. Proc Natl Acad Sci 109(50):20332–20337
Davis OK, Shafer DS (2006) *Sporormiella* fungal spores, a palynological means of detecting herbivore density. Palaeogeogr Palaeoclimatol Palaeoecol 237(1):40–50
De Cort G, Bessems I, Keppens E, Mees F, Cumming B, Verschuren D (2013) Late-Holocene and recent hydroclimatic variability in the central Kenya Rift Valley: the sediment record of hypersaline lakes Bogoria, Nakuru and Elementeita. Palaeogeogr Palaeoclimatol Palaeoecol 388:69–80
Deng H, Runger G (2013) Gene selection with guided regularized random forest. Pattern Recogn 46(12):3483–3489
Devane ML, Wood D, Chappell A, Robson B, Webster-Brown J, Gilpin BJ (2015) Identifying avian sources of faecal contamination using sterol analysis. Environ Monit Assess 187(10):625
Dietl GP, Flessa KW (2011) Conservation paleobiology: putting the dead to work. Trends Ecol Evol 26(1):30–37
Dietl GP, Kidwell SM, Brenner M, Burney DA, Flessa KW, Jackson ST, Koch PL (2015) Conservation paleobiology: leveraging knowledge of the past to inform conservation and restoration. Annu Rev Earth Planet Sci 43:79–103
Dutton CL, Subalusky AL, Hamilton SK, Bayer EC, Njoroge L, Rosi EJ, Post DM (2020) Alternative biogeochemical states of river pools mediated by hippo use and flow variability. Ecosystems 24:284–300
Farmer CG, Uriona TJ, Olsen DB, Steenblik M, Sanders K (2008) The right-to-left shunt of Crocodilians serves digestion. Physiol Biochem Zool 81(2):125–137
Funk C, Hoell A, Shukla S, Husak G, Michaelsen J (2016) The East African monsoon system: seasonal climatologies and recent variations. In: de Carvalho LMV, Jones C (eds) The monsoons and climate change: observations and modeling. Springer, Cham, pp 163–185
Gallo C, Landi S, d'Ippolito G, Nuzzo G, Manzo E, Sardo A, Fontana A (2020) Diatoms synthesize sterols by inclusion of animal and fungal genes in the plant pathway. Sci Rep 10(1):4204
Gaskell SJ, Eglinton G (1975) Rapid hydrogenation of sterols in a contemporary lacustrine sediment. Nature 254(5497):209–211
Gebrechorkos SH, Hülsmann S, Bernhofer C (2019) Regional climate projections for impact assessment studies in East Africa. Environ Res Lett 14(4):044031
Grubb P (1981) Equus burchelli. Mamm Species 157:1–9
Harrault L, Milek K, Jardé E, Jeanneau L, Derrien M, Anderson D (2019) Faecal biomarkers can distinguish specific mammalian species in modern and past environments. PLoS ONE 14(2):e0211119
Holdo RM, Holt RD, Fryxell JM (2009) Opposing rainfall and plant nutritional gradients best explain the wildebeest migration in the Serengeti. Am Nat 173(4):431–445
Humphries MS, Kirsten KL, McCarthy TS (2019) Rapid changes in the hydroclimate of southeast Africa during the mid- to late-Holocene. Quatern Sci Rev 212:178–186
Illius A, Gordon I (1992) Modelling the nutritional ecology of ungulate herbivores: evolution of body size and competitive interactions. Oecologia 89(3):428–434
Kartzinel TR, Pringle RM (2020) Multiple dimensions of dietary diversity in large mammalian herbivores. J Anim Ecol 89(6):1482–1496
Kartzinel TR, Hsing JC, Musili PM, Brown BRP, Pringle RM (2019) Covariation of diet and gut microbiome in African megafauna. Proc Natl Acad Sci 116(47):23588–23593

