

Experimental evidence that novel land management interventions inspired by history enhance biodiversity

Robert W. Hawkes^{1,2*}, Jennifer Smart^{2,3}, Andy Brown⁴, Helen Jones², Steve A. Lane⁵, Colin Lucas⁶,
James McGill⁷, Nick Owens⁸, Amanda Ratier Backes⁹, Jon R. Webb⁴, Doreen Wells¹⁰, and Paul M.
Dolman¹

¹ School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

² RSPB Centre for Conservation Science, The Lodge, Sandy, Bedfordshire, SG19 2DL, UK

³ School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

⁴ Natural England, Suite D, Unex House, Bourges Boulevard, Peterborough, PE1 1NG, UK

⁵ 5 Lancaster Way, East Winch, King's Lynn, PE32 1NY, UK

⁶ 49 Mill Road, Beccles, NR34 9UT, UK

⁷ 13 Cresswell Avenue, Taunton, TA2 6LS, UK

⁸ 22 Springfield Close, Weybourne, Holt, NR25 7TB, UK

⁹ Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg,
Halle (Saale), Germany

¹⁰ 1 Cobden Terrace, Great Yarmouth, NR30 1BT, UK

* Corresponding author

E-mail address: robert.hawkes@uea.ac.uk

Abstract

1. To address biodiversity declines within semi-natural habitats, land-management must cater for diverse taxonomic groups. Integrating our understanding of the ecological requirements of priority (rare, scarce or threatened) species through ‘biodiversity auditing’, with that of the intensity and complexity of historical land-use, encourages novel forms of management. Experimental confirmation is needed to establish whether this enhances biodiversity conservation relative to routine management.
2. Biodiversity auditing and historical land-use of dry-open terrestrial habitats in Breckland (Eastern England) both encourage management incorporating ground-disturbance and spatio-temporal variability. To test biodiversity conservation outcomes, we developed 40 4-ha management complexes over three successive winters, of which 20 were shallow-cultivated (rotovation) and 20 deep-cultivated (ploughing), stratified across 3,850-ha of closed-sward dry grassland and lowland heathland (collectively ‘dry grassland’). Complexes comprised four 1-ha sub-treatments: repeat-cultivation, first-time-cultivation, one-year-old fallow and two-year-old fallow. We examined responses of vascular plants; spiders; true bugs; ground, rove and ‘other’ beetles; bees and wasps; ants; and true flies on treatment complexes and 21 4-ha untreated controls. Sampling gave 132,251 invertebrates from 877 species and 28,846 plant observations from 167 species.
3. Resampling and rarefaction analyses showed shallow- and deep-cultivation both doubled priority species richness (pooling sub-treatments within complexes) compared to controls. Priority spider, ground beetle, other beetle, and true bug richness were greater on both treatments than controls. Responses were strongest for those priority dry-open-habitat associated invertebrates initially predicted (by biodiversity auditing) to benefit from heavy physical-disturbance.

4. Assemblage composition (pooling non-priority and priority species) varied between sub-treatments for plants, ants, true bugs, spiders, ground, rove and other beetles; but only one-year-old fallowed deep-cultivation increased priority richness across multiple taxa.
5. Treatments produced similar biodiversity responses across various dry grassland ‘habitats’ that differed in plant composition, allowing simplified management guidance.
6. *Synthesis and applications.* Our landscape-scale experiment confirmed the considerable biodiversity value of interventions inspired by history and informed by systematic multi-taxa analysis of ecological requirements across priority biota. Since assemblage composition varied between sub-treatments, providing heterogeneity in management will support the widest suite of species. Crucially, the intended recipients responded most strongly, suggesting biodiversity audits could successfully inform interventions within other systems.

KEYWORDS biodiversity audit, cultural landscape, dry grassland, ground-disturbance, landscape-scale conservation, lowland heathland, multi-taxa, semi-natural habitat

1. INTRODUCTION

Across Europe, conservation tends to focus on semi-natural habitats shaped by a long history of human management (Ratcliffe 1977; EC 1992), yet priority plants (Hülber *et al.* 2017) and invertebrates (Seibold *et al.* 2019) continue to disappear from such areas. Within surviving habitat remnants, land management often mimics elements of historic (pre-industrial c 1200 – 1750) practices on the assumption that this will support assemblages that persisted through human activity (Wright, Lake & Dolman 2012; Fuller, Williamson, Barnes & Dolman 2017), or is justified by reference to the ecology of a taxonomically-biased species-subset (Clark & May 2002; Griffiths & Dos Santos 2012). Many interpretations of ‘traditional’ management are incomplete and potentially suboptimal for threatened biodiversity, but the promotion of alternative novel approaches requires supporting evidence. A new emphasis on ‘re-wilding’ (Pettorelli *et al.* 2018) and a progressive shift from biodiversity conservation for the intrinsic value of species, to ecosystems and the goods and services they provide (Mace 2014), further increase the need for approaches which can quantify and predict biodiversity responses to landscape-scale interventions. Where biodiversity is well-characterised and autecological knowledge is strong (e.g. in much of Europe), the biodiversity audit approach - a bio-regional process where biodiversity records are collated and priority species with shared autecological requirements are subsequently grouped into cross-taxa ‘management guilds’ - provides an objective way of informing and optimising the conservation benefits of management interventions (Dolman, Panter & Mossman 2012). However, whilst recent audits support the importance of historical management to priority biota (Dolman *et al.* 2012), there is a pressing need for experimental confirmation involving multiple taxa.

Crucially, perceptions and implementation of ‘traditional management’ tend to be simplified and homogenised (e.g. the notion of conservation grazing, Fuller *et al.* 2017). Historical management was, in reality, characterised by repeated biomass removal and physical-disturbance through complex multi-layered land-use that often overlapped grazing with other forms of resource harvest and varied spatio-temporally within sites and across landscapes (Linnell, Kaczensky, Wotschikowsky,

Lescureux & Boitani 2015; Fuller *et al.* 2017). Synthesising autecological knowledge through biodiversity audits (Dolman *et al.* 2012) and a detailed understanding of historic land-use complexities (Linnell *et al.* 2015; Fuller *et al.* 2017) both inspire novel interventions (hereafter, ‘enhanced management’) that emphasise physical-disturbance, grazing, nutrient removal, spatio-temporal variability, early-successional habitats and structural complexity (Fuller *et al.* 2017). This might involve near-accurate replication of specific pre-industrial practices (e.g. coppicing, Merckx *et al.* 2012) or the use of wild or domestic herbivores to create and maintain dynamic mosaics (consistent with some principles of rewilding, Van Klink & WallisDeVries 2018); but in other circumstances it may be appropriate to adopt new approaches that provide the resources needed by the widest range of species, particularly priority species not helped by routine management.

Despite recent calls for strategies that deploy novel forms of enhanced management (Linnell *et al.* 2015; Fuller *et al.* 2017), this approach is untested. First, it is unclear whether target priority species are able to colonise newly-established suitable habitats (Thomas 1994) and whether the benefits of management are offset by negative impacts on species intolerant of the intervention. Second, because knowledge of land-use history is not exact, and modern techniques offer interventions that differ from historical methods, experiments are needed to establish whether treatment detail matters. In many cases accelerated succession (from increased rates of nitrogen-deposition, Tipping *et al.* 2019) may reduce the duration and pattern of colonisation of suitable micro-habitats, such that more severe longer-lived interventions may be beneficial (Härdtle, Niemeyer, Niemeyer, Assmann & Fottner 2006; Pedley, Franco, Pankhurst & Dolman 2013). Third, conservation advocacy would be streamlined if evidence supports consistent interventions across habitats that share similar ecological processes despite differing in plant composition. Last, most tests of intervention efficacy within semi-natural habitats focus either on vegetation structure as a proxy for biodiversity, or on single species or a limited subset of taxa (e.g. Lepidoptera, Goodenough & Sharp 2016; or birds, Żmihorski, Pärt, Gustafson & Berg 2016). Given that semi-natural habitats are

109 especially valued for their diverse assemblages (Ratcliffe 1977), such studies are unlikely to be good
110 substitutes for robust multi-taxa experiments.

111 Here, we test the multi-taxa consequences of enhanced management interventions across
112 an extensive, semi-natural mosaic (3,850-ha) of calcareous and acidic dry grasslands of varying age
113 and long-established lowland heathland (hereafter collectively 'grassland'). Grassland conservation
114 practices emphasise the role of grazing (Wells 1969; Bakker, De Bie, Dallinga, Tjaden & De Vries
115 1983), but the needs of many priority species may be better met by temporally- and spatially-
116 dynamic physical-disturbance (e.g. Denton 2013). A biodiversity audit of the Breckland region of
117 Eastern England (characterised by low-rainfall, sandy soils and internationally important grassland
118 habitats, Dolman *et al.* 2012), suggests 61% of the 629 priority species associated with dry-open
119 habitats require heavy physical-disturbance, fallows, or habitat juxtaposition. Such management was
120 characteristic of pre-industrial landscapes generally (Fuller *et al.* 2017) and particularly in Breckland,
121 where grassland habitats were disturbed by infrequent cultivation of long-rotation fallows, rabbit
122 farming and resource extraction (Bailey 1989; Dolman & Sutherland 1991). This combination of
123 autecology and history justifies creating mosaics of cultivations that vary in disturbance intensity and
124 fallow age. However, because previous ground-disturbance experiments have examined
125 homogenous, even-aged interventions (e.g. Pywell *et al.* 2007; Pedley *et al.* 2013; Hawkes *et al.*
126 2019b), the multi-taxa consequences of enhanced management are unclear.

