General and Comparative Endocrinology 244 (2017) 146–156

Contents lists available at ScienceDirect

General and Comparative Endocrinology

journal homepage: www.elsevier.com/locate/ygcen

One size does not fit all: Monitoring faecal glucocorticoid metabolites in marsupials

Kerry V. Fanson [a,b,*], Emily C. Best [c], Ashley Bunce [c,d], Benjamin G. Fanson [b], Lindsay A. Hogan [e,f], Tamara Keeley [e], Edward J. Narayan [g], Rupert Palme [h], Marissa L. Parrott [i], Trudy M. Sharp [j], Kim Skogvold [k,l], Lisa Tuthill [m], Koa N. Webster [n], Meredith Bashaw [a,o]

[a] Wildlife Reproductive Centre, Taronga Conservation Society Australia, Dubbo, NSW, Australia
[b] Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Geelong, VIC, Australia
[c] School of Biological Sciences, The University of Queensland, Brisbane, QLD, Australia
[d] Department of Environment and Heritage Protection, Brisbane, QLD, Australia
[e] Wildlife Biology Unit, School of Agriculture and Food Sciences, The University of Queensland, Gatton, QLD, Australia
[f] Native Species Breeding Program (NSBP), Perth Zoo, South Perth, WA, Australia
[g] Graham Centre for Agricultural Innovation & School of Animal and Veterinary Sciences, Charles Sturt University, Wagga Wagga, NSW, Australia
[h] Department of Biomedical Sciences, University of Veterinary Medicine, Vienna, Austria
[i] Wildlife Conservation and Science, Zoos Victoria, Parkville, VIC, Australia
[j] Fowlers Gap Arid Zone Research Station, Centre of Ecosystem Science, School of Biological Earth and Environmental Sciences, University of New South Wales, Kensington, NSW, Australia
[k] Conservation Medicine Program, School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA, Australia
[l] Perth Zoo Veterinary Department, Perth Zoo, South Perth, WA, Australia
[m] Moonlit Sanctuary, Pearcedale, VIC, Australia
[n] Department of Biological Sciences, Macquarie University, Sydney, NSW, Australia
[o] Department of Psychology, Franklin and Marshall College, Lancaster, PA, USA

ARTICLE INFO

Article history:
Received 17 May 2015
Revised 7 October 2015
Accepted 14 October 2015
Available online 23 October 2015

Keywords:
Adrenal
Cortisol
Corticosterone
Enzyme immunoassay
Non-invasive
Validation

ABSTRACT

Marsupial research, conservation, and management can benefit greatly from knowledge about glucocorticoid (GC) secretion patterns because GCs influence numerous aspects of physiology and play a crucial role in regulating an animal's response to stressors. Faecal glucocorticoid metabolites (FGM) offer a non-invasive tool for tracking changes in GCs over time. To date, there are relatively few validated assays for marsupials compared with other taxa, and those that have been published generally test only one assay. However, different assays can yield very different signals of adrenal activity. The goal of this study was to compare the performance of five different enzyme immunoassays (EIAs) for monitoring adrenocortical activity via FGM in 13 marsupial species. We monitored FGM response to two types of events: biological stressors (e.g., transport, novel environment) and pharmacological stimulation (ACTH injection). For each individual animal and assay, FGM peaks were identified using the iterative baseline approach. Performance of the EIAs for each species was evaluated by determining (1) the percent of individuals with a detectable peak 0.125–4.5 days post-event, and (2) the biological sensitivity of the assay as measured by strength of the post-event response relative to baseline variability (*Z*-score). Assays were defined as successful if they detected a peak in at least 50% of the individuals and the mean species response had a $Z \geqslant 2$. By this criterion, at least one assay was successful in 10 of the 13 species, but the best-performing assay varied among species, even those species that were closely related. Furthermore, the ability to confidently assess assay performance was influenced by the experimental protocols used. We discuss the implications of our findings for biological validation studies.

* Corresponding author at: Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, 75 Pigdons Road, Waurn Ponds, VIC 3216, Australia.
E-mail address: kerry.fanson@deakin.edu.au (K.V. Fanson).

http://dx.doi.org/10.1016/j.ygcen.2015.10.011
0016-6480/

1. Introduction

Marsupials exhibit several unique physiological, reproductive, and ecological traits compared with other mammals, making them important for understanding basic endocrine function. Indeed, the potential influence of hormones on individual fitness is well illustrated in dasyurids, which provide a particularly unique example of endocrine-mediated life history strategy – the complete male die-off at the end of the breeding season. This die-off is a result of a failure in the negative feedback system of the hypothalamic–pituitary–adrenal axis resulting in extended, elevated GC production which becomes deleterious to the health of the animal (Bradley, 2003; Naylor et al., 2008). However, aside from a few well-studied examples, there is still relatively little known about basic endocrinology and particularly adrenal function for many marsupial species (Hing et al., 2014; McDonald, 1977). Furthermore, more than 30% of Australian mammals are classified as threatened (Woinarski et al., 2014), and nearly 50% of recent mammalian extinctions have occurred in Australia (Short and Smith, 1994). Conservation efforts are key to species survival, and relocation to predator-free sanctuaries and captive breeding programs have become important management practices to help sustain dwindling marsupial populations (Selwood and Cui, 2006; Sheean et al., 2012).

Glucocorticoids (GCs) are secreted by the adrenal glands and regulate several aspects of physiology, including stress physiology. Because GCs play a broad role in energy regulation and homeostasis (Sapolsky et al., 2000), their characterisation as only a "stress" hormone is a misnomer. At baseline levels, GCs help regulate circadian and circannual rhythms (Dallmann et al., 2006), promote healthy reproductive function (Whirledge and Cidlowski, 2013), and facilitate immune function (Padgett and Glaser, 2003). These diverse roles are illustrated by the dramatic physiological effects of chronic GC elevation, including reproductive suppression, impaired immune function, muscle wasting, and decreased cognitive function (McEwen, 1998; Sapolsky, 2002). Individual differences in GC have also been associated with variation in behaviour (Carere et al., 2010; Koolhaas et al., 2010), and may underlie individual differences in fitness or life history strategies. Therefore, monitoring patterns of adrenocortical activity in marsupials could facilitate the evaluation of conservation efforts on health, welfare, and reproductive success.

Until recently, blood samples were needed to monitor GC patterns, but advances in methods for monitoring faecal glucocorticoid metabolites (FGM) have made it possible to measure GC repeatedly in the same free-living individuals without the need for capture or restraint (Sheriff et al., 2011). Since changes in hormone concentrations (not just absolute values) are a key determinant of endocrine function, FGM monitoring allows us to develop an understanding of critical elements of endocrinology that are difficult to monitor using traditional techniques. In addition, FGM provide a pooled estimate of circulating hormone concentrations (Möstl and Palme, 2002; Sheriff et al., 2011; Wielebnowski and Watters, 2007), thereby reducing the pulsatile patterns present in serum and offering a more integrated measure of adrenal function. These features arguably make FGM more useful for answering large-scale questions about conservation biology, individual variation, and animal welfare (Palme, 2012; Sheriff et al., 2011).

While FGM monitoring has great potential, it also presents several methodological challenges deriving from the fact that circulating GCs are metabolized into a diverse array of molecules by liver and gut bacteria prior to excretion (Palme et al., 2005; Touma and Palme, 2005; Wielebnowski and Watters, 2007). Steroid metabolism can differ as a function of species, sex, diet, age, and season (Goymann, 2012), so one must first ensure that the FGM detected by a specific assay reflect biologically relevant changes in adrenocortical activity (Touma and Palme, 2005). Typically, this biological validation is achieved by stimulating adrenocortical GC production pharmacologically (ACTH injection) or biologically (environmental stressor), and monitoring changes in FGM following this event. Assay selection is an important but often undervalued aspect of biological validation (Palme, 2005). Assay performance varies based on the ligand against which the antibody was raised, the part of the ligand recognised by the antibody, the range of FGM to which the antibody binds, cross-reactivity with metabolites of other (particularly gonadal) hormones, and the match between the antibody binding and a species' FGM profile (Möstl et al., 2005; Palme et al., 2005). In recent years, an increasing number of studies have sought to biologically validate an assay for monitoring FGM in a specific marsupial species (Davies et al., 2013; Hogan et al., 2011, 2012; Johnston et al., 2013; Keeley et al., 2012; McKenzie and Deane, 2005; Narayan et al., 2014, 2013; Oates et al., 2007), but none have directly compared the performance of multiple assays across species.

In this study, we aimed to compare the performance of five enzyme immunoassays (EIAs) for monitoring adrenocortical activity in 13 species representing three orders of Australian marsupials. We first identified overall similarities and differences in the performance of the five assays, controlling for species and individual differences. We then assessed EIA performance within each species by (1) identifying the percent of individuals who had a detectable post-event peak in each assay, (2) quantifying the strength of the post-event response (signal) relative to the amount of variation (noise) in the baseline, and (3) evaluating timing of peaks relative to the event. Finally, we discuss which assays were most effective for each species and consider the pattern of assay performance across taxonomic categories. Our data facilitate discussion of key aspects of study design that influence the probability of detecting a peak and we outline strategies to follow when biologically validating FGM assays.

2. Materials and methods

2.1. Study species and stressors

This study included a total of 44 individuals from 13 marsupial species representing three Marsupialia Orders (see Table 1 for details, including the number of males and females of each species). All animals were permanently housed in captivity except for western grey kangaroos (wild-caught and temporarily held in holding pens) and the northern hairy-nosed wombat (undergoing rehabilitation). All animals were adults except the northern hairy-nosed wombat, which was a juvenile. All procedures were approved by the appropriate ethics committees and review boards.

