

Supplementary Information

Table of Contents

Part 1 – Comparative methods and simulation model

Supplementary Figure 1. Illustration of non-parametric method for phylogenetic comparison.

Supplementary Figure 2. Different approaches to inferring correlations between variables in simulated taxonomically broad, sparse, unsystematized data with reversible predictor evolution.

Supplementary Figure 3. Different approaches to inferring correlations between variables in simulated taxonomically broad, sparse, unsystematized data with irreversible predictor evolution.

Supplementary Figure 4. Example of PGLS, PLM, and PGLM sensitivity-specificity investigation with irreversible predictor evolution.

Supplementary Figure 5. PGLS performance with changing predictor observations in simulated data.

Supplementary Figure 6. Influence of explicit branch length accounting in the inference process for simulated data.

Supplementary Figure 7. Effect of different correlation models on PGLS performance in simulated evolution.

Part 2 – Organelle DNA evolution

Supplementary Figure 8. Illustration of outlier influence.

Supplementary Figure 9. Features correlated with oDNA gene counts using PGLM.

Supplementary Figure 10. Correlations with mtDNA gene count when metazoans are removed from the phylogeny.

Supplementary Figure 11. mtDNA predictors in alternative plant phylogeny.

Supplementary Figure 12. ptDNA predictors in alternative plant phylogeny.

Supplementary Figure 13. mtDNA predictors using nonparametric, non-phylogenetic approach blocking eukaryotic clade.

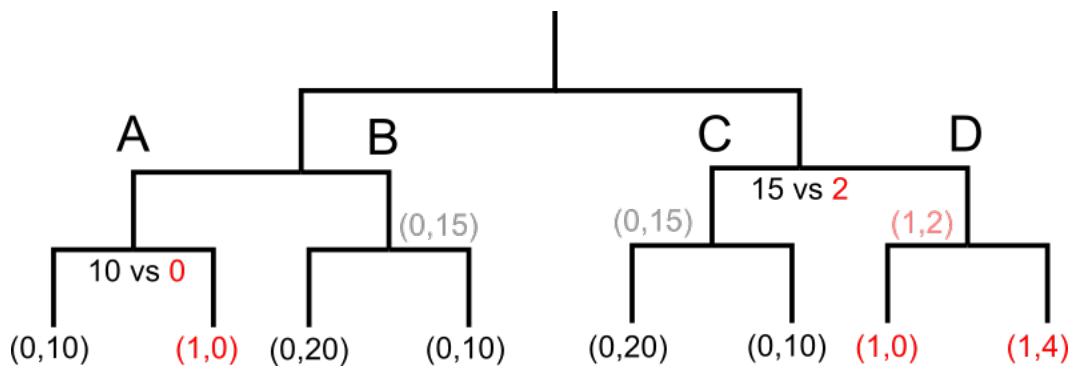
Supplementary Figure 14. ptDNA predictors using nonparametric, non-phylogenetic approach blocking eukaryotic clade.

Supplementary Figure 15. Mixed-model approaches assigning random effects to clade and analysing the remaining relationships.