127 To examine the effects of such management on grassland biodiversity we conducted a well-
128 replicated, landscape-scale experiment. We tested two contrasting ground-disturbance treatments
129 that provided structural complexity: shallow-cultivation with a rotatory rotavator and deep-
130 cultivation with an agricultural plough that, respectively, may disrupt vegetation less severely, or
131 more drastically, than historic cultivation by oxen-drawn plough. Treatments were built-up over
132 three years to create complexes comprising sub-treatments that varied in time since cultivation and
133 disturbance frequency (single or repeated cultivation) (Fig. 1). We quantified responses across nine

134 taxonomic groups, separately for non-priority and priority species, comparing treatments to areas of
135 grassland managed with light grazing and limited or no ground disturbance. To test the efficacy of
136 treatments based on autecological synthesis and historic land use, we compared responses to
137 shallow- and deep-cultivation at the 'complex level' (pooling across sub-treatments) examining
138 species richness (hereafter 'richness') and composition (quantified as the percentage of the species-
139 pool supported within, or unique to, each treatment) in contrast to controls. To test whether
140 biodiversity auditing usefully predicted responses to treatment interventions, we examined whether
141 those species associated with dry-open habitats whose autecology indicates an association with
142 physical-disturbance (the intended recipients) responded more strongly to treatment. Last, to refine
143 management recommendations, we examined responses to differing sub-treatments within
144 complexes, and whether efficacy differed with grassland type and composition.

2. MATERIALS AND METHODS

2.1 Study area

Our ground-disturbance interventions took place across the grasslands of the Stanford Military Training Area (STANTA) (52°51'N, 0°76'E, 3500-ha), Bridgham Heath (52°44'N, 0°83'E, 150-ha) and Brettenham Heath (52°43'N, 0°83'E, 200-ha) in Breckland. All three sites are situated close together (maximum distance 3.5 km, see Fig. S1) and managed similarly by light sheep grazing (c.1 ewe per ha) and occasional bracken *Pteridium aquilinum* control (for details, see Hawkes, Smart, Brown, Jones & Dolman 2019a).

2.2 Experimental treatments

In early 2015, 40 2-ha (100 x 200 m) ground-disturbance plots (20 shallow-cultivated using a rotary rotavator; 20 deep-cultivated using an agricultural plough) and 21 4-ha (\approx 200 x 200 m) control plots were established in grassland mostly excluding, but sometimes close to, scattered trees or scrub (for details, see Hawkes *et al.* 2019a). Treatments were repeated during each of the next two winters, again cultivating 2-ha (100 x 200 m), but with half overlapping a central repeatedly-treated sub-plot and half first-time-cultivation, to form a 4-ha treatment complex by 2017. Each complex comprised four 1-ha subplots: one-year-old fallow, two-year-old fallow, first-time-cultivated, and repeatedly-cultivated (Fig. 1). Less intensive vegetation disturbance such as cutting and removal was not considered, as structural effects are ephemeral (Dolman & Sutherland 1994) with little benefit for priority dry-open habitat species (Pedley *et al.* 2013).

Studies in the inner part of STANTA were precluded by potential unexploded ordnance (Fig. S1), but otherwise, treatments and controls were randomly distributed across the study area, stratified across four non-randomly distributed grassland strata (following Hawkes *et al.* 2019a; Hawkes *et al.* 2019b), defined by soil type, age since cultivation and plant composition as: calcareous grassland of any age (hereafter 'calcareous grassland'); young grassland (\leq 110 years old);

intermediate-aged grassland (111-167 years old); and ancient-acid grassland/lowland heath (most ≥ 168 years old and variably dominated by *Deschampsia flexuosa* and *Calluna vulgaris*) (see Table S1 for details).

2.3 Responses to treatment

Responses to treatment were assessed in 2017. Invertebrates were sampled across all 40 treatment complexes and 21 control plots, and vascular plants (hereafter 'plants') across 32 complexes (16 shallow-cultivated, 16 deep-cultivated) and 16 controls (randomly selected, constrained to strata). Invertebrate trapping intensity across each of the four 1-ha sub-treatments per complex and one central 1-ha plot within each control (hereafter 'samples') was consistent (see below). Greater sampling intensity per treatment complex than per control was accounted for subsequently by rarefaction (see below).

Pitfall traps sampled seven invertebrate groups [spiders (Araneae); ground beetles (Carabidae); rove beetles (Staphylinidae); other beetles (Coleoptera, excluding Carabidae and Staphylinidae); true bugs (Hemiptera, excluding aphids); ants (Formicidae); bees and wasps (Aculeata, excluding ants)] and pan traps sampled three [true bugs; bees and wasps; true flies (Diptera, resolved for: Asilidae, Rhagionidae, Stratiomyidae, Syrphidae, Tabanidae, Therevidae, Tipulidae)]. Each sample comprised: (i) six pitfall traps (each 11 cm deep, 8 cm diameter, covered by 12 x 12 mm mesh, with 50 ml of 33% propylene glycol) deployed 15 m apart in a central 15 x 30 m grid, for seven consecutive days, separately in each of May/June, July/August, and September; and (ii) four yellow pan traps (each 4 cm deep, 15 cm diameter, set at ground level, covered with 30 x 40 mm mesh, with 150 ml of water with a few drops of unscented detergent) deployed once in a central 15 x 15 m grid, for three consecutive days, between 1 July and 26 August. If efficiency was less than half the maximum trap-days per plot, the array was repeated, after which 96% of pitfall trap and 94% pan trap deployments were active for the whole exposure period (for details see

Appendix S1 of supplementary materials). Plant incidence was sampled between 10 April and 7 July, from 16 quadrats (1 x 1 m) distributed evenly (11-14 m apart) along two parallel 100 m transects (30-33 m apart), giving frequency (0-16) per species. Data were pooled across pitfall-months and sampling methods (pitfalls, pan traps, quadrats) giving one composite sample (n=21 control plots; and n=160 sub-treatment plots, nested within cultivated complexes).

Most sampled taxa were identified to species level, the few unresolved plants (0.3%), spiders (<0.1%), ground, rove and other beetles (<0.1%, 3.5%, 1.0%), true bugs (3.7%), and bees and wasps (0.9%) were not considered further. Species were considered as conservation priorities when classified as either: Threatened (IUCN Critically Endangered, Endangered and Vulnerable) or Near Threatened in Great Britain, or Nationally Rare, Nationally Scarce, or earlier designations of Red Data Book or Nationally Notable (see Table S3 for sources); remaining species were considered 'non-priority'.

2.4 Analysing richness and composition at the complex level

For each taxonomic group (separately for non-priority and priority species), we used sample completeness curves (derived from sample-based rarefactions, re-scaled to the number of individuals, using the Mao Tau function) from the package 'iNEXT' (Chao *et al.* 2014) to estimate sample coverage - a measure of sampling efficiency - at the observed sample size for each sub-treatment category, complex category (shallow- and deep-cultivation, pooling across sub-treatment), and controls. Next, cumulative richness (pooled across all nine taxa and per taxon, separately for non-priority and priority species) of both treatment complex categories and controls (hereafter, collectively 'regimes') was examined by sample-based rarefaction. As the number of individuals sampled differed among regimes, following Chao *et al.* (2014) we compared richness estimates at a base sample size (hereafter 'BSS') set as the smaller of: twice that of the regime with the smallest sample size, or the regime with the largest sample size. For comparison, each regime

was subsequently extrapolated (observed richness < BSS) or rarefied (observed richness > BSS) to the BSS. As extrapolation is unreliable beyond double the original sample size (Chao *et al.* 2014), setting the BSS to the largest sample size was inappropriate. Consistent with other studies (e.g. Schall *et al.* 2018), we also set the BSS to the smallest sample size (classical rarefaction) to ensure findings are robust. Richness estimates were considered to differ between regimes when pairwise 95% CIs obtained from 200 bootstrapping replications, did not overlap (Chao *et al.* 2014); following convention, no post-hoc correction for multiple comparisons was applied. To avoid unreliable extrapolation we did not analyse any taxonomic groups where the number of observed species, from any of the three regimes, was less than three (judged separately for non-priority and priority species). Last, because the overall (cross-taxa) non-priority and priority comparisons considered eight complexes (four deep-cultivated, four shallow-cultivated) and five controls that lacked plant data, we tested whether removing these samples entirely from both analyses altered inference. All analyses were carried out in R (R Core Team 2015).

Irrespective of relative richness, treatment complexes and controls may support distinctive assemblages, or unique species not recorded in any other regime. We quantified the percentage of the species-pool supported within, and the percentage unique to, each regime, separately for non-priority and priority species, for the overall assemblage and each taxon, using Euler diagrams (in package 'Eulerr', Larsson 2019). To simultaneously consider plant and invertebrate samples, we omitted the five controls and eight complexes sampled only for invertebrates. To control for the greater level of sampling effort in the treatment complexes than controls, we undertook 200 iterations, each drawing all 16 control samples and 16 samples from each treatment regime (of 64 available per treatment regime, selecting independently of complex-identity), drawing one per (four) sub-treatment x (four) strata combination per iteration. Although every iteration resampled the same 16 control plots, their percentage relative to the species-pool depended on the sub-treatments drawn in that iteration. For each metric (percentage of species supported within or unique to each regime), regimes were compared through pair-wise comparisons with Bonferroni correction (i.e. at

least 199 of the 200 iterations showed larger values for one regime). Note, because multiple sub-treatment samples could be selected from the same complex (mean 11 ± 1 SD separate complexes per treatment, per iteration), analyses were slightly biased in favour of the controls which always resampled 16 distinct locations (thus potentially having greater gamma diversity).

2.5 Analysing responses of habitat and management guilds

To establish whether the treatment complexes increased the richness of the intended recipients, we used two existing autecological classifications. First, using the online tool 'Pantheon' (Heaver *et al.* 2017) we classified non-priority and priority invertebrate species associated with dry-open habitats onto a composite ecological gradient of increasing physical-disturbance intensity (hereafter 'habitat guilds'): from 'tall swards and scrub', through 'short swards without exposed sand' (hereafter, 'short swards'), to 'short sward with exposed sand' (hereafter, 'short swards and bare ground'). Second, we used an earlier biodiversity audit (Dolman *et al.* 2012) to classify the same dry-open priority invertebrate species into management guilds requiring 'no', 'light', or 'heavy' ground-disturbance and either 'no/light' or 'heavy' grazing. Species with unknown or undifferentiated structural or management requirements and five species considered by Dolman *et al.* (2012) to be associated with wet/shaded grassland, were excluded (Table S2). Although priority plants were also classified by biodiversity auditing, analysis was restricted to invertebrates for consistency with Pantheon habitat guilds. For each guild, we compared overall invertebrate richness between regimes using rarefaction, separately for non-priority (habitat guilds) and priority (habitat or management guilds) species. As grazing was light across the study landscape, we predicted that species associated with heavy ground-disturbance would respond more strongly to the treatment, irrespective of their grazing classification.