For seven species, an ACTH challenge was used to pharmacologically stimulate the adrenal cortex (see Table 1 for manufacturer and dose). ACTH was administered as a single intramuscular injection of Synacthen (25 IU/ml) or Synacthen Depot (100 IU/ml; Novartis, Basel, Switzerland) or Corticotropin (80 IU/ml; Wedgewood Pharmacy, New Jersey, USA). This injection was administered under anaesthesia to two species (Tasmanian devils and koalas) and without anaesthesia to five species (numbats, bilbies, woylies, eastern grey kangaroo, and western grey kangaroo).

For the remaining six species, pre-scheduled biological stressors were used (Table 1). Three species were transferred from one institution to another within Australia and sample collection was initiated as soon as they arrived at the new institution. For yellow-bellied gliders, and southern bettongs, transit time was 5–6 h via plane and automobile. For the long-nosed potoroo, transit time was approximately 1 h via automobile. In these cases, peak values

Table 1
Information for species used in this study. "Abb." indicates the species abbreviation used in subsequent tables and figures (first letter of the Genus plus first two letters of the species). Reference provided if samples were previously used in another study (Note: one female Bpe was tested twice under different conditions and appears in both Bpe rows).

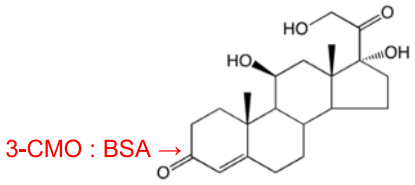
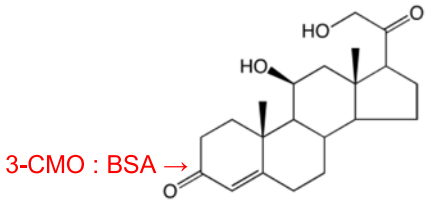
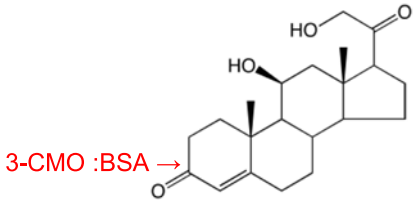
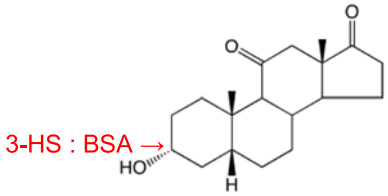
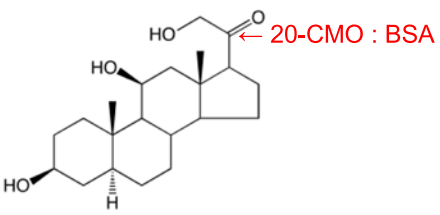
Order	Species	Abb.	# Animals (Male. Fem)	Location	Stressor (manufacturer)	ACTH dose (IU/kg)	Extraction (faeces/MeOH)	References
Dasyuromorphia	Numbat *Myrmecobius fasciatus*	Mfa	2.2	Perth Zoo	ACTH (Synacthen)	3	0.2 ± 0.01 g dry/4.5 mL	Hogan et al. (2012)
	Tasmanian devil *Sarcophilus harrisii*	Sha	2.2	Taronga Western Plains Zoo	ACTH (Synacthen)	4	0.2 ± 0.01 g dry/4 mL	Keeley et al. (2012)
Peramelemorphia	Bilby *Macrotis lagotis*	Mla	2.2	Dreamworld & Currumbin Sanctuaries	ACTH (Synacthen)	3	0.2 ± 0.01 g dry/5 mL	Narayan et al. (2014)
Diprotodontia	Koala *Phascolarctos cinereus*	Pci	2.2	Taronga Zoo	ACTH (Synacthen)	7	0.5 ± 0.01 g wet/5 mL	Narayan et al. (2013)
	Northern hairy-nosed wombat *Lasiorhinus krefftii*	Lkr	0.1	Department of Environment and Heritage Protection	Moved enclosures		0.5 ± 0.01 g wet/5 mL	
	Southern hairy-nosed wombat *Lasiorhinus latifrons*	Lla	0.3	Rockhampton Botanical Gardens and Zoo	Vet check/capture		0.5 ± 0.01 g wet/5 mL	Hogan et al. (2011)
	Mountain pygmy-possum *Burramys parvus*	Bpa	4.4	Healesville Sanctuary	Mating introductions		0.05 ± 0.001 g wet/1 mL	
	Yellow-bellied glider *Petaurus australis*	Pau	2.0	Moonlit Sanctuary	Transfer		0.5 ± 0.01 g wet/5 mL	
	Long-nosed potoroo *Potorous tridactylus*	Ptr	1.0	Moonlit Sanctuary	Transfer		0.05 ± 0.001 g wet/1 mL	
	Woylie *Bettongia penicillata*	Bpe	2.1	Perth Zoo	ACTH (Synacthen)	6	0.2 ± 0.01 g dry/4.5 mL	
		Bpe2	0.1	Perth Zoo	Moved enclosures + ACTH (Synacthen)	9	0.2 ± 0.01 g dry/4.5 mL	
	Southern bettong *Bettongia gaimardi*	Bga	2.0	Moonlit Sanctuary	Transfer		0.5 ± 0.01 g wet/5 mL	
	Eastern grey kangaroo *Macropus giganteus*	Mgi	1.3	Currumbin Wildlife Sanctuary	ACTH (Corticotropin)	10	0.5 ± 0.01 g wet/5 mL	
	Western grey kangaroo *Macropus fuliginosus*	Mfu	3.1	Fowlers Gap, Broken Hill, NSW	ACTH (Synact. depot)	10	0.5 ± 0.01 g wet/5 mL	

were expected immediately, so sample collection continued at least 10 days following transport to ensure baseline FGM values could be adequately measured. The northern hairy-nosed wombat was a juvenile female that was being hand-reared; once weaned, she was transferred to an outdoor enclosure, which served as the stressor for this study. For the mountain pygmy-possums, males were moved into females' enclosures for mating. Samples were collected before, during, and after the introduction when individuals were being weighed, so all samples are from known individuals. For the southern hairy-nosed wombats, animals were caught and anaesthetized for a routine vet check or to re-fit a radio-telemetry collar.

2.2. Collection and extraction of faecal samples

An average of 14.5 samples (range: 6–34 samples) were collected from each individual. Faecal samples were stored at a minimum of −20 °C for no more than four years. Samples were shipped overnight on ice to the Wildlife Reproductive Centre (Dubbo, NSW, Australia), where they were extracted just prior to the start of this study (but see below for two exceptions). Therefore, in most cases the extracts used in this study were not the same extracts that were used in the papers referenced in Table 1. FGM were extracted by suspension in 80% methanol, typically using 0.5 g faeces in 5 mL methanol (Palme et al., 2013). This protocol was adjusted (as stated in Table 1) when a species' samples were exceptionally small or had been previously dried (overnight at ~65 °C, pulverized, and sifted to remove debris). For individual samples that did not have sufficient material to reach the specified faecal volume, the volume of methanol was reduced accordingly (e.g., 0.25 g faeces in 2.5 mL methanol); this adjustment was made for no more than 5% of samples per individual. Faeces were weighed into a glass vial, 80% methanol was added, and samples were vortexed and placed on a rotator at ambient temperature overnight. The next morning, vials were centrifuged for 15 min at 1000 g. The supernatant was decanted into a clean glass scintillation vial and stored at −20 °C until analysis. In two cases, faecal material was not available for re-extraction, so we used existing extracts. For numbats, methanol extracts were stored for two years at −80 °C and then −20 °C

Table 2
Information for enzyme immunoassays used in this study. The reference is provided for the first description of the assay (commercial kit excepted).

Source	Assay ID	Immunogen		Sensitivity (ng/ml)	Intra-assay CV* (high, low)	References
Coralie Munro (*University of California Davis, California, USA*)	Cortisol (R4866)	Cortisol *3-CMO: BSA*		0.08	3.4, 6.3	Munro and Stabenfeldt (1985)
	Cs6 (CJM06)	Corticosterone *3-CMO: BSA*		0.098	4.2, 12.0	Watson et al. (2013)
Arbor Assays (*Ann Arbor, Michigan, USA*)	AAcort	Corticosterone *3-CMO: BSA*		0.08	3.8**	
Rupert Palme (*University of Vet Medicine Vienna, Austria*)	72a	11-Oxoaetiocholanolone *3-HS: BSA*		0.04	9.2, 14.5	Palme and Möstl (1997)
	37e	3β,5α-Tetrahydrocorticosterone *20-CMO: BSA*		0.02	9.7, 13.4	Touma et al. (2003)

* All samples for an individual were run on the same plate. %CV calculated from concentration data.
** Only one control in the middle of the curve was used for this assay.

before being shipped to the lab for analysis. For the northern hairy-nosed wombat, samples were extracted in our lab as described above, but the methanol extract was dried down and reconstituted in assay buffer prior to storage for one year prior to analysis.

2.3. Steroid assays

Samples were analysed using five previously described EIAs expected to target different metabolites (Table 2). Assays were biochemically validated for each species by demonstrating parallelism between a serially-diluted extract pool and the standard curve (Electronic supplementary material Fig. 3). FGM concentrations are expressed as ng/g faeces.

The antibodies provided by Coralie Munro (University of California Davis, Davis, CA, USA) were raised in white New Zealand rabbits to measure circulating GCs (either cortisol (hereafter Cortisol) or corticosterone (hereafter Cs6)). The corresponding hormone conjugate was labelled with horseradish peroxidase (HRP). Although these antibodies are fairly specific (i.e., have low cross-reactivity with other steroids tested), previous research has shown that they react with several FGM and can be used to monitor adrenocortical activity in a range of species (Watson et al., 2013; Young et al., 2004). Assay methods followed previously published protocols (Fanson et al., 2013). Briefly, 96-well plates (Nunc maxisorp) were coated with 0.05 ml antibody (working dilution: Cortisol = 1:9000; Cs6 = 1:12,500) and incubated at 4 °C overnight. After washing 3 times, 0.05 ml of steroid standard, diluted faecal extract, or control were added to each well followed by 0.05 ml of HRP-conjugate (working dilution: Cortisol = 1:20,000; Cs6 = 1:20,000). Plates were incubated for 1 h (Cortisol) or 2 h (Cs6), then washed 3 more times to remove unbound hormone and 0.10 ml ABTS substrate was added. Optical density was measured at 405 nm using a Dynex MRX Revelation plate reader (Dynex Technologies, Chantilly, VA, USA).