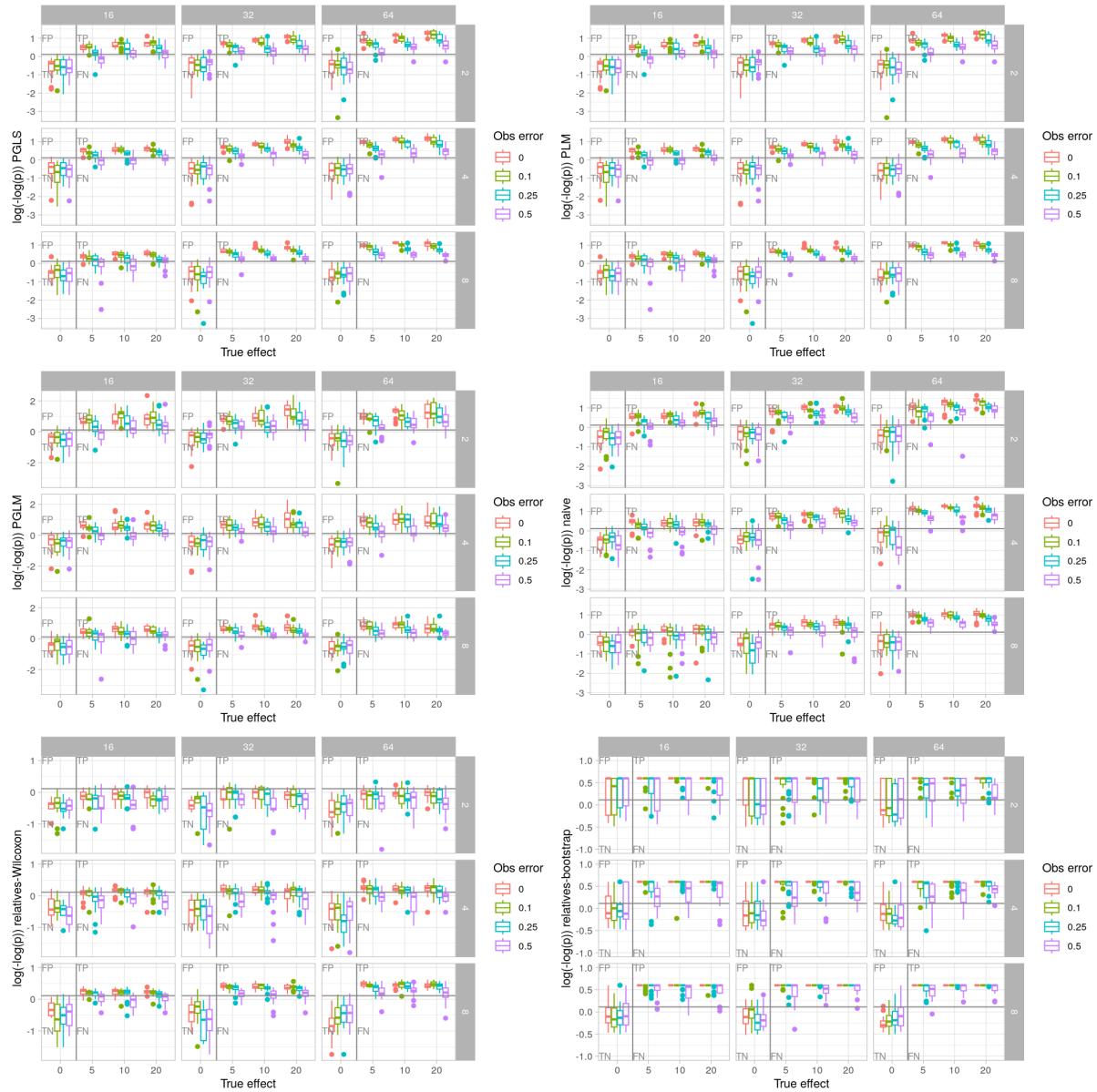
Supplementary Figure 16. Phylogenetic linear model results after normalizing clade oDNA gene counts.

Supplementary Table 1. Mitochondrial ecological predictors in our compiled database.

Supplementary Table 2. Plastid ecological predictors in our compiled database.