2.6 Analysing richness and composition at the sub-treatment level

271 Richness was compared between sub-treatments (nine levels: four shallow-cultivated sub-
 272 treatments, four deep-cultivated sub-treatments, plus controls) using Generalised Linear Mixed
 273 Models (GLMMs), separately for non-priority and priority species within each taxonomic group
 274 (though omitting any excluded from the complex rarefactions). Strata was included as a fixed effect,
 275 but with 'young' and 'calcareous' grasslands *a priori* merged as 'calcareous grassland' owing to the
 276 similarity of their vegetation structure (Hawkes *et al.* 2019b) and plant composition (Table S1), thus
 277 reducing model complexity. To determine whether treatment efficacy varied between strata, we
 278 examined a sub-treatment*strata interaction. Plot (for controls) or complex (sub-treatments)
 279 identity was included as a random effect to control for non-independence of sub-treatment samples
 280 within complexes. To account for slight variation in trap success, the total number of pitfall trap days
 281 (pooled across sampling rounds) and pan trap days were included as separate random effects for
 282 each invertebrate group sampled using that method (both were included for true bugs and bees and
 283 wasps). For each GLMM the appropriate error term (Poisson or negative binomial) was selected by
 284 examining the ratio of deviance / residual degrees of freedom of full (global) models. Candidate
 285 models comprising three possible variable combinations (sub-treatment, strata and the sub-
 286 treatment*strata interaction; additive sub-treatment and strata effects without the interaction
 287 term; sub-treatment alone) were examined using the package 'lme4' (Bates, Maechler, Bolker &
 288 Walker 2017). The top-ranked model was considered 'best' if $\Delta AICc > 2$ (Akaike's Information
 289 Criterion corrected for small sample size) relative to the next-ranked model (Burnham & Anderson
 290 2002); for competing models within 2 $\Delta AICc$ the most parsimonious was selected as additional
 291 variables lacked strong support (Burnham & Anderson 2002). Next, where strata was retained in the
 292 selected model, we merged strata levels with similar coefficients if this did not reduce model
 293 performance ($\Delta AICc \leq 2$). Last, the fixed effect of sub-treatment was considered to be supported if
 294 performance of the selected model deteriorated ($\Delta AICc > 2$) upon its removal; in these cases,
 295 category means were compared by Tukey's pairwise comparison using the package 'multcomp'

296 (Hothorn, Bretz & Westfall 2008). Spatial autocorrelation of model residuals was examined by
297 Moran's I using the package 'Ape' (Paradis, Claude & Strimmer 2004).

298 Assemblage composition of sub-treatments and the influence of strata were examined by
299 Redundancy Analyses (RDA, using Euclidean distance measures), separately for each taxonomic
300 group (pooling non-priority and priority species), with square-root transformed species matrixes
301 (that provided better-fitting models than Hellinger transformation) and down-weighting of rare
302 species (to reduce the influence of particularly abundant or rare species) using the 'Vegan' package
303 (Oksanen *et al.* 2018). Samples with fewer than ten observations for that taxonomic group were
304 omitted to avoid over-representing localities where the assemblage was poorly characterized. To
305 determine whether the fixed effects of sub-treatment and strata were important to species
306 composition, we used backwards stepwise selection from the full RDA model (using the 'ordistep'
307 function, Oksanen *et al.* 2018) with 1,000 permutations ($P < 0.05$, based on ANOVA-like tests).

3. RESULTS

Untreated vegetation on control plots was characterised by tall swards (mean drop-disk height: 4.9 cm, 95% CI 4.2-5.8) and virtually no bare ground; treatments reduced vegetation height (first-time-shallow-cultivated, 1.5 cm, 1.3-1.8; first-time-deep-cultivated, 1.3 cm, 1.1-1.5) and created extensive bare ground (48%, 41-56; 96%, 94-97) (Fig. S3). Sampling provided 28,846 observations of 150 non-priority and 12 priority plant species, and 132,251 records of invertebrates (121,968 from pitfall traps, 10,283 from pan traps) from 707 non-priority and 170 priority species (see Table S3 for species and Table S4 for numbers sampled per invertebrate group). Priority (but not non-priority) ants, true flies and plants were excluded from separate taxonomic analyses because fewer than three species were observed on controls (ants and plants) or shallow-cultivated complexes (ants and true flies). Pitfall traps sampled more true bug species, whilst pan traps sampled more bee and wasp species (the only taxonomic groups sampled with both trapping methods) (Fig. S4).

3.1 Richness at the complex level

Sampling completeness estimates for treatment complexes and controls exceeded 90% of estimated total non-priority and priority species richness for every taxon with the sole exception of priority bees and wasps on the controls (80%) (Table S5a). For non-priority species, overall richness was greater on both treatments (deep-cultivated: 610 species, 95% CI 599-620; shallow-cultivated: 553, 543-564) than controls (445, 417-472), and deep-cultivation supported more species than shallow-cultivation (Fig. 2a). For separate taxa, richness was greater on treatments than controls for eight (deep-cultivation: other beetles; rove beetles; ground beetles; true bugs; bees and wasps; ants; true flies; plants) and four (shallow-cultivation: rove beetles; ground beetles; true bugs; plants) of the nine groups. Deep-cultivation supported greater richness than shallow-cultivation for five groups (other beetles; true bugs; bees and wasps; true flies; plants).

For priority species, overall richness on both treatments (deep-cultivation: 114 species, 110-120 95% CI; shallow-cultivation: 107, 102-112) was approximately double that of controls (60: 49-70) (deep- and shallow-cultivation did not differ) (Fig. 2a). For separate taxa, richness was greater on both treatments than controls for four (other beetles; ground beetles; spiders; true bugs) of the six groups considered. Deep-cultivation supported more priority ground beetles but fewer priority true bugs than shallow-cultivation.

Lowering the BSS to the smallest sample size affected inferences for four of 17 rarefaction analyses (Table S6); in three, non-significant differences became significant (owing to narrower CIs): inference at twice the lowest sample size is therefore conservative. Inference from the overall cross-taxa analyses was not affected when complexes and control plots lacking plant data were excluded (Fig. S5). For the two groups sampled by two trapping methods, removing the least effective method meant deep-cultivated complexes no longer supported more priority true bugs than controls, but did not affect inference for bees and wasps and non-priority true bugs (Fig. S6).

3.2 Composition at the complex level

For non-priority species, both treatments supported a larger percentage of the overall non-priority species-pool (deep-cultivation: 77%, 95% CI 74-80%; shallow-cultivation: 71%, 67-75%) than controls (56%, 54-58%; Fig. 2b); more non-priority species were unique to deep-cultivation (17% relative to the species-pool, CI 14-20%) than were only recorded in either shallow-cultivation (10%, 8-12%) or controls (8%, 6-9%). For separate taxa, treatments held more unique non-priority species than controls, respectively for three groups (deep-cultivation: other beetles; ground beetles; bees and wasps) or one group (shallow-cultivation: ground beetles), while controls did not hold more unique non-priority species for any group.

For priority species, both treatments supported double the percentage of the overall priority species-pool (deep-cultivated: 65%, CI 57-71%; shallow-cultivated: 65%, 59-72%) than controls

(overall assemblage: 38%, 35-41%; Fig. 2b). Both treatments supported a greater unique percentage of priority species (deep-cultivated: 23%, 17-28%; shallow-cultivated: 20%, 14-26%) than controls (9%, 6-12%). For separate taxa, treatments only held more unique priority species than controls for one group (deep-cultivation: ground beetles; shallow-cultivation: other beetles), while controls did not hold more unique priority species for any group.

3.3 Responses of habitat and management guilds

Of the 707 non-priority and 170 priority invertebrate species, 551 non-priority and 135 priority species were associated with dry-open habitats according to Pantheon (75 and 14 were associated with wet/shaded habitats, while broad ecological requirements of 81 and 21 were unknown); of these, 518 (94%) and 123 (91%) were classified among the three habitat guilds (Table S3). For non-priority species, richness of the 'tall sward and scrub' associated species was greater on both complex treatments than controls, richness of the 'short sward' species was similar across both treatments and controls, and for the 'short sward and bare ground' species richness was greater for one treatment (deep-cultivation) than controls (Fig. 3). For priority species, richness of the 'tall sward and scrub' associated species was greater on one treatment (shallow-cultivation) than controls, richness of the 'short sward' species was again similar across both treatments and controls, and for the 'short sward and bare ground' species richness was nearly three times greater on both treatments than on controls.

Of the 135 dry-open priority invertebrate species, 105 (78%) were classified into five management guilds (Table S3). Response to both complex treatments was progressively greater for management guilds with more intense requirements (Fig. 3): for the 15 priority species autecology classified as 'no ground-disturbance and no/light grazing', richness was similar across both treatments and controls (indicating a lack of treatment penalty); for those classified as 'no ground-disturbance and heavy grazing' (17 species) or 'light ground-disturbance and no/light grazing' (15

species), richness was greater for shallow-cultivation or both treatments (respectively) than controls; for those classified as 'heavy ground-disturbance and no/light grazing' (33 species), richness on both treatments was double that on controls; while for those classified as 'heavy ground-disturbance and heavy grazing' (25 species), richness on both treatments was three times that on controls. Lowering the BBS to the smallest sample size affected inference for one of the six habitat guild analyses and none of the five management guild analyses (Table S6).

3.4 Richness at the sub-treatment level

Sub-treatments differed in vegetation structure; vegetation height was similar across the first-time and repeatedly-cultivated sub-treatments, irrespective of cultivation method (shallow or deep-cultivation; Fig. S3), but the deep-cultivations contained more abundant bare-ground than the shallow-cultivations. Vegetation height and cover recovered after fallowing, and more quickly after shallow-cultivation, while deep-cultivation still retained abundant bare ground as one-year-old fallows. Ewe abundance in May 2017 did not vary between controls and sub-treatments (Fig. S2); but appeared to increase on the sub-treatments through the summer (Hawkes pers. obs.).

Sample completeness was high (>90%) across all eight individual sub-treatments, for non-priority and priority rove and ground beetles, and non-priority other beetles, spiders, true bugs, bees and wasps, ants and plants (Table S5b). Where sampling completeness was less than 90% it remained strong (e.g. 80-90% for priority spiders and true bugs in one sub-treatment, for priority other beetles in two sub-treatments and for priority bees and wasps across five sub-treatments; and 78-88% for non-priority true flies across all deep-cultivated sub-treatments).