The assay provided by Arbor Assays (hereafter AAcort), used the same antigen and enzyme label as Cs6, but the antibody was raised in sheep. Assay methods followed manufacturer instructions.

The assays produced in Vienna (University of Veterinary Medicine, Vienna, Austria), specifically target FGM, and therefore are tailored to broad-spectrum FGM analysis (Möstl et al., 2005). The antibodies were raised in rabbits against 11-oxoaetiocholanolone (hereafter 72a; Palme and Möstl, 1997) and 3β,5α-tetrahydrocorticosterone (37e; Touma et al., 2003). Assays were run as described in the source papers. Briefly, 96-well plates were incubated with 0.25 ml of coating buffer containing Protein A for 24 h, emptied, then incubated overnight with 0.30 ml Trizma buffer solution rich in bovine serum albumin and containing sodium azide. Plates were emptied, blotted, and immediately frozen. Just before use, plates were thawed and washed 3 times. For the assay, 0.05 ml of steroid standard, diluted faecal extract, or control were added to each well, followed by 0.1 ml biotinylated steroid (working dilution: 72a = 1:500,000; 37e = 1:15,000) and 0.1 ml of primary antibody (working dilution: 72a = 1:40,000; 37e = 1:15,000). Equilibrium binding was achieved during overnight incubation at 4 °C. Plates were washed 3 times before adding 0.25 ml streptavidin-peroxidase to each well. Following a 45-min incubation at 4 °C, plates were washed 6 times and 0.25 ml TMB substrate was added. The reaction was stopped with 0.05 ml H_2SO_4 and optical density was measured at 450 nm using a Dynex MRX Revelation plate reader.

2.4. Statistical analysis

2.4.1. Overall comparison of assay performance

We first examined broad correlations among the different assays with the aim of identifying which assays yielded similar profiles. The value of this information is that it may offer predictive power for choosing alternative assays for a species. Overall, we expect strong intercorrelations among the EIAs because although they are all designed to pick up different FGM, these FGM are all derived from GCs (or at least from steroids produced by the adrenals) that increase after stimulation. A smaller intercorrelation might indicate that steroids from other sources (e.g., gonads) are being co-measured. In particular, 72a is known to co-measure some androgen metabolites (Ganswindt et al., 2003). Using principal component analysis with factor rotation can highlight which assays are more correlated and which differ from others by looking at each assay's loading score. This can provide a clue as to the most important characteristics affecting similarity of assay performance. One possibility is that assays will yield similar profiles (and consequently load strongly on the same factor) if antibodies are raised against similar antigens. If this is true, we would expect AAcort and Cs6 to be highly correlated since they are raised against the same antigen (see Table 2). Conversely, 37e and 72a should be the least correlated with other assays and each other, since the antigens used in these two assays are quite distinct. A second possibility is that assay performance may be strongly influenced by the manufacturing and purification process. In this case, we would expect assays produced by the same manufacturer to be more correlated (i.e., Cortisol and Cs6 should be correlated; 37e and 72a should be correlated; AAcort should be the least correlated with other assays). Third, assay performance may be most dependent on the enzyme label used. This would suggest that the three EIAs using horseradish peroxidase (HRP) as a label, Cortisol, Cs6, and AAcort, should be similar, while the two EIAs using biotinylated labels, 37e and 72a, might be similar. A final possibility is that the performance of any given assay is determined by the combination of a complex array of factors, making it difficult or impossible to predict which assays may be suitable alternatives.

In order to examine associations among assays, FGM concentrations (log-transformed) were first centred for each individual and EIA. This allowed us to focus on trends within an individual so that observed relationships were not confounded by among-individual variation. Values were then standardised across species by calculating the *Z*-score for each species and EIA. This both centred the data so that observed relationships were not driven by species variation and standardised the variance so that no species or assay was particularly influential in driving the results. Scatterplots of pairwise comparisons (both overall and within each species) were used to get a sense of how assays were related (see Electronic supplementary material Fig. 1).

Principal component analysis was conducted in SAS 9.4 (Cary, NC; PROC FACTOR) to examine overall associations among the assays. The analysis only included species for which there was data for all five assays (i.e., the three species that did not demonstrate parallelism on the cortisol assay were excluded). Varimax rotation

Table 3
Results of principal components analysis illustrating the correlation among assays. The fact that all assays loaded heavily of PC1 in the unrotated factor loading pattern shows that all assays were positively correlated. However, the varimax rotation, which maximises loading of each variable on one PC, highlights differences among the assays. Bold indicates loading scores greater than 0.6.

Assay	Unrotated		Rotated		
	PC1	PC2	PC1	PC2	PC3
37e	**0.74**	0.62	0.26	0.23	**0.90**
72a	**0.75**	−0.13	0.24	**0.89**	0.28
AAcort	**0.88**	0.01	**0.73**	0.29	0.43
Cortisol	**0.78**	−0.42	**0.73**	0.49	0.01
Cs6	**0.79**	−0.06	**0.83**	0.09	0.32
Eigenvalue	3.12	0.58	1.86	1.19	1.18

was used to maximise an assay's loading on one factor and thereby simplify interpretation (O'Rourke et al., 2005).

2.4.2. Assay performance within each animal

We then assessed the performance of each of the five assays at the level of the individual animal. To determine whether any of the post-stressor values qualified as 'peaks', we used the iterative baseline approach was used to identify baseline and peak values (Brown et al., 1994). In this approach, the baseline is calculated through an iterative process excluding points greater than mean + 2 SD until no points fall above this threshold. The remaining points are considered to provide the baseline value, which is described using M and SD. Samples that fall above the threshold (i.e., are excluded from the final baseline calculation) are classified as peaks. This calculation was performed separately for each assay for each individual using the R package hormLong (Fanson and Fanson, 2015).

Biological sensitivity of the assay was tested by looking at the signal-to-noise ratio. To determine the magnitude of the post-event increase (signal) relative to the within-baseline variation (noise) detected by that assay, we calculated a *Z*-score for each individual: $Z = (\text{Maximum} - \text{Mean}_{\text{baseline}})/\text{SD}_{\text{baseline}}$. For the purposes of this paper, we identified the maximum value that occurred between 0.125 days (3 h) and 4.5 days after the event, even if the point was not classified as a 'peak' (above). While most studies of FGM responses to acute events in eutherian mammals use a window of 2 days following the event (e.g., Young et al., 2004), previous marsupial studies have found the FGM peak can occur 3 days after ACTH injection (e.g., bilbies (Narayan et al., 2014); southern hairy-nosed wombats (Hogan et al., 2011)). This may be due to the relatively long gut passage times seen in some marsupial species (50–100 h; Barboza, 1993; Davies et al., 2013; Munn and Dawson, 2006). Therefore, we chose a 4-day window to ensure no peaks were missed. In addition, we used the maximum FGM value during this window rather than the initial increase in FGM because it was more objective to quantify.

To calculate the time-to-peak interval for each assay for each individual, we identified the maximum FGM value between 0.125 and 4.5 days post-stressor, as described above. The maximum FGM concentration in the post-event timeframe was used for this calculation even if that sample was not sufficiently different from baseline variation to be classified as a 'peak' by the iterative method. We then calculated the time between the stressor and the occurrence of the maximum value. This should not be confused with excretion lag time, which is calculated as the time from onset of stressor to the initial increase in FGM. Due to the number of different methods included in this study and variation in sampling frequency, we felt it would be misleading to estimate excretion lag time. Nonetheless, we feel time-to-peak gives useful information about when to increase sampling intensity.

2.4.3. Assay performance for each species

Using the results of the individual assessment, we used two criteria to assess assay performance for each species: (1) the percent of individuals with a peak identified using the iterative method, and (2) the average *Z*-score for each species. The first measure pro-

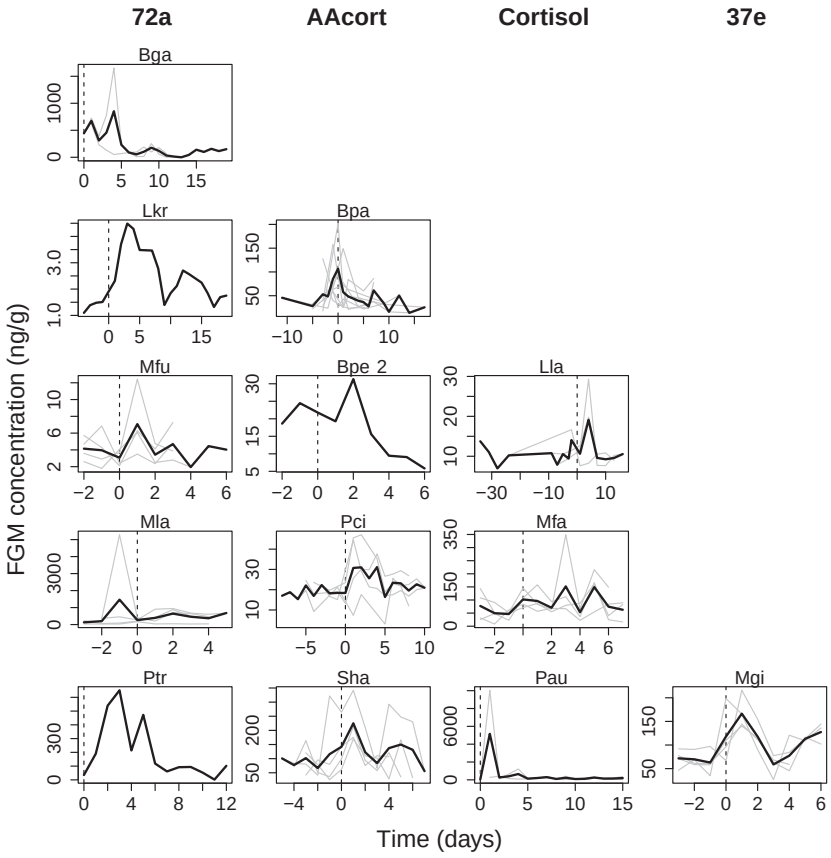


Fig. 1. Best performing assay for each species (see Table 1 for full species names). Assays were determined to be the "best" if they (1) detected peaks in the most individuals, and (2) had the greatest biological sensitivity (i.e., highest *Z*-score; see Table 4). Species plots are ordered in columns based on the best performing assay (assay name at top of each column). Assay 72a was identified as the best assay in the most species ($n = 5$), whereas Assay Cs6 was not identified as the best for any species. Within each species' plot, the grey lines represent individual profiles (FGM values averaged by day), the black line represents species mean, and the vertical dashed line indicates time of stressor.