Kaufman L, Rousseeuw PJ (2005) Finding groups in data: an introduction to cluster analysis. Wiley-Interscience, Hoboken, NJ, USA
Leeming R, Ball A, Ashbolt N, Nichols P (1996) Using faecal sterols from humans and animals to distinguish faecal pollution in receiving waters. Water Res 30(12):2893–2900
Leeming R, Stark JS, Smith JJ (2015) Novel use of faecal sterols to assess human faecal contamination in Antarctica: a likelihood assessment matrix for environmental monitoring. Antarct Sci 27(1):31–43
Liaw A, Wiener M (2018) randomForest. In.
Liebezeit G, Wöstmann R (2010) Coprostanol in Siak River sediments, E Sumatra. Indonesia Bull Environ Contam Toxicol 85(6):585–588
Linseele V, Riemer H, Baeten J, De Vos D, Marinova E, Ottoni C (2013) Species identification of archaeological dung remains: A critical review of potential methods. Environ Archaeol 18(1):5–17
Lyon B, DeWitt DG (2012) A recent and abrupt decline in the East African long rains. Geophysical Research Letters 39(2):
Malhi Y, Doughty CE, Galetti M, Smith FA, Svenning J-C, Terborgh JW (2016) Megafauna and ecosystem function from the Pleistocene to the Anthropocene. Proc Natl Acad Sci 113(4):838–846
Martin-Creuzburg D, Merkel P (2016) Sterols of freshwater microalgae: potential implications for zooplankton nutrition. J Plankton Res 38(4):865–877
Nasri N, Elfalleh W, Tlili N, Hannachi H, Triki S, Khaldi A (2012) Minor lipid components of some Acacia species: potential dietary health benefits of the unexploited seeds. Lipids Health Dis 11:49–49
Ogutu JO, Piepho HP, Dublin HT, Bhola N, Reid RS (2008) Rainfall influences on ungulate population abundance in the Mara-Serengeti ecosystem. J Anim Ecol 77(4):814–829
Ogutu JO, Owen-Smith N, Piepho HP, Said MY (2011) Continuing wildlife population declines and range contraction in the Mara region of Kenya during 1977–2009. J Zool 285(2):99–109
Parnell AC, Phillips DL, Bearhop S, Semmens BX, Ward EJ, Moore JW, Jackson AL, Grey J, Kelly DJ, Inger R (2013) Bayesian stable isotope mixing models. Environmetrics 24(6):387–399
Phillips DL, Inger R, Bearhop S, Jackson AL, Moore JW, Parnell AC, Semmens BX, Ward EJ (2014) Best practices for use of stable isotope mixing models in food-web studies. Can J Zool 92(10):823–835
Prost K, Birk JJ, Lehndorff E, Gerlach R, Amelung W (2017) Steroid biomarkers revisited - improved source identification of faecal remains in archaeological soil material. PLoS ONE 12(1):e0164882–e0164882
Sandom C, Faurby S, Sandel B, Svenning J-C (2014) Global late Quaternary megafauna extinctions linked to humans, not climate change. Proceedings of the Royal Society b: Biological Sciences 281(1787):20133254
Schroeter N, Lauterbach S, Stebich M, Kalanke J, Mingram J, Yildiz C, Schouten S, Gleixner G (2020) Biomolecular evidence of early human occupation of a high-altitude site in Western Central Asia during the Holocene. Front Earth Sci 8:20
Shah VG, Hugh Dunstan R, Geary PM, Coombes P, Roberts TK, Von Nagy-Felsobuki E (2007) Evaluating potential applications of faecal sterols in distinguishing sources of faecal contamination from mixed faecal samples. Water Res 41(16):3691–3700
Stuart C, Stuart M (2013) Stuart's Field Guide to the Tracks anad Signs of Souther, Central and East African Wildlife. Penguin Random House South Africa,
Subalusky AL, Dutton CL, Njoroge L, Rosi EJ, Post DM (2018) Organic matter and nutrient inputs from large wildlife influence ecosystem function in the Mara River. Africa Ecology 99(11):2558–2574
Tierney JE, deMenocal PB (2013) Abrupt shifts in Horn of Africa hydroclimate since the Last Glacial Maximum. Science 342(6160):843–846
Tierney JE, Smerdon JE, Anchukaitis KJ, Seager R (2013) Multidecadal variability in East African hydroclimate controlled by the Indian Ocean. Nature 493(7432):389–392
Tierney JE, Ummenhofer CC, deMenocal PB (2015) Past and future rainfall in the Horn of Africa. Science Advances 1(9): e1500682
Tyagi P, Edwards DR, Coyne MS (2007) Use of selected chemical markers in combination with a multiple regression model to assess the contribution of domesticated animal sources of fecal pollution in the environment. Chemosphere 69(10):1617–1624
Vane CH, Kim A, McGowan S, Leng M, Heaton T, Kendrick C, Coombs P, Yang H, Swann G (2010) Sedimentary records of sewage pollution using faecal markers in contrasting peri-urban shallow lakes. Sci Total Environ 409(2):345–356
Venkatesan MI, Santiago CA (1989) Sterols in ocean sediments: novel tracers to examine habitats of cetaceans, pinnipeds, penguins and humans. Mar Biol 102(4):431–437

Wundervald B, Parnell AC, Domijan K (2020) Generalizing gain penalization for feature selection in tree-based models. IEEE Access 8:190231–190239

Yang W, Seager R, Cane MA, Lyon B (2015) The Annual Cycle of East African Precipitation. J Clim 28(6):2385–2404

Zocatelli R, Lavrieux M, Guillemot T, Chassiot L, Le Milbeau C, Jacob J (2017) Fecal biomarker imprints as indicators of past human land uses: Source distinction and preservation potential in archaeological and natural archives. J Archaeol Sci 81:79–89

Authors and Affiliations

Andrew C. Kemp[1] · Christopher H. Vane[2] · Alexander W. Kim[2] · Christopher L. Dutton[3,4] · Amanda L. Subalusky[4] · Stuart K. Kemp[5] · Andrew C. Parnell[6]

[1] Department of Earth and Ocean Sciences, Tufts University, Medford, MA 02176, USA

[2] British Geological Survey, Environmental Science Centre, Keyworth, Nottingham NG12 5GG, UK

[3] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA

[4] Department of Biology, University of Florida, Gainesville, FL 32611, USA

[5] Reducer, London EC2Y 5JA, UK

[6] Hamilton Institute, Insight Centre for Data Analytics, Maynooth University, Maynooth, Co. Kildare, Ireland