For non-priority species, richness was greater on all eight sub-treatments, compared to controls, for rove beetles, ground beetles, and bees and wasps (Fig. 4). Non-priority other beetle richness was greater on two sub-treatments (one-year-old-fallow deep-cultivated and repeatedly-

shallow-cultivated), non-priority plant richness was greater on one sub-treatment (one-year-old-fallowed deep-cultivated), and non-priority true bug richness was lower on two sub-treatments (repeatedly- and first-time deep-cultivated). Non-priority spider, ant, and true fly richness did not differ between the sub-treatment categories and controls. For non-priority other beetles and plants, richness was greater on calcareous and intermediate-aged grassland (pooled) than ancient-acid grassland; whilst for non-priority spiders, richness was greater on intermediate-aged grassland than calcareous and ancient-acid grassland (pooled) (though no interactions between sub-treatment and strata were supported, Table S7).

For the priority species, richness was greater on the one-year-old-fallowed deep-cultivated sub-treatment, compared to controls, for other beetles, ground beetles, true bugs and bees and wasps (Fig. 4). For priority other beetles and bee and wasps, richness was also greater on the repeatedly-shallow-cultivated sub-treatment, and the two-year-old fallowed sub-treatments (both shallow- and deep-cultivated), respectively. Priority rove beetle and spider richness did not differ between any of the sub-treatment categories and controls. For priority ground beetles and other beetles, richness was greater on calcareous and intermediate-aged grassland (pooled) than ancient-acid grassland (again, interactions between sub-treatment and strata were not supported, Table S7).

Residuals from six of the fifteen richness models (non-priority and priority spiders, non-priority ground beetles, priority rove beetles, non-priority true bugs, and non-priority ants) were significantly, though weakly, spatially autocorrelated (Moran's $I = 0.03, -0.02, 0.08, 0.08, 0.04$ and 0.06 , respectively), suggesting some variation attributable to geographic factor(s) not considered in the models. Nevertheless, we consider inference for sub-treatment effects to be robust, as treatments and controls were distributed randomly.

3.5 Composition at the sub-treatment level

Redundancy analyses (Fig. 5) supported differences in composition among sub-treatment and strata categories for other beetles, rove beetles, ground beetles, spiders, true bugs, ants, and plants (bees and wasps and true flies were omitted from this analysis). RDA models explained between 25% (other beetles) and 43% (ground beetles, Fig. 5) of overall variance in sample composition. For spiders, ground beetles, rove beetles, true bugs and ants, sub-treatment explained more variance than strata; for other beetles and plants, sub-treatment and strata explained similar variance (Table S8).

For the three beetle groups, the composition differed markedly between the repeatedly- / first-time-cultivated sub-treatments (both shallow- and deep-cultivated) and controls, whilst the one- and then two-year-old-fallowed sub-treatments converged towards controls (but did not overlap). Spiders showed a similar pattern, but with less convergence towards the controls. For plants and true bugs, the four deep-cultivated sub-treatments and the repeatedly-shallow-cultivated sub-treatment differed markedly from controls, whilst the remaining three shallow-cultivated sub-treatments converged towards controls (but remained distinct). Ants showed a less clear pattern, but the first-time-deep-cultivated sub-treatment was distinct from controls and all other sub-treatments. For other beetles, rove beetles, ground beetles and plants, calcareous grassland and ancient-acid grassland supported markedly different assemblages, with intermediate-aged grassland positioned between the two; whilst for spiders, true bugs and ants, calcareous grassland and intermediate-aged grassland assemblages were similar, and ancient-acid grassland was distinct.

4. DISCUSSION

Through one of the largest multi-taxa land management experiment yet conducted in a European grassland, we quantified consequences of management interventions inspired by *a priori* knowledge of the intensity and complexity of historic land-use, and informed by systematic, cross-taxa analysis of priority species and their ecological requirements. Sampling over 130,000 invertebrates and using 28,000 observations of plants showed that using ground-disturbance to create treatment complexes increased the overall richness of both non-priority and priority species, while complexes held more unique priority species than controls. Within treatment complexes, the one-year-old deep-cultivated fallow sub-treatment increased priority species richness across multiple taxa, but assemblage composition varied between all sub-treatments. Our studies demonstrate that providing the full sub-treatment complement, through different establishment methods, will cater for the needs of the widest range of species.

Both treatment complexes increased structural complexity and supported a greater overall richness of non-priority and priority species than controls, consistent with the well-established benefits of habitat heterogeneity (Stein, Gerstner & Kreft 2014). More surprising was the magnitude of response - especially of priority species - which nearly doubled in richness with the complex treatment. Although priority species responded particularly well to the one-year-old deep-cultivated fallows (but not the shallow-cultivated equivalent), the shallow- and deep-cultivated complexes were as effective at enhancing overall priority species richness, potentially as responses to each establishment method varied between taxa. For example, priority ground beetles responded more strongly to the barer deep-cultivated complexes; probably because some important ruderal food plants (e.g. *Chenopodium album*) were more abundant on this treatment (Table S3; most priority ground beetles were granivores). Priority true bugs, by contrast, responded more strongly to the more vegetated shallow-cultivated complexes; possibly because some species were less tolerant of the more intensive deep-cultivated treatment (subtle differences in management can drastically alter leafhoppers communities, Biedermann, Achtziger, Nickel & Stewart 2005). Interestingly,

priority bee and wasp richness was similar across regimes; it is likely that many of these species utilised the bare-open sub-treatments on the treatment complexes for nesting, and the fallowed sub-treatments and controls for foraging. Pan traps may have also sampled large numbers of wide-ranging individuals.

Within treatment complexes, composition was distinct between individual sub-treatments for nearly every taxonomic group considered, suggesting the efficacy of the complex was not attributable to one sub-treatment *per se* but the full sub-treatment complement. Juxtaposition of cultivation sub-treatments within complexes may have further increased richness by providing, in close proximity, the contrasting micro-habitats needed by species whose requirements vary during their life-cycle. Irrespective of this, the provision of additional structural complexity by overlapping and abutting treatments was simpler and less costly than creating an equivalent set of independent, isolated, sub-treatment plots.

At the complex level, treatment effects on richness were greater for priority than non-priority spiders, ground beetles, other beetles, and true bugs (shallow-cultivation only), perhaps because for these groups species associated with 'short sward' or 'short sward and bare ground' comprised a greater proportion of the priority than of the non-priority species (Fig. S7). As complexes were optimised to inferred cross-taxa requirements of the largest number of priority species (Dolman *et al.* 2012), it is encouraging that benefits from treatments were much greater for priorities than non-priorities across most groups, but also appeared to avoid penalty to species not classified as requiring physical disturbance.

Across most taxonomic groups, both complex treatments and controls supported a similar proportion of unique non-priority and priority species. Supported by our analysis of assemblage composition at the sub-treatment level, this demonstrates that, whilst complexes may increase richness across most taxonomic groups, no single establishment method can deliver the resource requirements of the whole species-pool (a central finding of Fuller *et al.* 2017). To cater for the

broadest range of species, efforts to implement enhanced management should adopt different establishment methods to create complex nested heterogeneity, while also retaining some untreated habitat; however further work would be required to optimise the relative extent of treatments to untreated resource. Whilst the present study focused on maximising cross-taxa richness of priority species, as this is an appropriate goal for conserving biodiversity, further research would be needed to determine the consequence of these treatments for the phylogenetic or functional diversity that may be relevant to other ecosystem services.

The mosaic of grasslands on different soils support contrasting plant assemblages including calcareous grassland, acid grassland or lowland heathland (Rodwell 1991; Rodwell 1992), that differed in composition for most invertebrate groups. Nevertheless, they were subject to similar historic land-use (Sheail 1979; Fuller *et al.* 2017) and are characterised by similar ecological processes of nutrient limitation, drought and physical disturbance (Dolman & Sutherland 1992). Crucially, treatment efficacy for biodiversity never varied between strata, indicating results are applicable to a wide range of dry grassland habitats, regardless of compositional differences. Similarly, across fen habitat and the superficially very different wet grassland-ditch complexes of grazing marsh habitat, biodiversity audit analysis of priority species' requirements showed similar functional dependence on littoral margins, undulating topography, early and late succession and extended management rotations (Mossman, Panter & Dolman 2012). Together, these findings challenge a long-held paradigm that difference in plant composition between functionally similar habitats predicate differences in conservation practice.

Based on a bioregional analysis of priority species requirements and knowledge of land-use history we predicted that dry-open habitat species associated with heavy disturbance would respond more strongly to treatment than those associated with little or no disturbance. For priority invertebrate species, those *a priori* thought to be associated with the heaviest forms of physical-disturbance responded most strongly to treatment, regardless of whether these were classified

along a single composite physical-disturbance gradient (habitat guilds) or on independent gradients of grazing intensity and ground-disturbance (management guilds). This confirms the importance of prevailing historic management to extant priority species (Fuller *et al.* 2017) and the success of the Biodiversity Audit approach in targeting enhanced management interventions appropriate to the ecological requirements of priority species (Dolman *et al.* 2012). For non-priority invertebrates, those apparently associated with short swards and bare ground responded positively to deep-cultivations, whilst those associated with tall swards and scrub responded positively to both treatments. This again reflects the structural heterogeneity of treatment complexes, which provided open, short swards on recent cultivations and one-year-old fallows, and taller swards on the two-year-old fallows. Whilst this experiment focused on mechanical interventions, other more natural approaches – such as the use of wild boar *Sus scrofa* or large herbivores - can promote dynamic mosaics with resulting benefits for some priority species and taxa (De Schaetzen, Van Langevelde & WallisDeVries 2018; Garrido *et al.* 2019) and may be equally effective for priority species associated with heavy forms of disturbance.