Table 4
Comparison of assay performance for each species. The top number indicates the percentage of individuals for which a peak was detected between 0.125 days (3 h) and 4.5 days post-stressor using the iterative baseline approach. The bottom number represents the average species *Z*-score calculated for the maximum post-stressor value and provides a measure of the strength of the peak relative to the noise in the baseline. Colour-coding was determined using the following approach. Initial colour was based on the percentage of individuals with peaks (<50% (unsuccessful) = white; 50–75% = blue 1 (lightest); 100% = blue 2). If the peak *Z*-score was >5 (good signal-to-noise ratio), shading was increased one level. Thus darker blue represents better performing assays.

Family	spp. abb (Table 1)		37e	72a	AAcort	Cortisol	Cs6
Dasyuromorphia							
	Mfa	% with peaks	50%	0%	50%	50%	25%
		Z-score	3		4	4.5	2.8
	Sha		50%	25%	50%	50%	50%
			3.5	2.9	4	3.5	3.2
Peramelemorphia							
	Mla		0%	25%	0%	0%	0%
				2			
Diprotodontia							
	Pci		25%	50%	50%	25%	25%
			3.1	2.5	6.4	1.6	1.7
	Lkr		0%	100%	100%	NA	100%
				2.8	2.7		2.5
	Lla		33%	33%	33%	66%	33%
			2.2	1.4	3.2	6.7	5.7
	Bpa		0%	0%	12.5%	0%	0%
					1.6		
	Pau		100%	100%	100%	100%	100%
			16.6	8.1	11.2	43.1	6.3
	Ptr		0%	100%	0%	100%	0%
				8.4		3.4	
	Bpe		0%	0%	0%	NA	0%
	Bpe2		100%	100%	100%	NA	0%
			5.4	8.8	11.1		
	Bga		50%	100%	100%	50%	100%
			2.2	7.5	3.6	3	5.1
	Mgi		100%	50%	0%	50%	25%
			4	3.4		4.4	2.5
	Mfu		25%	75%	75%	NA	25%
			2.4	6	4.9		2.1

vides an index of how robust the assay is at detecting peaks across individuals, while the second provides an index of how sensitive the assay is. Given the strict criteria used in the iterative method, our minimum standard for a successful assay was when a post-event peak was detected in 50% or more of the individuals tested. We further rated the quality of these successful assays by factoring in the average signal/noise ratio (i.e., *Z*-score). Thus, if two assays both detected peaks in 100% of the individuals in a species, the assay with the higher *Z*-score (i.e., more sensitive assay) was identified as the best performing assay for a species.

3. Results

3.1. Overall comparison of assay performance

In general, assays were positively correlated with each other, indicating that they capture similar broad-scale patterns of FGM. This was reflected in the unrotated factor loading scores (Table 3). All factors loaded heavily on Component 1, which accounts for 60% of the variance. However, the varimax rotation highlighted some differences in assay alignment, which were also generally apparent in the pairwise comparisons (Electronic supplementary material Fig. 1). Assays AAcort, Cortisol and Cs6 all loaded heavily on rotated PC1. AAcort and Cs6 were raised against the same antigen, and the within-species pairwise comparisons indicated that this was the strongest correlation for most species. Assays 37e and 72a, which have very different antigens, loaded heavily on separate PCs.

3.2. Assay performance within each animal

Of the 212 total combinations of assay type and individual animal, the iterative method identified post-event peaks in 74 (35%). Of the peaks identified, 55% were observed in only a single sample, which highlights how transient peaks can be and thus the importance of rigorous sample collection. *Z*-scores from these peaks had a mean of 7.6 ± 9.7. The post-event time at which the maximum value (not the initial increase) for each animal was observed varied from 0.3 to 4 days (Electronic supplementary material Table 1). For each species, mean and individual profiles for the best performing assay are presented in Fig. 1. Data for all species and assays are available on an interactive webpage (https://bfanson.shinyapps.io/ausFaunaApp/). See Electronic supplementary material for species-specific discussion of results and assay recommendations. In this paper, we focus on summary data and evaluation of assay performance.

3.3. Assay performance for each species

For most species, one or more successful EIAs were identified (i.e., at least one EIA detected a post-event peak in 50% or more of the individuals tested). The best performing assay for each species was identified as the assay that detected peaks in the most individuals (based on the iterative approach) and had the greatest biological sensitivity (i.e., highest *Z*-score). Fig. 1 shows profiles and average FGM concentrations for each species as measured by that species' most successful assay. Only for yellow-bellied gliders did all five assays identify a peak in both animals. No successful assay was identified using our iterative baseline approach for bilbies or mountain pygmy-possums. For woylies, no EIAs were successful in the first test (Bpe: 6 IU/kg ACTH, 1 sample/day), but three EIAs were successful in the second test (Bpe2: 9 IU/kg ACTH plus enclosure change, 3 samples/day; Electronic supplementary material Fig. 2). There was considerable variation in assay performance among species, even closely related species (Table 4, interactive data from individual species/assay combinations available at https://bfanson.shinyapps.io/ausFaunaApp/). For example, different assays were successful for the two *Lasiorhinus* spp. (northern and southern hairy-nosed wombats), as well as the two *Bettongia* spp. (southern bettong and woylie). From the perspective of the assays, the 72a and AAcort assays were successful in the greatest number of species (8 species), followed by Cortisol (7 species), then by 37e and Cs6 (6 species each).

4. Discussion

4.1. Overall comparison of assay performance

Monitoring faecal glucocorticoid metabolites (FGM) provides a valuable, non-invasive tool for assessing an animal's condition and well-being. However, quantification of FGM is very dependent on the assay that is used. This study offers a large comparative examination of the performance of five assays for monitoring changes in adrenocortical activity via FGM in 13 marsupial species. We identified assays that reflected biologically relevant changes in adrenocortical activity for 10 species of marsupials. However, there were no apparent phylogenetic trends with regards to which assays were successful (Table 4). This reiterates that biological validation must be conducted for each species; identification of an appropriate assay cannot be assumed based on what works for closely related species (Touma and Palme, 2005).

When subjected to a principal component analysis, these five assays covaried highly, but following rotation they were divisible into three distinct families (Cortisol/Cs6/AAcort, 37e, and 72a). This

indicates that all five assays measure similar gross changes in adrenocortical activity, but their separation after rotation suggests that they measure different FGM. The three assays in which antibodies were raised to measure a circulating GC (Cs6, Cortisol, and AAcort) continued to load highly on a single factor after rotation, while each of the group-specific antibodies (72a, 37e) shifted to its own factor. This suggests that the antigen used to develop an assay, and perhaps the type and attachment of the enzyme label, has a greater influence on assay performance than other aspects of the EIA set-up (Möstl et al., 2005). While the Cortisol, Cs6, and AAcort antibodies all target FGM with similar side-chains on ring D of FGM, the 72a antibody targets FGM in which the side-chain has been cleaved from ring D, and the 37e antibody targets ring A. Thus it is likely these assays are binding different subsets of FGM with different structures.

Although assays that load on the same factor may yield similar profiles, they are still likely to differ in key aspects such as the amount of noise in the baseline and the magnitude of the peaks. The three assays that load together can best be thought of as members of a family; they are related but not always interchangeable. As an illustration, note that the number of marsupial species for which all three of these assays were successful (3 species) is greatly outnumbered by the number of species where only one or two of these EIAs were successful (8 species). However, this information about assay performance may highlight useful next steps in the biological validation process. If a particular assay shows the right trend but the biological sensitivity to stressors is not optimal, it may be beneficial to test an assay designed to detect a similar metabolite structure but which may yield a slightly stronger signal.

4.2. Considerations for experimental design

The results of this study highlighted several elements of experimental design that contribute to successful peak detection. Two critical determinants of whether peaks were detected were (1) stimulus used to stimulate adrenocortical activity (type, magnitude, and duration), and (2) sampling intensity. Because methodological factors can be so influential, it is particularly important during biological validation to adopt a robust experimental design that maximises the likelihood of detecting changes in adrenocortical activity.

The animals in this study that were transferred between institutions (yellow-bellied gliders, long-nosed potoroo, and southern bettong) tended to have a higher percentage of individuals in which a peak was detected and a greater number of assays that detected the peak (Table 4). Transport is intensely stressful and lasts for some time (Grandin, 1997), so it is likely to provoke a more dramatic and sustained increase in GC than an acute stressor or an ACTH injection. Since transport between facilities is often required for maintaining populations of captive animals, collecting faeces before and after transport should be a reliable and minimally invasive way to validate FGM assays in captive populations (e.g., Palme et al., 2000). However, animals may still show individual differences in GC response to transport based on their previous experience and temperament (Grandin, 1997). Unlike transported animals, those that experienced husbandry changes within an institution (southern hairy-nosed wombat and mountain pygmy-possum) were less likely to have peaks detected by any of the assays. This is consistent with findings by Hogan et al. (2012) that for numbats, interactions among animals and health issues resulted in significantly higher FGM concentrations than environmental events or husbandry changes, which included moving animals to a new enclosure within the facility. We suggest that most within-institution husbandry procedures are unlikely to be consistently stressful enough to be used in validation experiments (but see Hogan et al., 2011 and McKenzie and Deane, 2005 for counter-examples).