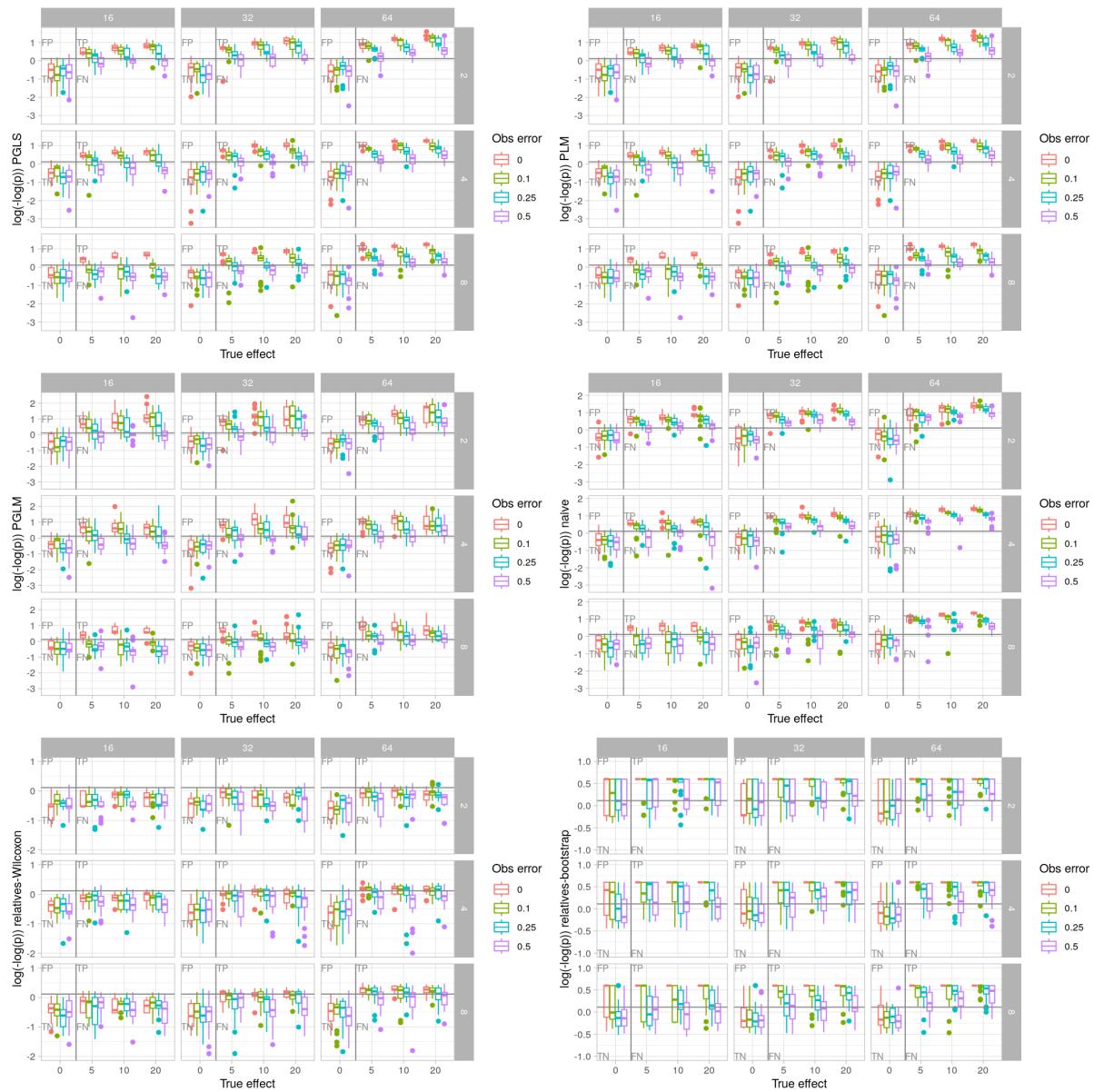


Supplementary Figure 1. Illustration of non-parametric method for phylogenetic comparison. Each tip is labelled with (predictor, response) values – for example (parasite, ptDNA count). In clade A, the two tips have different predictor values, so their response values (10 and 0) are added to a list of pairs for comparison. In clade B, the two tips have the same predictor value, and as the sister clade A has already been used in the dataset, no further comparisons are possible. In clade C, the tips have the same predictor value, so we label the ancestor with this predictor value and the mean of the response values (0,15). We do the same for the ancestor of clade D, obtaining (1,2). Now these ancestor nodes have different predictor values, so we add their response values (15 and 2) to the list of pairs for comparison. We then use a non-parametric test to compare the recorded 0-predictor responses (10 and 15) with the 1-predictor responses (0 and 2).