4.1 Synthesis and applications

Fuller *et al.* (2017) showed that a better appreciation of the complexity and intensity of historical management, combined with knowledge of priority species requirements, encourages novel forms of enhanced intervention within cultural landscapes. Through an unprecedented landscape-scale biodiversity experiment, we confirm that restoring structural complexity and nested heterogeneity to grassland, irrespective of fine-scale differences in vegetation structure and composition, both increased non-priority species richness and, crucially, doubled priority species richness. To maximise cumulative richness, complexity should be created through a range of establishment methods, as shallow- or deep-cultivation each supported unique species. Additionally, within the complexes, the full sub-treatment complement is needed to support the widest suite of species (one- and two-year

old fallows, repeat and first-time cultivations), as assemblage composition varied with sub-treatment. This management should be implemented in such a way that treatment complexes are surrounded by untreated habitat, which supports its own distinct assemblage and unique species.

Responses to treatments varied considerably across taxa, for example differing markedly between priority ground beetles and rove beetles at the complex level; this emphasises the value of multi-taxa sampling when evaluating the biodiversity consequence of management interventions (e.g. Vessby, Söderström, Glimskär & Svensson 2002; Yong *et al.* 2020). While taxonomic surrogacy may therefore be inadvisable, multi-taxa species groups defined by their ecological associations (habitat- and management-guilds) usefully predicted responses to interventions. Thus, systematic analysis of the relative frequency of species with contrasting resource requirements, across the full complement of priority species, can also inform management strategies and prescriptions in other biogeographical regions, without the need for costly multi-taxa experiments. Where biodiversity is well-characterised and autecological knowledge strong (e.g. much of Europe) we recommend such regional Biodiversity Audits, synthesised with a detailed understanding of historic land-use, to better inform conservation interventions. These findings could be extrapolated to other biogeographical regions that comprise similar semi-natural habitats but lack equivalent levels of biodiversity data and autecological knowledge.

AUTHORS' CONTRIBUTIONS

RH, JS, AB and PD conceived the study and designed methodology; RH, HJ, SL, CL, JM, NO, ARB, JW and DW collected the data; RH conducted all analyses; RH, JS, AB, and PD led the writing of the manuscript. All authors contributed critically to the drafts.

ACKNOWLEDGEMENTS

The Royal Society for the Protection of Birds and Natural England (NE) funded this work through the Action for Birds in England program, with additional support from Deference Infrastructure Organisation and Breaking New Ground. Funding from the Natural Environment Research Council (NE/G524095/1) contributed to the coding of species attributes. We thank Sheep Enterprise, the STANTA conservation groups, Shadwell Estate, D. Ash, J. Black, T. Cowan, R. Evans, S. Gilham, C. Hainsworth, I. Levett, K. Marsden, T. Pankhurst, J. Symonds and M. Taylor for valuable assistance. Cranfield University provided soil data under license. We are grateful to Norfolk Biodiversity Information Services, NE, and the many volunteers who assisted with invertebrate processing.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study will be available on Dryad upon publication.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

REFERENCES

- Bailey, M. (1989). *A marginal economy? East Anglian Breckland in the later Middle Ages*. Cambridge University Press, Cambridge.
- Bakker, J.P., De Bie, S., Dallinga, J.H., Tjaden, P. & De Vries, Y. (1983). Sheep-grazing as a management tool for heathland conservation and regeneration in the Netherlands. *Journal of Applied Ecology*, 20(2), 541-560. <https://doi.org/10.2307/2403525>
- Bates, D., Maechler, M., Bolker, B. & Walker, S. (2017). Linear mixed-effects models using 'Eigen' and S4. *R package Version 1.1-1.5*. Available at: <https://cran.r-project.org/web/packages/lme4/> (accessed May 2018).
- Biedermann, R., Achtziger, R., Nickel, H. & Stewart, A.J.A. (2005). Conservation of grassland leafhoppers: A brief review. *Journal of Insect Conservation*, 9(4), 229-243. <https://doi.org/10.1007/s10841-005-0531-z>
- Burnham, K.P. & Anderson, D.R. (2002). *Model selection and multi-model inference: a practical information-theoretic approach*. Springer, New York.
- Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. & Ellison, A.M. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, 84(1), 45-67. <https://doi.org/10.1890/13-0133.1>
- Clark, J.A. & May, R.M. (2002). Taxonomic bias in conservation research. *Science*, 297(5579), 191-192. <https://doi.org/10.1126/science.297.5579.191b>
- De Schaetzen, F., Van Langevelde, F. & WallisDeVries, M.F. (2018). The influence of wild boar (*Sus scrofa*) on microhabitat quality for the endangered butterfly *Pyrgus malvae* in the Netherlands. *Journal of Insect Conservation*, 22(1), 51-59. <https://doi.org/10.1007/s10841-017-0037-5>
- Denton, J. (2013). Comment: Conservation grazing of heathland – where is the logic? *British Wildlife*, 24, 339–347.

614 Dolman, P.M., Panter, C.J. & Mossman, H.L. (2012). The biodiversity audit approach challenges
615 regional priorities and identifies a mismatch in conservation. *Journal of Applied Ecology*, 49,
616 986-997. <https://doi.org/10.1111/j.1365-2664.2012.02174.x>

617 Dolman, P.M. & Sutherland, W.J. (1991). Historical clues to conservation: Management techniques
618 that reproduce old farming practices may yet save one of the last remaining heathlands in
619 Britain. *New Scientist*, 1751, 40-43.

620 Dolman, P.M. & Sutherland, W.J. (1992). The ecological changes of Breckland grass heaths and the
621 consequences of management. *Journal of Applied Ecology*, 29(2), 402-413.
622 <https://doi.org/10.2307/2404509>

623 Dolman, P.M. & Sutherland, W.J. (1994). The use of soil disturbance in the management of Breckland
624 grass heaths for nature conservation. *Journal of Environmental Management*, 41(2), 123-
625 140. <https://doi.org/10.1006/jema.1994.1038>

626 EC (1992). *Council Directive 92/43/EEC of 21 May 1992 on the Conservation of Natural Habitats and*
627 *of Wild Fauna and Flora*. European Commission, Brussels.

628 Fuller, R.J., Williamson, T., Barnes, G. & Dolman, P.M. (2017). Human activities and biodiversity
629 opportunities in pre-industrial cultural landscapes: Relevance to conservation. *Journal of*
630 *Applied Ecology*, 54(2), 459-469. <https://doi.org/10.1111/1365-2664.12762>

631 Garrido, P., Mårell, A., Öckinger, E., Skarin, A., Jansson, A. & Thulin, C.-G. (2019). Experimental
632 rewilding enhances grassland functional composition and pollinator habitat use. *Journal of*
633 *Applied Ecology*, 56(4), 946-955. <https://doi.org/10.1111/1365-2664.13338>

634 Goodenough, A.E. & Sharp, M.H. (2016). Managing calcareous grassland for the declining Duke of
635 Burgundy *Hamearis lucina* butterfly: Effects of grazing management on *Primula* host plants.
636 *Journal of Insect Conservation*, 20(6), 1087-1098. [https://doi.org/10.1007/s10841-016-9946-](https://doi.org/10.1007/s10841-016-9946-y)
637 [y](https://doi.org/10.1007/s10841-016-9946-y)

638 Griffiths, R.A. & Dos Santos, M. (2012). Trends in conservation biology: Progress or procrastination in
 639 a new millennium? *Biological Conservation*, 153, 153-158.
 640 <https://doi.org/10.1016/j.biocon.2012.05.011>
 641 Härdtle, W., Niemeyer, M., Niemeyer, T., Assmann, T. & Fottner, S. (2006). Can management
 642 compensate for atmospheric nutrient deposition in heathland ecosystems? *Journal of*
 643 *Applied Ecology*, 43(4), 759-769. <https://doi.org/10.1111/j.1365-2664.2006.01195.x>
 644 Hawkes, R.W., Smart, J., Brown, A., Jones, H. & Dolman, P.M. (2019a). Experimental evidence that
 645 ground-disturbance benefits Woodlark *Lullula arborea*. *Ibis*, 161, 447-452.
 646 <https://doi.org/doi:10.1111/ibi.12696>
 647 Hawkes, R.W., Smart, J., Brown, A., Jones, H., Lane, S., Wells, D. & Dolman, P.M. (2019b). Multi-taxa
 648 consequences of management for an avian umbrella species. *Biological Conservation*, 236,
 649 192-201. <https://doi.org/10.1016/j.biocon.2019.05.039>
 650 Heaver, D., Webb, J., Roy, D., Dean, H., Harvey, M., Macadam, C. & Curson, J. (2017). Pantheon: A
 651 new resource for invertebrate survey standards and analysis. *In Practice: Bulletin of the*
 652 *Chartered Institute of Ecology and Environmental Management*, 98, 25-29.
 653 <http://nora.nerc.ac.uk/id/eprint/518573>
 654 Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous inference in general parametric models.
 655 *Biometrical Journal*, 50(3), 346-363. <https://doi.org/10.1002/bimj.200810425>
 656 Hülber, K., Moser, D., Sauberer, N., Maas, B., Staudinger, M., Grass, V., . . . Willner, W. (2017). Plant
 657 species richness decreased in semi-natural grasslands in the Biosphere Reserve Wienerwald,
 658 Austria, over the past two decades, despite agri-environmental measures. *Agriculture,*
 659 *Ecosystems & Environment*, 243, 10-18. <https://doi.org/10.1016/j.agee.2017.04.002>
 660 Larsson, J. (2019). eulerr: Area-proportional euler and venn diagrams with ellipses. R package version
 661 5.1.0. Available at: <https://cran.r-project.org/package=eulerr> (accessed on 05/01/2020).