Among species that received an ACTH injection, we found that the dose of the injection influenced the likelihood of detecting a signal. FGM peaks were more likely to be detected when species received a higher dose of ACTH (e.g., eastern and western grey kangaroos – 10 IU/kg) than when they received a lower dose (e.g., numbat, bilby – 3 IU/kg). GC release in response to ACTH is dose-dependent up to a plateau of maximum GC release (Patel and Clayton, 1999). Identifying an appropriate ACTH dose for validation studies is challenging, particularly because there is often little information available for exotic (non-domesticated) species (Hunter, 2010). The dosage depends on (1) species or individual responsiveness to synthetic ACTH, (2) the biological sensitivity required in a particular study, and (3) defecation frequency, or the length of time over which circulating GCs are pooled, which will dilute signals in circulating GCs. Dose–response experiments or comparisons with response to actual stressors could help identify an ideal ACTH dose to use (Martin and McDonald, 1986; Narayan et al., 2014), but are often difficult to conduct in exotic species. It is also important to remember that dose rates generally scale allometrically, such that larger animals tend to require a smaller dose rate, but this scaling factor is complicated by physiological differences among species (Hunter, 2010). In order to elicit a change in circulating GCs that is sufficient enough to be detected in faecal samples, it may be necessary to administer a larger dose, use a slow-release gel, or give multiple injections. This is particularly true in species that defecate less frequently, in which case the long window of pooling will minimise acute changes in adrenocortical activity. To ensure that an assay is biologically sensitive enough to measure the stressor of interest, we recommend testing actual events rather than using an ACTH challenge, unless there is sufficient information for a species to determine an appropriate dose.

Sampling intensity is even more important in validation studies. Species with poor sampling coverage (e.g., southern hairy-nosed wombats and mountain pygmy-possums – every few days) were less likely to have peaks detected than species with really good coverage (e.g., eastern and western grey kangaroos – several times per day). This may be a bigger issue for most marsupial species, which are herbivorous and defecate frequently, than for other taxa (e.g., Young et al., 2004). Even with relatively intense stressors, such as transfer between institutions, the peak was sometimes present in only a single sample. This highlights the need for frequent sampling when biologically validating assays or monitoring responses to acute stressors (Palme, 2012; Touma and Palme, 2005). Sampling intensity is important not only for post-stressor samples, but also for baseline estimates of adrenocortical activity. In order to identify peaks, it is important to have a good assessment of undisturbed baseline. This can be somewhat challenging since stressors are unpredictable and often unnoticed by researchers or keepers. Therefore, the baseline sampling period must be sufficient to accommodate these unpredictable increases in adrenocortical activity and still obtain a robust baseline estimate.

Woylies provide a good illustration of the importance of both stressor magnitude and sampling frequency (Electronic supplementary material Fig. 2). In the first validation attempt, all three individuals received a dose of 6 IU/kg and samples were collected once per day (Bpe). None of the assays detected post-ACTH peaks in any individual. One of the individuals was subsequently retested using a dose of 9 IU/kg given at the same time as an enclosure change and samples were collected three times per day (Bpe2). This revealed a very distinct peak in a single sample, particularly on assays 72a and AAcort. The contrast between these two sets of results suggests that either the small magnitude of the event

or the low sampling frequency caused validation failure in the first study, rather than an inability of the EIAs to detect woylie FGM.

4.3. Considerations for data analysis

There are a number of peak detection methods that can be used in biological validation studies. Most validation studies use one of two peak detection approaches: (1) assessing the average population response, or (2) examining individual animal responses. Assessing average response focuses on the magnitude of the increase in FGM for the whole population following a challenge, typically with parametric statistics like related-samples *t*-tests (Hogan et al., 2011, 2012) or mixed linear models including the individual as a random factor (Keeley et al., 2012). Parametric techniques are excellent for detecting overall changes in FGM in a group of animals and include an explicit assessment of the probability that the observed difference was due to chance. However, an individual with a large response can have an undue influence on the mean, obscuring individuals with small or no responses. Conversely, if the timing of the peak varies among individuals, then such analyses may fail to detect a significant response at the population level, even if there are pronounced peaks at the individual level. Therefore, null results may reflect limited power of the test rather than failure of the validation attempt.

Individual-centred techniques such as the iterative approach excel at identifying whether an assay detects a change for each individual, and can be combined (as we have here) to determine how consistently effects are detected across individuals. These techniques are most appropriate when sex or individual differences in response patterns preclude combining individuals, as done in parametric tests. For example, Narayan et al., (2013, 2014) used visual inspection of plots to validate the cortisol assay used in this study for measuring ACTH-induced increases in FGM in koalas, where males had longer latencies to peak than females (Narayan et al., 2013), and in bilbies, where large individual differences in latency were observed (Narayan et al., 2014). The disadvantage of visual inspection is that it can be subjective, whereas the iterative approach provides a more objective assessment. Ideally, some quantitative confirmation that a peak value is unlikely to occur by chance should be sought either using an indicator that the peak magnitude is unusual in the population of baseline values (*Z*-scores, this study), or using a control condition [e.g., comparing ACTH to a saline injection (Narayan et al., 2014); comparing transport to a stationary vehicle (Palme et al., 2000)].

These different calculations have consequences for the comparability of results. This study included a subset of samples that had been previously analysed using one of the assays for biological validation: numbat (Cortisol: Hogan et al., 2012), Tasmanian devil (Cs6: Keeley et al., 2012), bilby (Cortisol: Narayan et al., 2014), and koala (Cortisol: Narayan et al., 2013). The numbat and Tasmanian devil studies used parametric statistics. Using our iterative baseline approach, we confirmed that Cortisol was the most successful assay for numbat and Cs6 was one of the most successful assays for Tasmanian devils. However, in each species we found that the assay identified peaks in only half of the individuals. This is consistent with the suggestion that parametric statistics may allow large responses in some individuals to mask a lack of response in others. For numbats, it is also possible that the FGM had degraded in storage since faecal extracts had been stored (first at −80 °C, then at −20 °C) for a total of 2 years, the oldest of any extracts used in this study. In bilby and koala, which had previously been analysed using visual inspection and percent increase in FGM, our iterative baseline approach did not successfully identify peaks in the Cortisol assay even though our Cortisol graphs looked similar to the published figures. Although increased FGM were observed following injection for these species/assay combinations, variation throughout the sampling period was too large for any peaks to be identified. We recommend using the iterative baseline approach because it combines the advantages of parametric and individual-centred approaches, accommodating individual differences in responsiveness while taking into account the signal-to-noise ratio and defining peaks relative to the amount of variance present in the baseline.

4.4. Concluding remarks

This study highlights the importance of biologically validating assays used for non-invasive FGM monitoring for each species, and illustrates the value of comparing multiple assays. The lack of a consistent relationship between assay success and taxonomic relatedness reiterates earlier suggestions that a positive validation must be obtained for each species (Touma and Palme, 2005). In addition, because there are several methodological factors that can affect the outcome of a validation study (highlighted above), we suggest caution be used in interpreting negative results in a validation attempt. If a validation attempt is unsuccessful, that could indicate the assay is not appropriate for that species, but it could also indicate insufficient stressor magnitude/ACTH dose, low sampling frequency, or overly conservative peak-detection techniques. If validation of one assay is unsuccessful, measuring FGM in the same samples using different assays may produce different results. Our study provides a good starting point for determining which assays are appropriate for several marsupial species, even though we encountered some limitations. Small sample sizes within each species prevented us from assessing the effects of sex, reproductive state, or unique individual circumstances, all of which have been documented to affect FGM profiles in some species (Touma and Palme, 2005).

Furthermore, it is important to recognise individual variation in biological validation studies and adopt an analysis approach that accommodates this variability. Individuals often differ in their response to a stressor (peak magnitude and timing). This may reflect differences in perception of the stressor, time of day, metabolic rate, and/or patterns of steroid metabolism based on sex, age, or diet (Goymann, 2012). While in some species a single assay may work well for many questions (e.g., 37e in mice; Touma et al., 2004), other species may not have a single "perfect" assay; rather, there may be several assays that perform well in different regards. Researchers need to identify which assay characteristics are most relevant to their questions of interest and select an assay accordingly. For example, if a researcher is only studying females, then they can choose the assay that detected the biggest peak in females even if it was not particularly effective at detecting peaks in males. Conversely, if trying to monitor both sexes, then it might be necessary to settle for an assay that detects peaks more consistently, even if the signal-to-noise ratio is not as good.

Faecal steroid analysis provides a valuable tool for assessing the physiological status of an animal non-invasively. Marsupial studies examining basic physiology, conservation, or management approaches can all benefit from such a technique, but it is crucial that assays be properly biologically validated. Here, we provide the most extensive assessment of FGM assays in marsupials to date. Researchers can help ensure robust outcomes of validation studies by: (1) testing multiple assays in order to optimise signal-to-noise ratio and match biological sensitivity to the scale of the event studied (i.e., relative increase following a stressor), (2) ensuring that the stressor or ACTH dose is of sufficient magnitude to produce a measureable increase in FGM, and (3) adopting a fairly intensive sampling regime during the validation process to improve the estimate of baseline, ensure that peaks are not missed, and that lag time is accurately identified. In addition, we suggest researchers adopt the iterative baseline approach (Brown et al.,

1994), which provides a quantitative peak detection method while simultaneously accommodating individual variation. In addition to peaks, researchers should also evaluate the amount of variance present in the baseline, attempt to minimise variance from sources that are not of interest, and use a metric such as *Z*-score to assess signal-to-noise ratio.