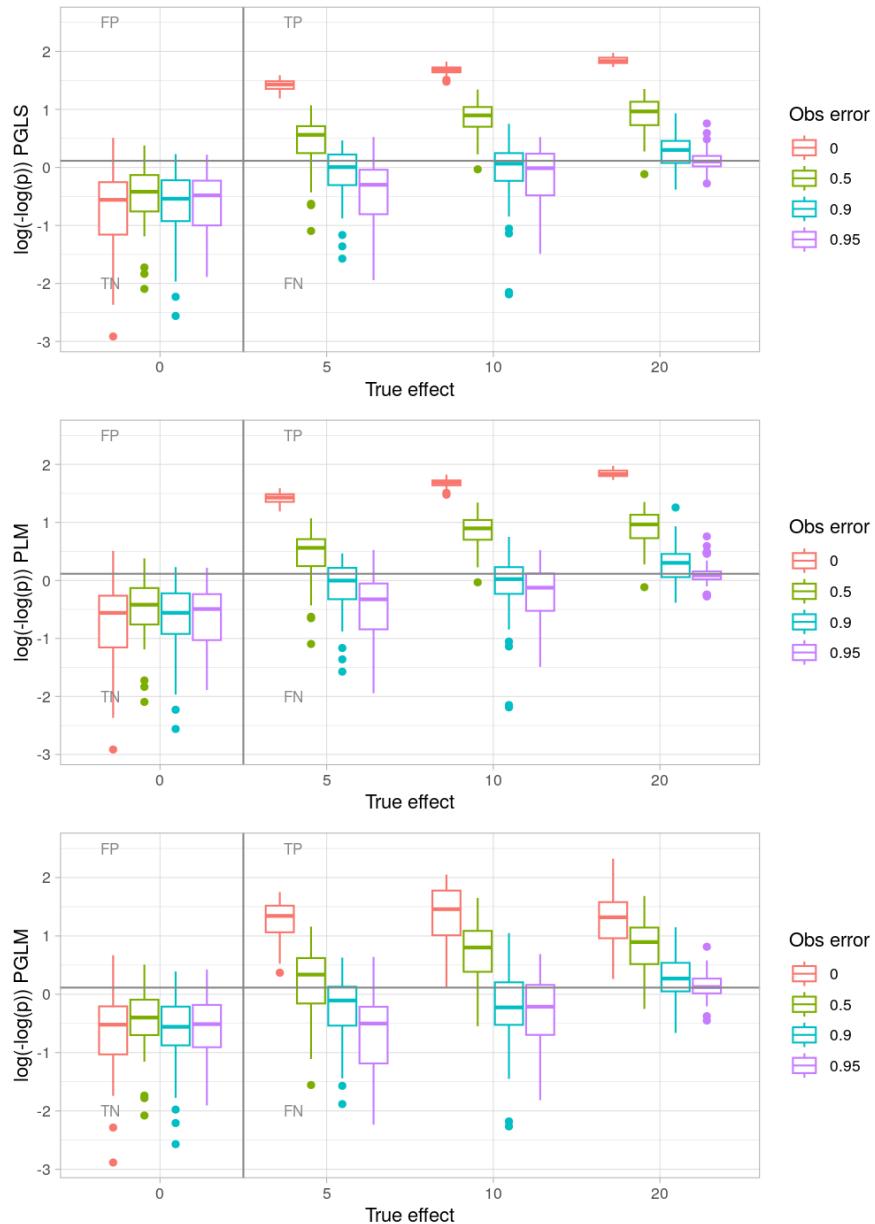


Supplementary Figure 2. Different approaches to inferring correlations between variables in simulated taxonomically broad, sparse, unsystematized data with reversible predictor evolution. In each subplot, the x-axis describes c , the influence of the predictor on the response variable, and the y-axis gives the distribution of p-values from an analysis. On the horizontal axis, 0 is no influence (null hypothesis); increasing influence gives a stronger signal. The different colours (noise labels) correspond to different proportions of occluded observations; 0 is perfect observations of the predictor, nonzero values are the probability of observing a negative value for a positive case. The facets in each frame give the size of the tree (columns) and the average number of evolutionary events giving a positive predictor value (rows). Finally, the different frames give different inference approaches: PGLS, no phylogenetic accounting, a non-parameteric approach comparing relatives with the Mann-Whitney test, and the same principle with bootstrap resampling.