662 Linnell, J.D.C., Kaczensky, P., Wotschikowsky, U., Lescureux, N. & Boitani, L. (2015). Framing the
 663 relationship between people and nature in the context of European conservation.
 664 *Conservation Biology*, 29(4), 978-985. <https://doi.org/10.1111/cobi.12534>
 665 Mace, G.M. (2014). Whose conservation? *Science*, 345(6204), 1558-1560.
 666 <https://doi.org/10.1126/science.1254704>
 667 Merckx, T., Feber, R.E., Hoare, D.J., Parsons, M.S., Kelly, C.J., Bourn, N.A.D. & Macdonald, D.W.
 668 (2012). Conserving threatened Lepidoptera: Towards an effective woodland management
 669 policy in landscapes under intense human land-use. *Biological Conservation*, 149(1), 32-39.
 670 <https://doi.org/10.1016/j.biocon.2012.02.005>
 671 Mossman, H.L., Panter, C.J. & Dolman, P.M. (2012). Fens biodiversity audit: Part 1 & 2 - Methodology
 672 and results. School of Environmental Sciences, University of East Anglia, Norwich.
 673 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., . . . Wagner, H. (2018).
 674 Community ecology package version 2.5-2. [https://cran.r-](https://cran.r-project.org/web/packages/vegan/vegan.pdf)
 675 [project.org/web/packages/vegan/vegan.pdf](https://cran.r-project.org/web/packages/vegan/vegan.pdf) (accessed July 2018).
 676 Paradis, E., Claude, J. & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R
 677 language. *Bioinformatics*, 20(2), 289-290. <https://doi.org/10.1093/bioinformatics/btg412>
 678 Pedley, S.M., Franco, A.M.A., Pankhurst, T. & Dolman, P.M. (2013). Physical disturbance enhances
 679 ecological networks for heathland biota: A multiple taxa experiment. *Biological*
 680 *Conservation*, 160, 173-182. <https://doi.org/10.1016/j.biocon.2013.01.006>
 681 Pettorelli, N., Barlow, J., Stephens, P.A., Durant, S.M., Connor, B., Schulte to Bühne, H., . . . du Toit,
 682 J.T. (2018). Making rewilding fit for policy. *Journal of Applied Ecology*, 55(3), 1114-1125.
 683 <https://doi.org/10.1111/1365-2664.13082>
 684 Pywell, R.F., Bullock, J.M., Tallowin, J.B., Walker, K.J., Warman, E.A. & Masters, G. (2007). Enhancing
 685 diversity of species-poor grasslands: An experimental assessment of multiple constraints.
 686 *Journal of Applied Ecology*, 44(1), 81-94. <https://doi.org/10.1111/j.1365-2664.2006.01260.x>

687 R Core Team (2015). R: a language and environment for statistical computing. Version 3.2.2. R
688 Foundation for Statistical Computing, Vienna.

689 Ratcliffe, D. (1977). *A nature conservation review: Volume 1: The selection of biological sites of*
690 *national importance to nature conservation in Britain*. Cambridge University Press, New
691 York.

692 Rodwell, J.S. (1991). *British plant communities. Volume 2: Mires and heaths*. Cambridge University
693 Press, Cambridge.

694 Rodwell, J.S. (1992). *British plant communities. Volume 3: Grasslands and montane vegetation*.
695 Cambridge University Press, Cambridge.

696 Schall, P., Gossner, M.M., Heinrichs, S., Fischer, M., Boch, S., Prati, D., . . . Ammer, C. (2018). The
697 impact of even-aged and uneven-aged forest management on regional biodiversity of
698 multiple taxa in European beech forests. *Journal of Applied Ecology*, 55(1), 267-278.
699 <https://doi.org/10.1111/1365-2664.12950>

700 Seibold, S., Gossner, M.M., Simons, N.K., Blüthgen, N., Müller, J., Ambarlı, D., . . . Weisser, W.W.
701 (2019). Arthropod decline in grasslands and forests is associated with landscape-level
702 drivers. *Nature*, 574, 671-674. <https://doi.org/10.1038/s41586-019-1684-3>

703 Sheail, J. (1979). Documentary evidence of the changes in the use, management and appreciation of
704 the grass-heaths of Breckland. *Journal of Biogeography*, 6(3), 277-292.
705 <https://doi.org/10.2307/3038180>

706 Stein, A., Gerstner, K. & Kreft, H. (2014). Environmental heterogeneity as a universal driver of species
707 richness across taxa, biomes and spatial scales. *Ecology Letters*, 17(7), 866-880.
708 <https://doi.org/10.1111/ele.12277>

709 Thomas, C.D. (1994). Extinction, colonization, and metapopulations: environmental tracking by rare
710 species. *Conservation Biology*, 8(2), 373-378. [https://doi.org/10.1046/j.1523-](https://doi.org/10.1046/j.1523-1739.1994.08020373.x)
711 [1739.1994.08020373.x](https://doi.org/10.1046/j.1523-1739.1994.08020373.x)

- Tipping, E., Davies, J.A.C., Henrys, P.A., Jarvis, S.G., Rowe, E.C., Smart, S.M., . . . Pakeman, R.J. (2019). Measured estimates of semi-natural terrestrial NPP in Great Britain: Comparison with modelled values, and dependence on atmospheric nitrogen deposition. *Biogeochemistry*, 144(2), 215-227. <https://doi.org/10.1007/s10533-019-00582-5>
- Van Klink, R. & WallisDeVries, M.F. (2018). Risks and opportunities of trophic rewilding for arthropod communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1761). <https://doi.org/doi:10.1098/rstb.2017.0441>
- Vessby, K., Söderström, B., Glimskär, A. & Svensson, B. (2002). Species-richness correlations of six different taxa in Swedish seminatural grasslands. *Conservation Biology*, 16(2), 430-439. <https://doi.org/10.1046/j.1523-1739.2002.00198.x>
- Wells, T.C.E. (1969). Botanical aspects of conservation management of chalk grasslands. *Biological Conservation*, 2(1), 36-44. [https://doi.org/10.1016/0006-3207\(69\)90113-X](https://doi.org/10.1016/0006-3207(69)90113-X)
- Wright, H.L., Lake, I.R. & Dolman, P.M. (2012). Agriculture—a key element for conservation in the developing world. *Conservation Letters*, 5(1), 11-19. <https://doi.org/10.1111/j.1755-263X.2011.00208.x>
- Yong, D.L., Barton, P.S., Okada, S., Crane, M., Cunningham, S.A. & Lindenmayer, D.B. (2020). Conserving focal insect groups in woodland remnants: The role of landscape context and habitat structure on cross-taxonomic congruence. *Ecological Indicators*, 115, 106391. <https://doi.org/10.1016/j.ecolind.2020.106391>
- Żmihorski, M., Pärt, T., Gustafson, T. & Berg, Å. (2016). Effects of water level and grassland management on alpha and beta diversity of birds in restored wetlands. *Journal of Applied Ecology*, 53(2), 587-595. <https://doi.org/10.1111/1365-2664.12588>