Acknowledgments

We would like to thank all the people and marsupials who made this project possible. Thanks to Russ Hart from Arbor Assays for his generous contribution to this project and to Rosy Kirkup for her assistance in the lab. Funding was provided by Taronga Conservation Society Australia, Deakin University, and Franklin and Marshall College. Acknowledgements for specific species are as follows:

Bilbies: Thanks to the veterinarians Dr. Vere Nicolson and Dr. Michael Pyne who assisted with applying the ACTH treatments to greater bilbies sampled from the Dreamworld Theme Park and the Currumbin Wildlife Sanctuary. Staff of the Dreamworld Theme Park, Currumbin Wildlife Sanctuary and Honours student (Nicole Evans) collected faecal samples, and Griffith University provided funding.

Eastern grey kangaroos: Thanks to Currumbin Wildlife Sanctuary for supporting this project and to Dr Adrian Bradley for use of his lab. The research was funded by The University of Queensland.

Koalas: The koala ACTH challenge was performed at the Taronga Zoo veterinary facilities, under the supervision of Dr. Larry Vogelnest. Staff of the Australian Mammals section at Taronga Zoo collected faecal samples, and Taronga Conservation Society Australia provided funding.

Long-nosed potoroos, southern bettongs, & yellow-bellied gliders: Thanks to the keeping staff at Moonlit Sanctuary for collection of samples.

Northern hairy-nosed wombat: Thanks to Alan Horsup from the Department of Environment and Heritage Protection and Tina Jansen from Project Kial for their contributions.

Numbats: Thanks to Caroline Lawrence at Perth Zoo for organizing and shipping samples.

Mountain pygmy-possums: Thanks to Healesville Sanctuary, Paula Watson and Matt West who assisted with possum care and collection of samples.

Tasmanian devils: Thanks to the Taronga Western Plains Zoo veterinary staff and zookeepers for their assistance in the ACTH challenge and faecal sample collection. Funding was provided by the Morris Animal Foundation, USA, and the SeaWorld and Busch Gardens Conservation Fund, USA.

Western grey kangaroos: Thanks to Dr Steve McLeod and the assistants who worked with us at the UNSW Fowlers Gap Arid Zone Research Station. Funding and support was provided by the RIRDC and NSW Department of Primary Industries.

Woylies: Thanks to Dr Simone Vitali and Dr Kris Warren for their supervision of this project. Thanks to the veterinary staff of the Perth Zoo Veterinary Department and the keeping staff of the Native Species Breeding Program and the Nocturnal House for their assistance in ACTH administration, animal care, and collection of faecal samples. Funding was provided by Wildlife Conservation Action, Perth Zoo, and Murdoch University.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ygcen.2015.10.011.

References

Barboza, P.S., 1993. Digestive strategies of the wombats: feed intake, fiber digestion, and digesta passage in two grazing marsupials with hindgut fermentation. Physiol. Zool., 983–999

Bradley, A.J., 2003. Stress, hormones and mortality in small carnivorous marsupials. In: Jones, M., Dickman, C., Archer, M. (Eds.), Predators with Pouches – The Biology of Carnivorous Marsupials. CSIRO Publishing, Victoria, Australia, pp. 255–267.

Brown, J.L., Citino, S.B., Shaw, J., Miller, C., 1994. Endocrine profiles during the estrous cycle and pregnancy in the Baird's tapir (*Tapirus bairdii*). Zoo Biol. 13, 107–117.

Carere, C., Caramaschi, D., Fawcett, T.W., 2010. Covariation between personalities and individual differences in coping with stress: converging evidence and hypotheses. Curr. Zool. 56, 728–740.

Dallmann, R., Touma, C., Palme, R., Albrecht, U., Steinlechner, S., 2006. Impaired daily glucocorticoid rhythm in Per1(Brd) mice. J. Comp. Physiol. A. 192, 769–775.

Davies, N., Gillett, A., McAlpine, C., Seabrook, L., Baxter, G., Lunney, D., Bradley, A., 2013. The effect of ACTH upon faecal glucocorticoid excretion in the koala. J. Endocrinol. 219, 1–12.

Fanson B.G., Fanson K.V., 2015. hormLong: an R package for the longitudinal analysis of hormone data. <https://github.com/bfanson/hormLong>.

Fanson, K.V., Lynch, M., Vogelnest, L., Miller, G., Keeley, T., 2013. Response to long-distance relocation in Asian elephants (*Elephas maximus*): monitoring adrenocortical activity via serum, urine, and feces. Eur. J. Wildl. Res. 59, 655–664.

Ganswindt, A., Palme, R., Heistermann, M., Borragan, S., Hodges, J.K., 2003. Non-invasive assessment of adrenocortical function in the male African elephant (*Loxodonta africana*) and its relation to musth. Gen. Comp. Endocrinol. 134, 156–166.

Goymann, W., 2012. On the use of non-invasive hormone research in uncontrolled, natural environments: the problem with sex, diet, metabolic rate and the individual. Meth. Ecol. Evol. 3, 757–765.

Grandin, T., 1997. Assessment of stress during handling and transport. J. Anim. Sci. 75, 249–257.

Hing, S., Narayan, E., Thompson, R.A., Godfrey, S., 2014. A review of factors influencing the stress response in Australian marsupials. Cons. Physiol. 2, cou027.

Hogan, L.A., Johnston, S.D., Lisle, A.T., Keeley, T., Wong, P., Nicolson, V., Horsup, A.B., Janssen, T., Phillips, C.J.C., 2011. Behavioural and physiological responses of captive wombats (*Lasiorhinus latifrons*) to regular handling by humans. Appl. Anim. Behav. Sci. 134, 217–228.

Hogan, L.A., Lisle, A.T., Johnston, S.D., Robertson, H., 2012. Non-invasive assessment of stress in captive numbats, *Myrmecobius fasciatus* (Mammalia: Marsupialia), using faecal cortisol measurement. Gen. Comp. Endocrinol. 179, 376–383.

Hunter, R., 2010. Interspecies allometric scaling. In: Cunningham, F., Elliott, J., Lees, P. (Eds.), Comparative and Veterinary Pharmacology. Springer, Berlin Heidelberg, pp. 139–157.

Johnston, S.D., Booth, R.A., Pyne, M., Keeley, T., Mackie, J.T., Hulse, L., Ellis, W., 2013. Preliminary study of faecal cortisol and corticosterone as an index of acute cortisol secretion in the koala (*Phascolarctos cinereus*). Aust. Vet. J. 91, 534–537.

Keeley, T., O'Brien, J., Fanson, B., Masters, K., McGreevy, P., 2012. The reproductive cycle of the Tasmanian devil (*Sarcophilus harrisii*) and factors associated with reproductive success in captivity. Gen. Comp. Endocrinol. 176, 182–191.

Koolhaas, J.M., de Boer, S.F., Coppens, C.M., Buwalda, B., 2010. Neuroendocrinology of coping styles: towards understanding the biology of individual variation. Front. Neuroendocrinol. 31, 307–321.

Martin, I.K., McDonald, I.R., 1986. Adrenocortical functions in a macropodid marsupial *Thylogale billardierii*. J. Endocrinol. 110, 471–480.

McDonald, I., 1977. Adrenocortical functions in marsupials. In: Stonehouse, B., Gilmore, D. (Eds.), The Biology of Marsupials. Macmillan, London, pp. 345–377.

McEwen, B.S., 1998. Protective and damaging effects of stress mediators. N. Engl. J. Med. 338, 171–179.

McKenzie, S., Deane, E.M., 2005. Faecal corticosteroid levels as an indicator of well-being in the tammar wallaby, *Macropus eugenii*. Comp. Biochem. Physiol. A 140, 81–87.

Möstl, E., Palme, R., 2002. Hormones as indicators of stress. Domest. Anim. Endocrinol. 23, 67–74.

Möstl, E., Rettenbacher, S., Palme, R., 2005. Measurement of corticosterone metabolites in birds' droppings: an analytical approach. Ann. N. Y. Acad. Sci. 1046, 17–34.

Munn, A.J., Dawson, T.J., 2006. Forage fibre digestion, rates of feed passage and gut fill in juvenile and adult red kangaroos *Macropus rufus* Desmarest: why body size matters. J. Exp. Biol. 209, 1535–1547.

Munro, C., Stabenfeldt, G., 1985. Development of a cortisol enzyme-immunoassay in plasma. Clin. Chem. 31, 956-956.

Narayan, E.J., Webster, K., Nicolson, V., Mucci, A., Hero, J.-M., 2013. Non-invasive evaluation of physiological stress in an iconic Australian marsupial: the Koala (*Phascolarctos cinereus*). Gen. Comp. Endocrinol. 187, 39–47.

Narayan, E.J., Evans, N., Hero, J.-M., 2014. Monitoring physiological stress in semi-free ranging populations of an endangered Australian marsupial, the Greater Bilby (*Macrotis lagotis*). Eur. J. Wildl. Res. 60, 727–735.

Naylor, R., Richardson, S.J., McAllan, B.M., 2008. Boom and bust: a review of the physiology of the marsupial genus *Antechinus*. J. Comp. Physiol. B. 178, 545–562.

Oates, J.E., Bradshaw, F.J., Bradshaw, S.D., Stead-Richardson, E.J., Philippe, D.L., 2007. Reproduction and embryonic diapause in a marsupial: insights from captive female Honey possums, *Tarsipes rostratus* (Tarsipedidae). Gen. Comp. Endocrinol. 150, 445–461.

O'Rourke, N., Hatcher, L., Stepanski, E.J., 2005. A Step-by-step Approach to Using SAS for Univariate & Multivariate Statistics. SAS Institute.

Padgett, D.A., Glaser, R., 2003. How stress influences the immune response. Trends Immunol. 24, 444–448.

Palme, R., 2005. Measuring fecal steroids – guidelines for practical application. Ann. N. Y. Acad. Sci. 1046, 75–80.