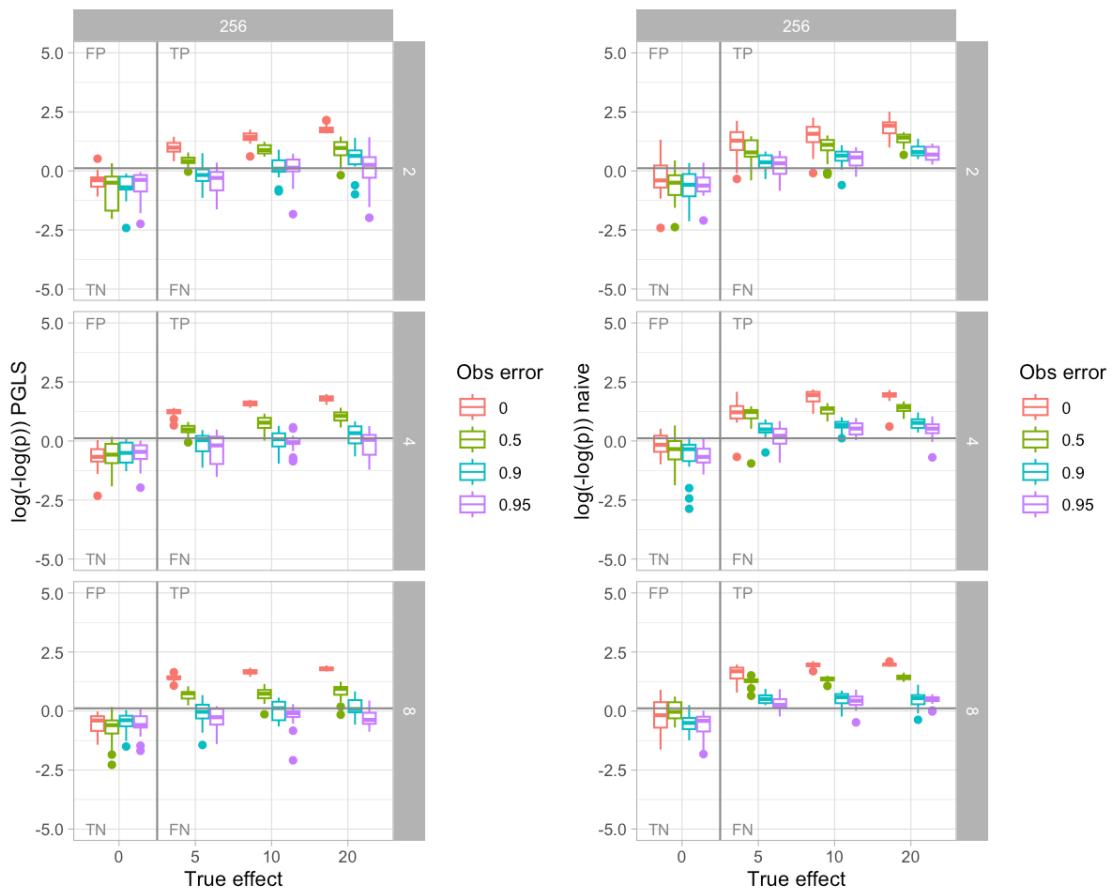
Comparative analyses with challenging (oDNA) data



Supplementary Figure 3. Different approaches to inferring correlations between variables in simulated taxonomically broad, sparse, unsystematized data with irreversible predictor evolution. In each subplot, the x-axis describes c , the influence of the predictor on the response variable, and the y-axis gives the distribution of p-values from an analysis. On the horizontal axis, 0 is no influence (null hypothesis); increasing influence gives a stronger signal. The different colours (noise labels) correspond to different proportions of occluded observations; 0 is perfect observations of the predictor, nonzero values are the probability of observing a negative value for a positive case. The facets in each frame give the size of the tree (columns) and the average number of evolutionary events giving a positive predictor value (rows). Finally, the different frames give different inference approaches: PGLS, no phylogenetic accounting, a non-parameteric approach comparing relatives with the Mann-Whitney test, and the same principle with bootstrap resampling.