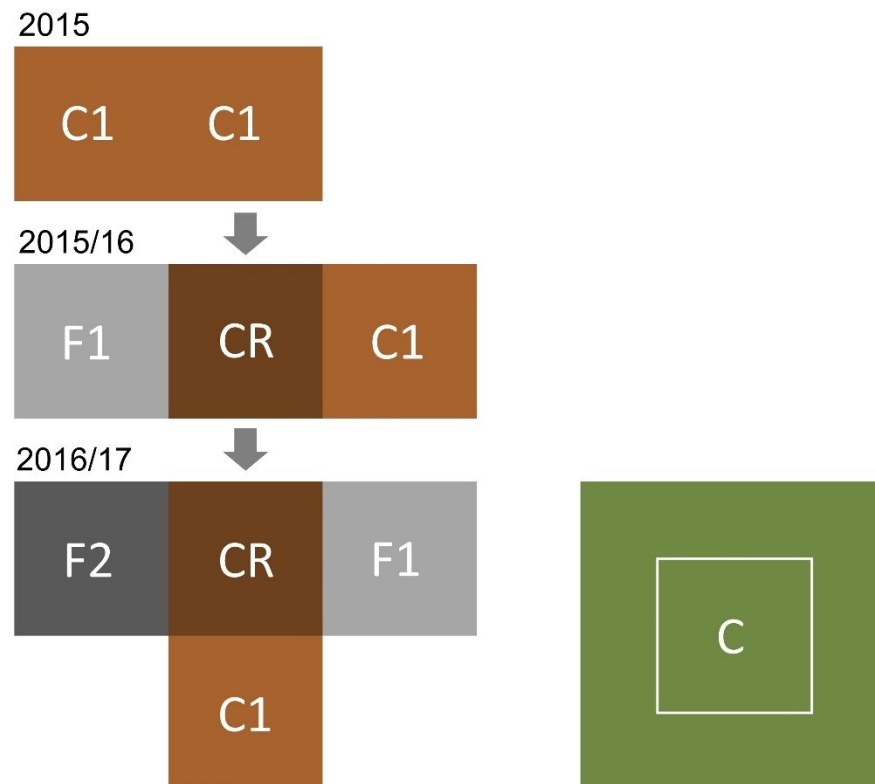
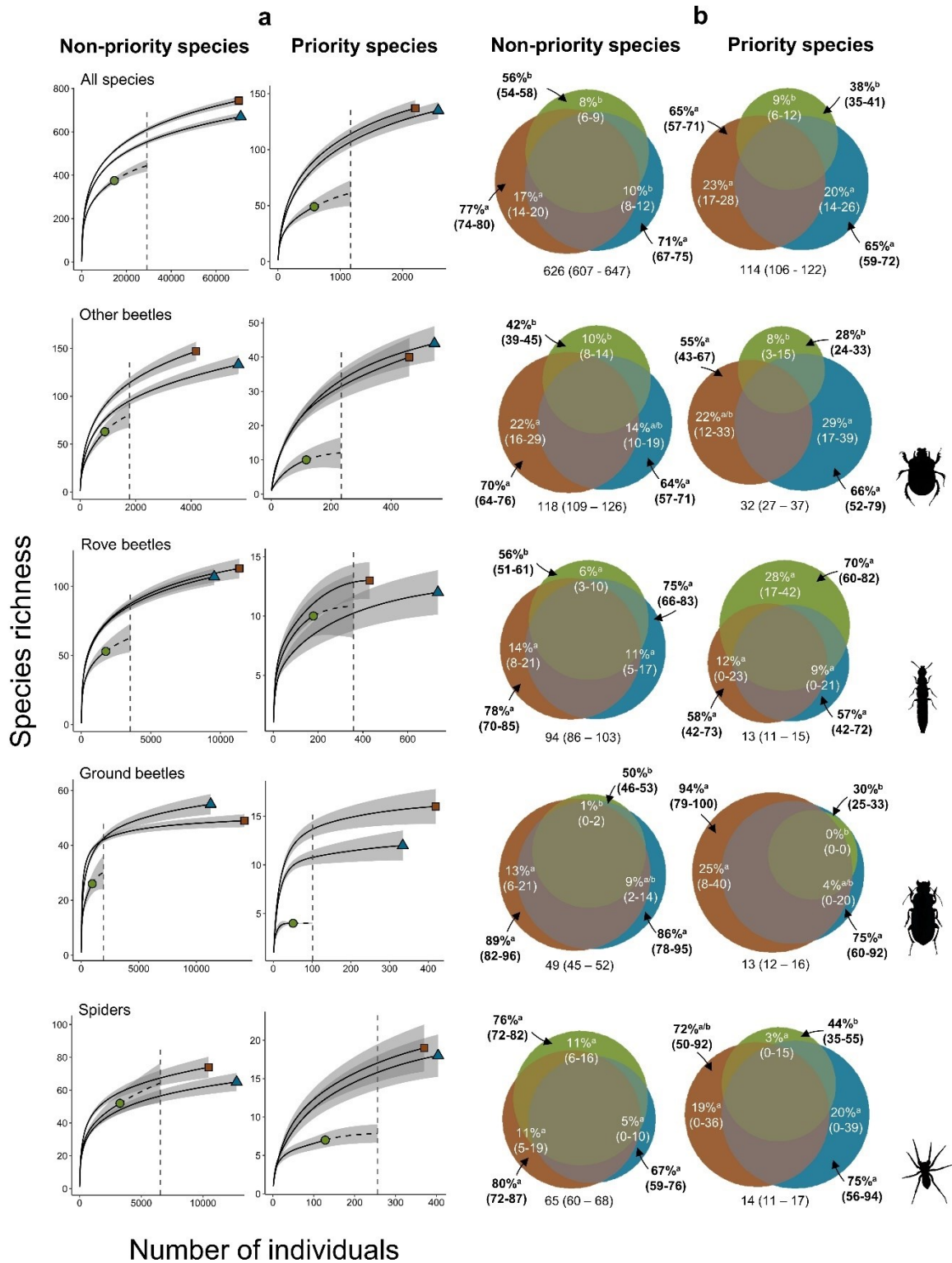
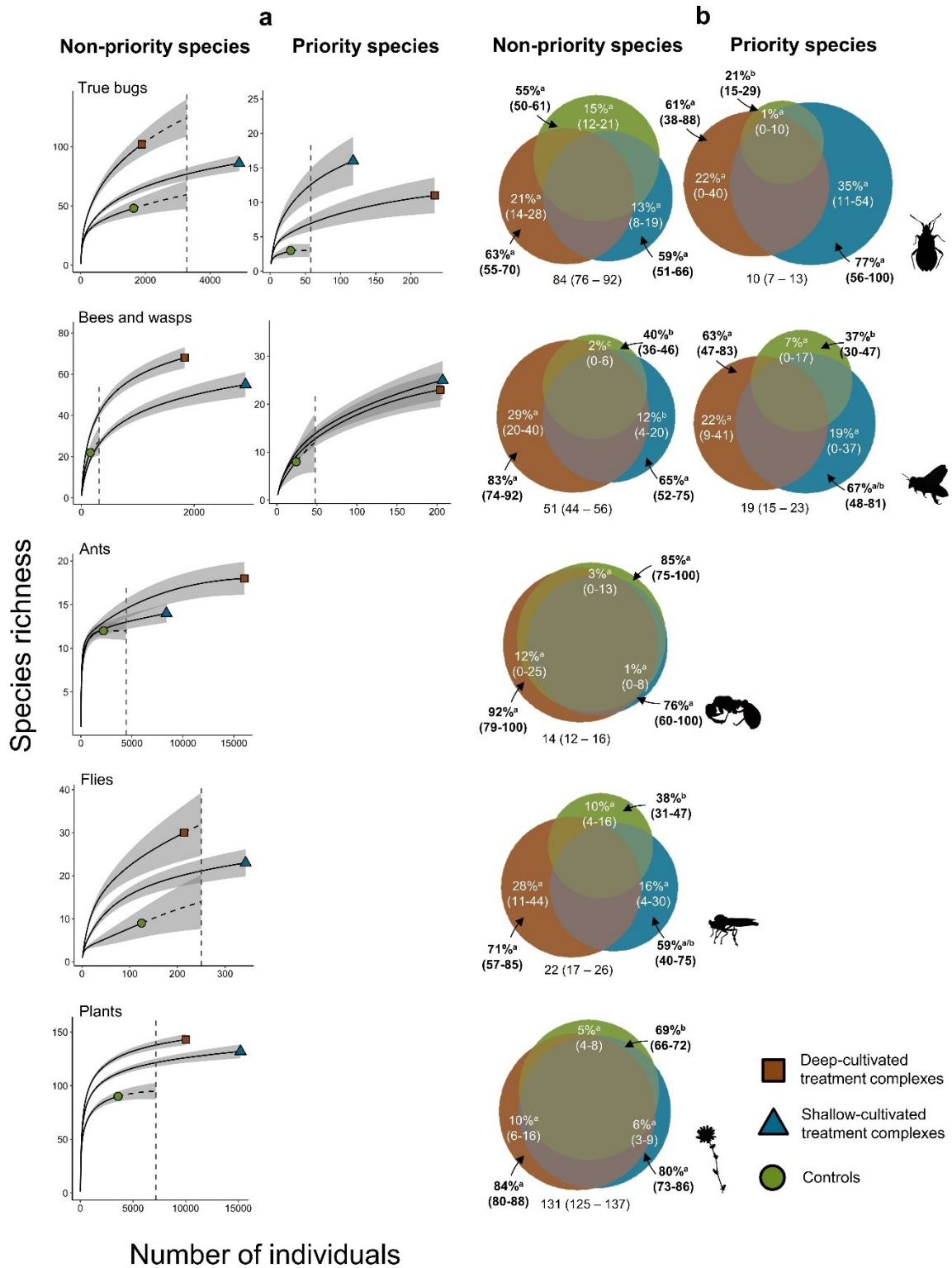


FIGURE 1 Development of a treatment complex over three successive winters to the final 4-ha complex (in 2017), comprising four 1-ha sub-treatments: CR, repeatedly-cultivated (brown); C1, first-time-cultivated (light brown); F1, one-year-old fallow (light grey); F2, two-year-old fallow (grey), also shown is a single 4-ha control plot (C, green) of which the central 1-ha (white outline) was sampled. See supplementary material Fig. S3 for mean vegetation height and bare ground extent in sub-treatments and controls.



744

745



746

747

FIGURE 2 Richness and composition of non-priority and priority species in shallow- or deep-cultivated treatment complexes (the enhanced management) and controls, shown for all species pooled and separately for each of nine taxonomic groups. Left panels (a) contrast richness between regimes through sample-based rarefaction (rescaled to numbers of individuals), for invertebrate groups using 20 complexes per treatment and 21 control plots for pooled species and for plants using 16 complexes per treatment and 16 controls. Symbols and solid lines denote observed and interpolated richness respectively, shading represents 95% CI bounds, the vertical dashed line denotes the base sample size (twice the smallest sample size) where richness was compared. Euler's (b) show the mean and 95% CI of total species richness across regimes (below each panel), and the percentage of this pool recorded within (outer bold values) and unique to (internal white values) each regime, based on 200 resampling iterations each comprising 16 subplots per treatment and all 16 control plots. Separately for both responses (percentage of the species pool and unique species) regimes that share a superscript do not differ (following pairwise comparisons with Bonferonni correction). For mean percentage overlap between each regime see Table S9. For plants, true flies and ants, limited numbers of priority species prohibited separate examination.

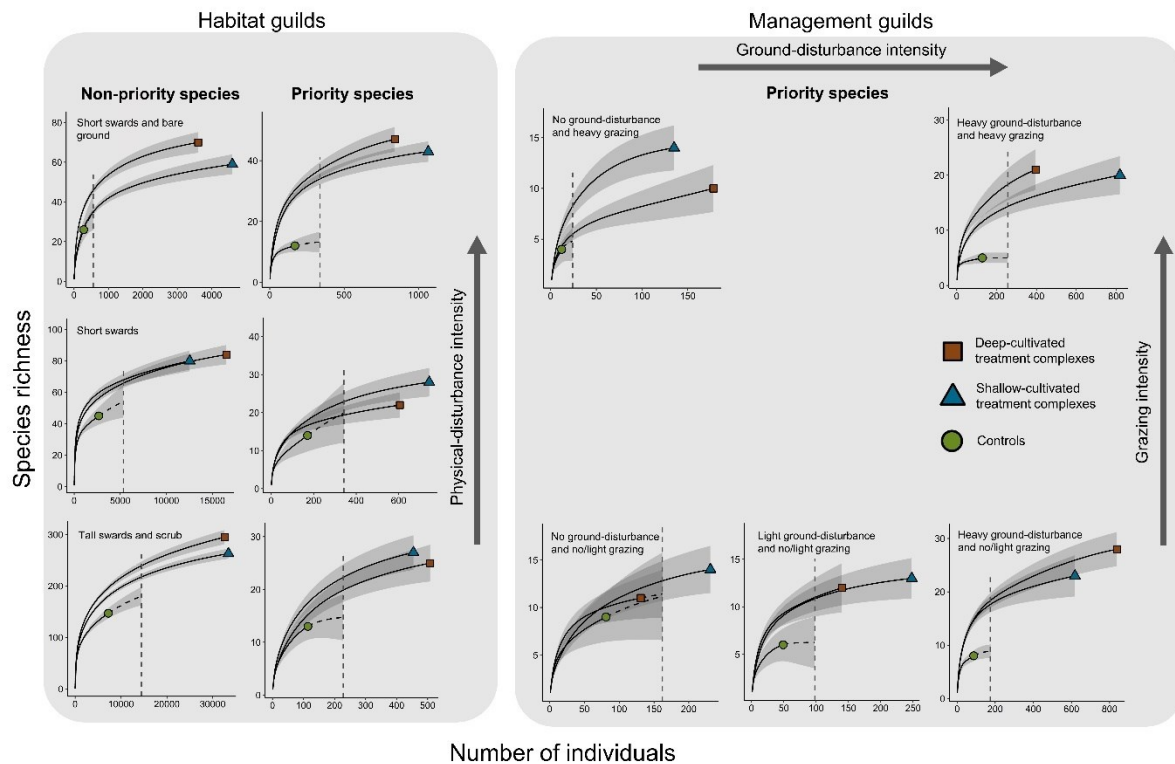


FIGURE 3 Response to enhanced grassland management of multi-taxa invertebrate guilds with differing habitat association and management requirements. Left panels consider three habitat guilds (from the Pantheon database) ranked along a composite gradient of increasing disturbance intensity: from tall swards and scrub, through short sward, to short sward with bare ground, separately for non-priority and priority species. Right panels (priority species only) consider biodiversity auditing classification in relation to independent gradients of grazing and ground-disturbance intensity. For each habitat or management guild, sample-based rarefaction (rescaled to numbers of sampled individuals) contrasts richness between shallow- ($n = 20$) or deep-cultivated ($n = 20$) treatment complexes and controls ($n = 21$). Symbols and solid lines denote observed and interpolated richness respectively, and shading represents 95% CI bounds. The vertical dashed line denotes the base sample size (twice the smallest sample size), where richness was compared.