Palme, R., 2012. Monitoring stress hormone metabolites as a useful, non-invasive tool for welfare assessment in farm animals. Anim. Welfare 21, 331–337.

Palme, R., Möstl, E., 1997. Measurement of cortisol metabolites in faeces of sheep as a parameter of cortisol concentration in blood. Z. Saugetierkd. (Germany).

Palme, R., Robia, C., Baumgartner, W., Möstl, E., 2000. Transport stress in cattle as reflected by an increase in faecal cortisol metabolite concentrations. Vet. Rec. 146, 108–109.

Palme, R., Rettenbacher, S., Touma, C., El-Bahr, S.M., Möstl, E., 2005. Stress hormones in mammals and birds - Comparative aspects regarding metabolism, excretion, and noninvasive measurement in fecal samples. Ann. N. Y. Acad. Sci. 1046, 162–171.

Palme, R., Touma, C., Arias, N., Dominchin, M.F., Lepschy, M., 2013. Steroid extraction: get the best out of faecal samples. Wien. Tierärztl. Monat. – Vet. Med. Austria 100, 238–246.

Patel, L., Clayton, P., 1999. Clinical usefulness of the low dose ACTH test. J. Endocrinol. Investig. 22, 401–404.

Sapolsky, R.M., 2002. Endocrinology of the stress-response. In: Becker, J.B., Breedlove, S.M. (Eds.), Behavioral Endocrinology, 2nd ed. MIT Press, Cambridge, Massachusetts, pp. 409–450.

Sapolsky, R.M., Romero, L.M., Munck, A.U., 2000. How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. Endocr. Rev. 21, 55–89.

Selwood, L., Cui, S., 2006. Establishing long-term colonies of marsupials to provide models for studying developmental mechanisms and their application to fertility control. Aust. J. Zool. 54, 197–209.

Sheean, V.A., Manning, A.D., Lindenmayer, D.B., 2012. An assessment of scientific approaches towards species relocations in Australia. Austral Ecol. 37, 204–215.

Sheriff, M., Dantzer, B., Delehanty, B., Palme, R., Boonstra, R., 2011. Measuring stress in wildlife: techniques for quantifying glucocorticoids. Oecologia 166, 869–887.

Short, J., Smith, A., 1994. Mammal decline and recovery in Australia. J. Mammal. 75, 288–297.

Touma, C., Palme, R., 2005. Measuring fecal glucocorticoid metabolites in mammals and birds: the importance of validation. Ann. N. Y. Acad. Sci. 1046, 54–74.

Touma, C., Sachser, N., Mostl, E., Palme, R., 2003. Effects of sex and time of day on metabolism and excretion of corticosterone in urine and feces of mice. Gen. Comp. Endocrinol. 130, 267–278.

Touma, C., Palme, R., Sachser, N., 2004. Analyzing corticosterone metabolites in fecal samples of mice: a noninvasive technique to monitor stress hormones. Horm. Behav. 45, 10–22.

Watson, R., Munro, C., Edwards, K.L., Norton, V., Brown, J.L., Walker, S.L., 2013. Development of a versatile enzyme immunoassay for non-invasive assessment of glucocorticoid metabolites in a diversity of taxonomic species. Gen. Comp. Endocrinol. 186, 16–24.

Whirledge, S., Cidlowski, J.A., 2013. A role for glucocorticoids in stress-impaired reproduction: beyond the hypothalamus and pituitary. Endocrinology 154, 4450–4468.

Wielebnowski, N., Watters, J., 2007. Applying fecal endocrine monitoring to conservation and behavior studies of wild mammals: important considerations and preliminary tests. Israel J. Ecol. Evol. 53, 439–460.

Woinarski, J., Burbidge, A., Harrison, P., 2014. Action plan for Australian Mammals 2012. CSIRO.

Young, K.M., Walker, S.L., Lanthier, C., Waddell, W.T., Monfort, S.L., Brown, J.L., 2004. Noninvasive monitoring of adrenocortical activity in carnivores by fecal glucocorticold analyses. Gen. Comp. Endocrinol. 137, 148–165.

Electronic Supplementary Material

Results and assay recommendations for individual species

Note: Species ordered according to Table 1

Numbat (Mfa)

Four numbats (2 females, 2 males) underwent an ACTH challenge (3 IU/kg, Synacthen). Samples were collected twice daily from the time the ACTH injection was given to three days later, and once per day the remainder of the time (i.e., up to 7 days post-injection). Three assays (37e, AAcort, and Cortisol) detected peaks in half the individuals, but the signal-to-noise ratio was fairly low for all assays. This may be in part due to a relatively low ACTH dose. However, it is also possible that the numbat extracts had degraded in storage. For this species, only faecal extracts were available and these had been stored (first at -80 ^{o}C, then at -20 ^{o}C) for a total of 2 years, the oldest of any extracts used in this study.

Tasmanian devil (Sha)

Four devils (2 females, 2 males) underwent an ACTH challenge (4 IU/kg, Synacthen). Samples were collected daily prior to the ACTH challenge. After the injection, all observed faecal samples were collected for two individuals, and daily samples were collected for the other two. Devils generally defecate only 1-2 times per day. Assay performance was fairly similar across assays, but 72a was the least successful at detecting peaks.

Bilby (Mla)

Four bilbies (2 females, 2 males) were given an ACTH injection (3 IU/kg, Synacthen) and samples were collected once daily. Only one assay (72a) detected a peak for one individual (Mla-3), but most assays failed to detect a peak following ACTH injection. This may be in part due to the relatively low ACTH dose and/or insufficient sampling frequency. Consequently, it was not possible to assess assay performance, but that is *not* to say that these assays are not suitable for monitoring adrenal activity in bilbies. More frequent sampling and/or larger stressors may reveal the suitability of one or more of these assays. Narayan et al. (2014) used the Cortisol EIA (R4866) and reported that FGMs increased between 24-72 h post-ACTH challenge in 3 bilbies (1 male and 2 female). The male bilby did not exhibit a strong response to the ACTH challenge. Narayan et al. (2014) also echoed the points that dose response experiments may be needed to determine the appropriate dose of ACTH for stimulation of the adrenocortical stress axis in this species.

Koala (Pci)

Four koalas (2 females, 2 males) were given an ACTH injection (7 IU/kg, Synacthen) and samples were collected once daily. Both 72a and AAcort detected peaks in half the individuals, but AAcort had much higher sensitivity. The other three assays only detected a peak in one individual. Therefore, we would recommend AAcort as the preferred assay for this species. Narayan et al. (2013) used the Cortisol EIA (R4866) to assay the same samples (but different extracts) and discovered that the FGM EIA detected a rise in FCM in koala faecal extracts within 24h (for females) and 48 h (for males) since the ACTH challenge. Narayan et al. (2013) further discussed that complexity of faecal cortisol metabolite detection in the koala is with the prolonged excretory lag-time of FCM (delayed second peak for some koalas up to 9 days) due to its excessively long gut system.

Northern hairy-nosed wombat (Lkr)

One juvenile female was hand-reared before being moved to an outdoor exhibit. Sampling started just prior to her move to the outdoor exhibit. Following the move, peaks were detected on 3 assays and peak magnitudes were similar (Table 4). However, Assay 72a showed the most expected pattern, with FGM concentrations low prior to the move, increasing just after the move, and then decreasing after ~5 days. For the other two assays (AAcort and Cs6), FGM concentrations were quite high prior to the move, peaked just after the move, and then fell fairly quickly to levels much lower than the initial samples. Therefore, we would recommend using assay 72a, with the caveat that this validation only included a single juvenile female.

Southern hairy-nosed wombat (Lla)

Three females were exposed to various husbandry events that may have been potentially stressful (e.g., veterinary exam with anaesthesia or re-fitting a radio-telemetry collar). Samples were collected approximately twice per week before and after the event. The Cortisol assay was the most successful, detecting post-event peaks in two of the three individuals. However, the low sampling frequency made it difficult to assess assay performance since some peaks may have been missed.

Mountain pygmy-possums (Bpa)

Eight mountain pygmy-possums (4 females, 4 males) were monitored surrounding breeding introductions. Once estrous was detected in females (based on behaviour, pouch condition and vaginal cytology) males were introduced to the female's home enclosure for several days. There were very few peaks detected for this species (Table 4). However, only six samples were collected from each individual, and sample collection was infrequent (every few days, with 2 samples each prior to, during, and post introduction). Therefore, it is possible that acute peaks were missed. Additionally, since they were exposed to handling and transfer between enclosures fairly regularly, this may not have been a significant stressor.

Interestingly, we saw a significant peak in 3 of the 4 females just *prior* to the introduction. Based on vaginal cytology, this peak would have occurred just prior to ovulation, and glucocorticoids are known to peak just prior to ovulation in many vertebrates (Fanson and Parrott, 2015). Based on these peaks, the Cortisol assay seemed to be the most sensitive and have the least variability in baseline, but female profiles for all assays were quite similar and further validation is necessary.

Yellow-bellied glider (Pau)

Two male gliders were transferred between institutions, and sample collection started immediately when they arrived at the new institution. All 5 assays detected a peak within the first two days after arriving, and the peak was much stronger in Glider2. These results indicate that for large stressors, any of these assays would be capable of detecting a response, but their sensitivity to smaller stressors is unknown. The Cortisol assay showed the strongest increase (Table 4), but this was largely driven by Glider2. When considering the profiles for both individuals, we would recommend Assay 37e, which captured a much stronger initial response to transfer for Glider1 than the rest of the samples.

Long-nosed potoroo (Ptr)

One male potoroo was transferred between institutions, and sample collection started immediately when he arrived at the new institution. Assay 72a detected the greatest increase in FGM in the first three days post-arrival. Although the Cortisol assay also

detected a peak in the first few days, the peak was much smaller and there was considerably more noise in the baseline. Therefore, we would recommend Assay 72a for long-nosed potoroos.