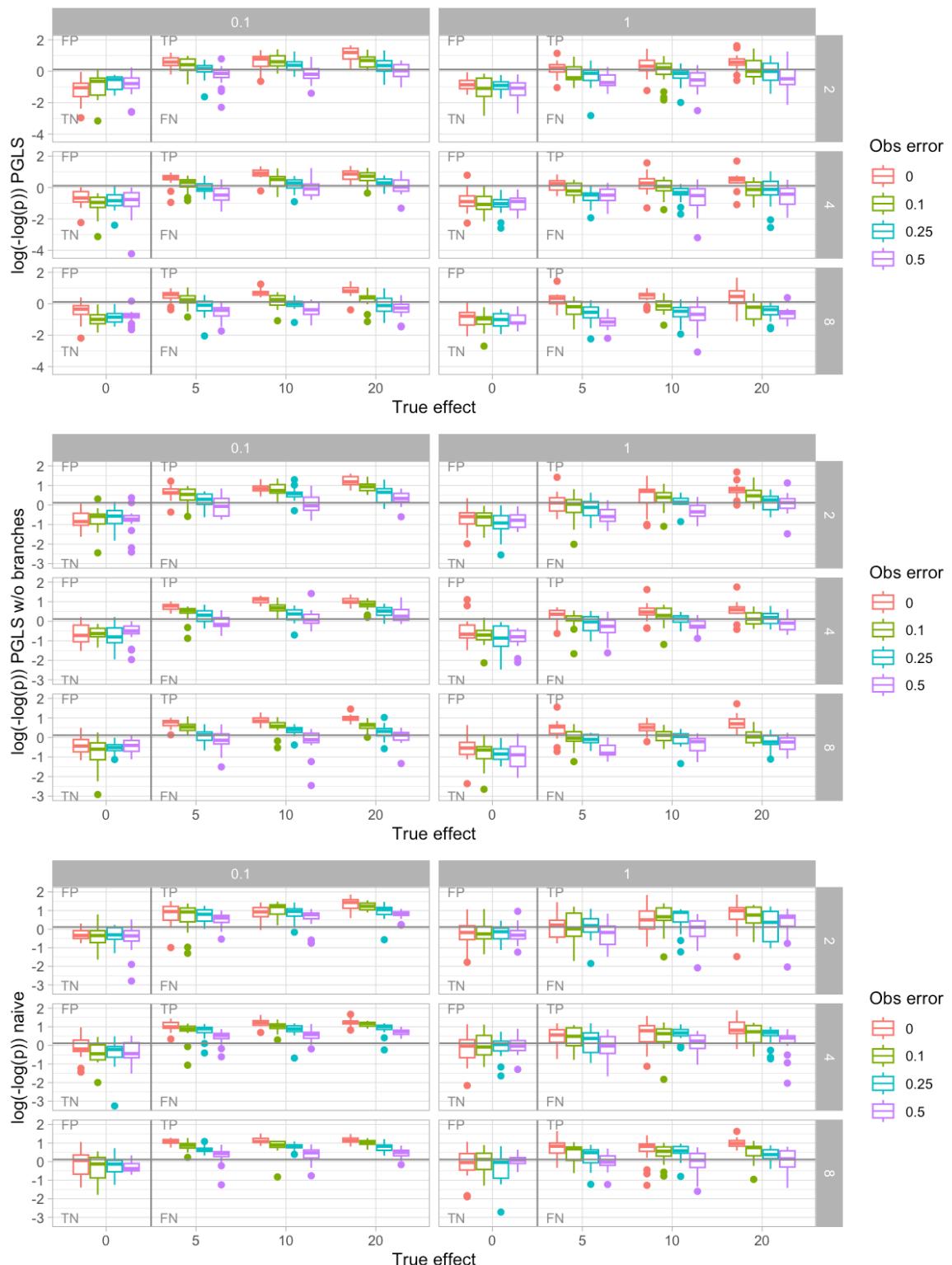


Supplementary Figure 4. Example of PGLS, PLM, and PGLM sensitivity-specificity investigation with irreversible predictor evolution. Evolutionary dynamics were simulated on a tree with 256 leaves, with differing true effect c linking predictor value and response evolution. Predictor value changed irreversibly through evolution. Observations of the predictor value were occluded with an observation error parameter giving the probability that a positive value is observed as a negative. The grey line corresponds to $p=0.05$; points above would be interpreted as the presence of a signal (without multiple hypothesis correction), points below as the absence of a signal. PGLS with covariance structure derived from a Brownian model rarely gives false positive (FP) correlations and has substantial power to detect true positive (TP) correlations even for high observation error probabilities; PLM is almost identical in performance. PGLM likewise limits FP and retains power to detect TP, although the spread of p-values reported by PGLM for positive cases is rather broader. *nleme* (Pinheiro et al. 2020) was used for PGLS; *phylolm* (Tung Ho and Ané 2014) for PLM and PGLM with generalized estimating equations. Here, an average of 8 evolutionary events innovated a positive predictor value.