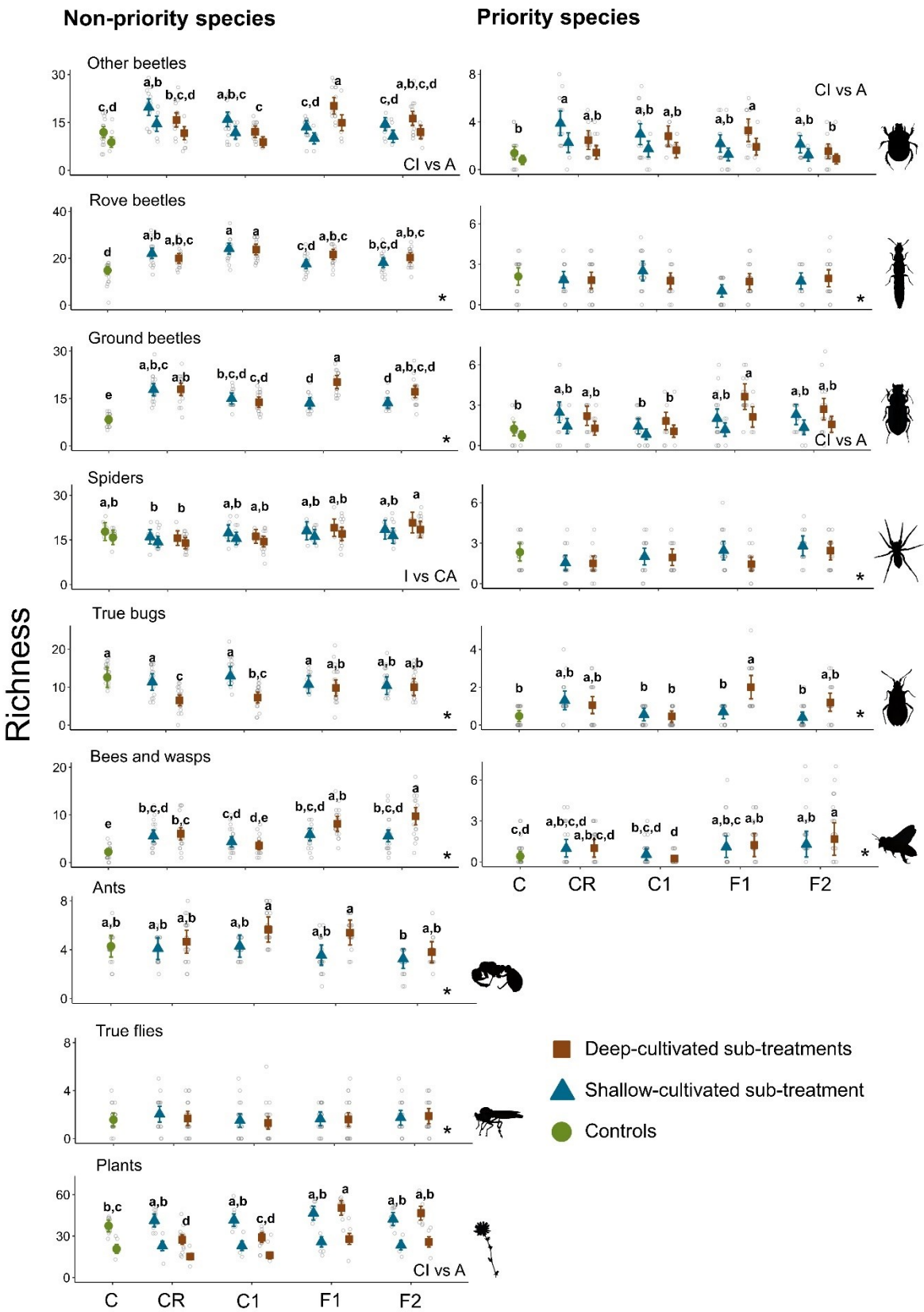


FIGURE 4 Richness of non-priority and priority species, within each of nine taxonomic groups, across repeatedly-cultivated (CR), first-time-cultivated (C1), one-year-old fallow (F1), and two-year-old fallow (F2) sub-treatments, within shallow- and deep-cultivated complexes, and in untreated controls (C). Sub-treatments and controls were compared by Generalized Linear Mixed Models that controlled for the fixed effect of strata (when retained during model simplification, see Table S10 for model coefficients), with strata merged either as ‘calcareous and intermediate-aged grassland’ (left offset) relative to ancient-acid grassland (right offset) (‘CI vs A’), or intermediate-aged grassland (left offset) relative to ‘calcareous and ancient-acid grassland’ (right offset) (‘I vs CA’), or with strata excluded from the model (indicated by *). Symbols denote predicted richness, bars 95% CIs, and open circles individual data points. Sub-treatment means that share a superscript (homogenous sub-sets, a–e) did not differ significantly (Tukey pairwise comparisons $P > 0.05$). Where no pairwise comparisons are reported the effect of treatment was not supported. For plants, true flies and ants, limited numbers of priority species prohibited separate examination.

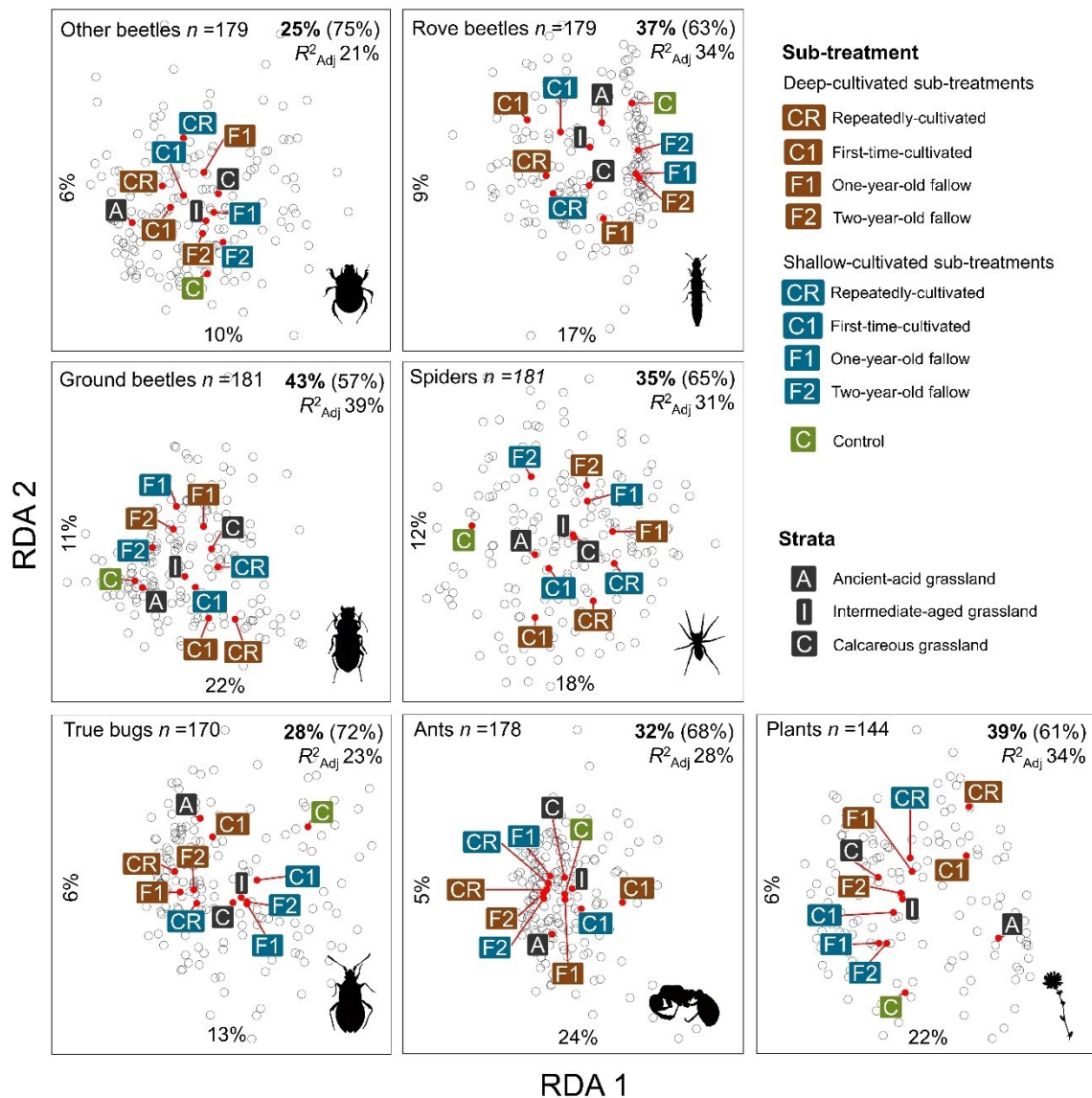


FIGURE 5 Redundancy analysis (RDA) relating assemblage composition to sub-treatment and strata, separately for each of seven taxa (pooling non-priority and priority species). See Table S8 for model details. For each model, the sample size is reported at the top left of the panel, the overall percentage variance explained (constrained in bold type; unconstrained in parentheses), adjusted R^2 , and the constrained variance explained by each axis (adjacent to each axis) are also shown; open circles show individual samples. Red points denote the centroid of each sub-treatment and strata category. For bees and wasps and true flies, limited numbers of sampled individuals prohibited separate examination.