Woylie (Bpe & Bpe2)

Three woylies (1 female, 2 males) were exposed to an ACTH challenge. All three individuals were initially given a dose of 6 IU/kg (Synacthen) and samples were collected once daily. No clear peaks were detected in any of the individuals, so the ACTH challenge was repeated for the female (Bpe2) using a dose of 9 IU/kg and collecting samples three-times daily. The female was also moved to a new enclosure when the ACTH was administered. Using this validation design, clear peaks were detected on 3 of the 4 assays (Table 4; also see ESM Fig 2). Based on the results from this final ACTH challenge, we recommend Assay AAcort because it had the strongest signal-to-noise ratio, but Assay 72a would be a suitable alternative. The results from this species highlight the importance of using a sufficient ACTH dose and sampling adequately, especially post-stressor.

Southern bettong (Bga)

Two male bettongs were transferred between institutions, and sample collection started immediately when they arrived at the new institution. All 5 assays showed the expected trend, with FGM concentrations being elevated shortly after the transfer and then decreasing between days 3-5 post-transfer. The difference in time-to-peak between the two individuals (Bettong1 = day 4; Bettong2 = day 1; ESM Table 1) may reflect differences in perception of the transfer or differences in body condition. Bettong2 died of lumpy jaw (Oral necrobacillosis) towards the end of the quarantine period, which may be related to the large FGM peak ~day 9-11 (especially in Cortisol). Assay 72a yielded the greatest fold-increase as well as z-score (Table 4). We would recommend using Assay 72a, but Assays Cs6 and AAcort would offer suitable (though less sensitive) alternatives.

Eastern grey kangaroo (Mgi)

Four eastern grey kangaroos (3 males, 1 female) underwent an ACTH challenge (10 IU/kg, Corticotropin). Samples were collected daily prior to the ACTH challenge, after which all faecal samples were collected. Four days after the ACTH challenge, sampling was reduced to once daily. Assay 37e was the most consistent at detecting post-ACTH peaks, and demonstrated moderate signal-to-noise ratio. Therefore, we recommend 37e as the preferred assay for eastern grey kangaroos, with 72a and Cortisol as suitable alternatives.

Western grey kangaroo (Mfu)

Four wild-caught western grey kangaroos (1 female, 3 males) were given an ACTH injection (10 IU/kg, Synacthen Depot) after acclimating to holding pens. Samples were collected every four hours for five days and then twice daily for a further five days. Both 72a and AAcort detected a peak in 3 of the 4 animals, but 72a had a stronger signal-to-noise ratio. The other two assays (37e and Cs6) only detected a peak on one individual. Therefore, we recommend 72a as the preferred assay.

References

Fanson, K.V., Parrott, M.L., 2015. The value of eutherian–marsupial comparisons for understanding the function of glucocorticoids in female mammal reproduction. Horm. Behav. doi:10.1016/j.yhbeh.2015.05.012

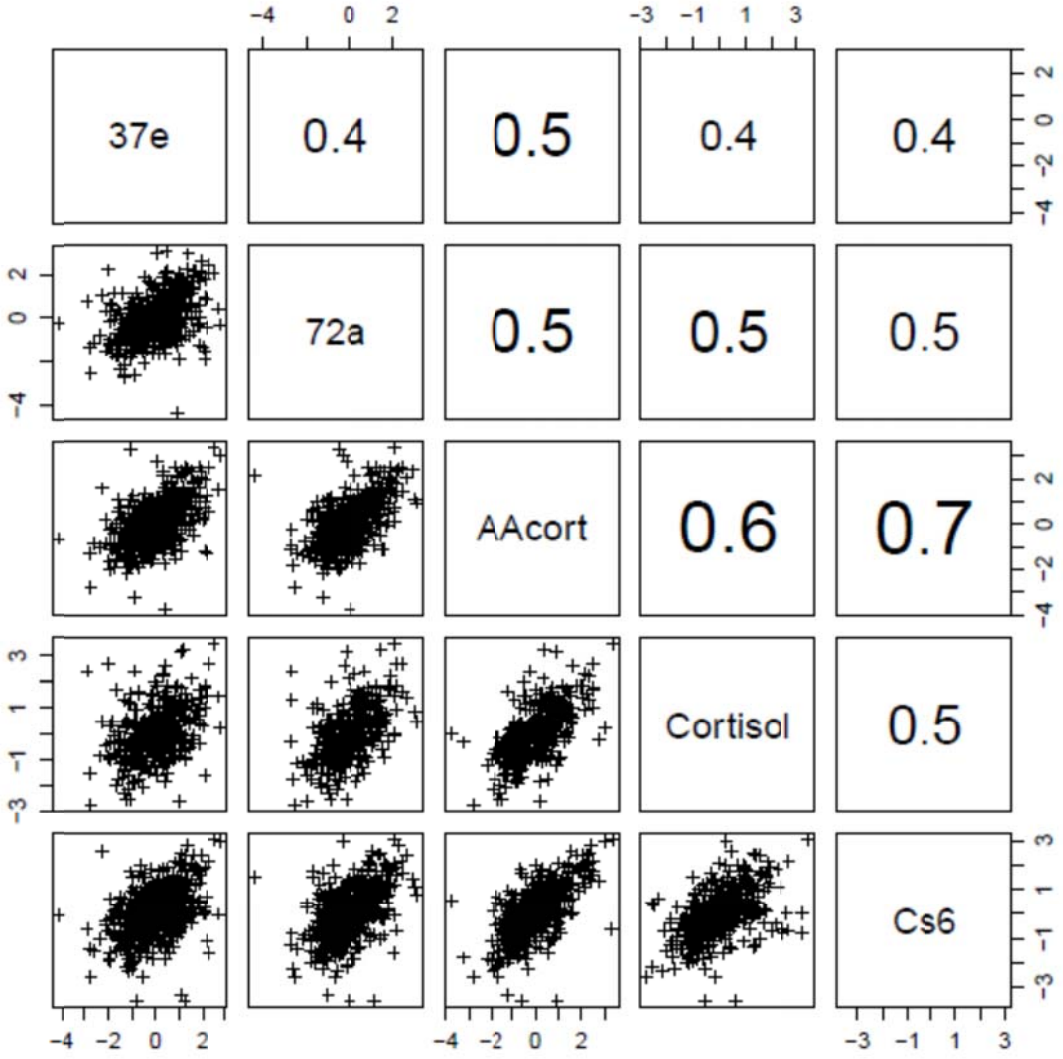
Narayan, E.J., Evans, N., Hero, J.-M., 2014. Monitoring physiological stress in semi-free ranging populations of an endangered Australian marsupial, the Greater Bilby (*Macrotis lagotis*). Eur. J. Wildl. Res. 60, 727-735.

Narayan, E.J., Webster, K., Nicolson, V., Mucci, A., Hero, J.-M., 2013. Non-invasive evaluation of physiological stress in an iconic Australian marsupial: The Koala (*Phascolarctos cinereus*). Gen. Comp. Endocrinol. 187, 39-47.

Supplementary Table 1. Delay in time (days) to maximum post-stressor value (within 4.5 days). Top = species mean; bottom = range (note that if there is only 1 individual, no range is indicated).

spp	37e	72a	AAcort	Cortisol	Cs6
Dasyuromorphia					
Mfa	1.4	2.7	1.9	1.7	2.7
	(0.9 - 1.9)	(1.9 - 2.9)	(0.9 - 2.9)	(0.9 - 2.9)	(0.9 - 3.9)
Sha	1.7	1.2	0.9	1.7	0.9
	(0.9 - 3.9)	(0.9 - 1.9)	(0.9 - 0.9)	(0.9 - 3.9)	(0.9 - 0.9)
Peramelemorphia					
Mla	1.2	2	1.5	1.8	2.2
	(1 - 2)	(2 - 2)	(1 - 2)	(1 - 4)	(2 - 3)
Diprotodontia					
Pci	2	2	2.2	1.8	2.2
	(1 - 3)	(1 - 3)	(1 - 4)	(1 - 4)	(1 - 4)
Lkr	3	3	2	-	1
Lla	2	2	3	4	2
	(1 - 4)	(1 - 4)	(1 - 4)	(4 - 4)	(1 - 4)
Bpa	1.7	1.7	1.3	1.6	1.7
	(0.7 - 3.7)	(0.7 - 3.7)	(0.7 - 3.7)	(0.7 - 3.7)	(0.7 - 3.7)
Pau	1	1.5	1.5	1.5	1.5
	(1 - 1)	(1 - 2)	(1 - 2)	(1 - 2)	(1 - 2)
Ptr	2	3	2	2	1
Bpe	1.3	1	1.9	-	2.6
	(0.9 - 1.9)	(0.9 - 1.3)	(1.9 - 1.9)		(0.9 - 3.9)
Bpe2	1.3	1.3	1.9	-	3
Bga	3.5	2.5	2.5	1	2.5
	(3 - 4)	(1 - 4)	(1 - 4)	(1 - 1)	(1 - 4)
Mgi	0.8	1.4	0.6	1.4	0.9
	(0.4 - 1.4)	(0.7 - 2.2)	(0.4 - 0.9)	(0.6 - 2.1)	(0.3 - 2.2)
Mfu	0.8	0.8	0.8	-	0.9
	(0.6 - 1)	(0.6 - 1)	(0.6 - 1)		(0.6 - 1.6)

Supplementary Figure 1. Pairwise comparisons among assays. Data were centered for each individual, assay, and species (see text for details). Therefore, axes represent standardized z-scores. The graphs below the diagonal are scatterplots showing the relationship between 2 assays. The numbers above the diagonal represent the correlation coefficient, and the size of the number reflects the p-value (lower p-values are shown in larger font).



Supplementary Figure 2. Profiles for a female woylie that received two different ACTH doses. Both profiles represent results obtained using the AAcort assay. A) The female received 6 IU/kg ACTH and samples were collected once daily. B) The female (Bpe2) received 9 IU/kg ACTH and samples were collected three times per day. Vertical grey bar indicates when the ACTH was administered.