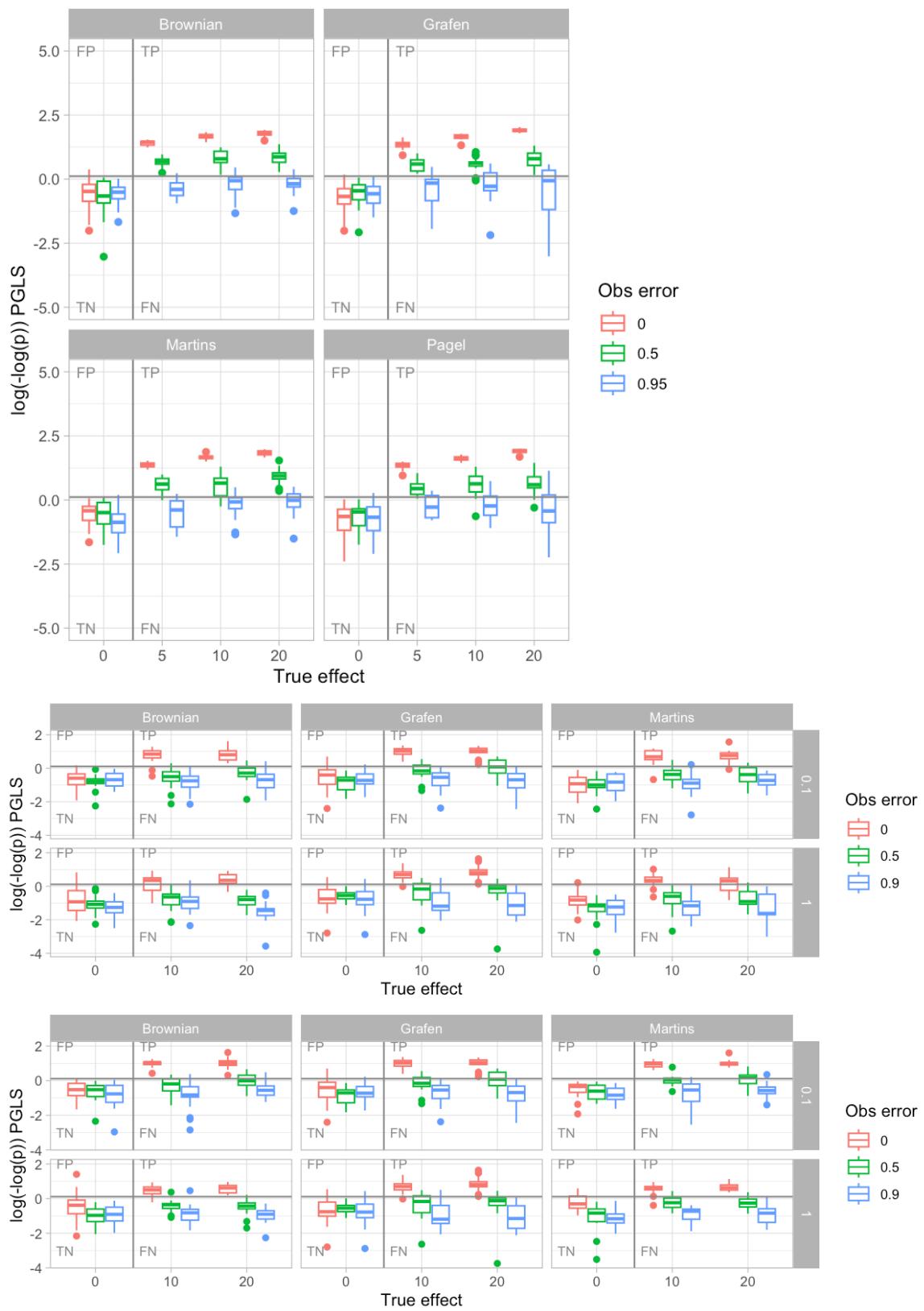
Supplementary Figure 5. PGLS performance with changing predictor observations in simulated data. As in Supplementary Figs 2-4, the x-axis gives the strength of relationship between predictor and response, and the y-axis gives a distribution of p-values from an analysis. The average number of events generating a positive predictor value is varied across rows, and different colours give different proportions of occluded predictor observations. PGLS rarely gives false positives, and retains some statistical power even when 95% of positive predictor values are observed as negative.

Comparative analyses with challenging (oDNA) data

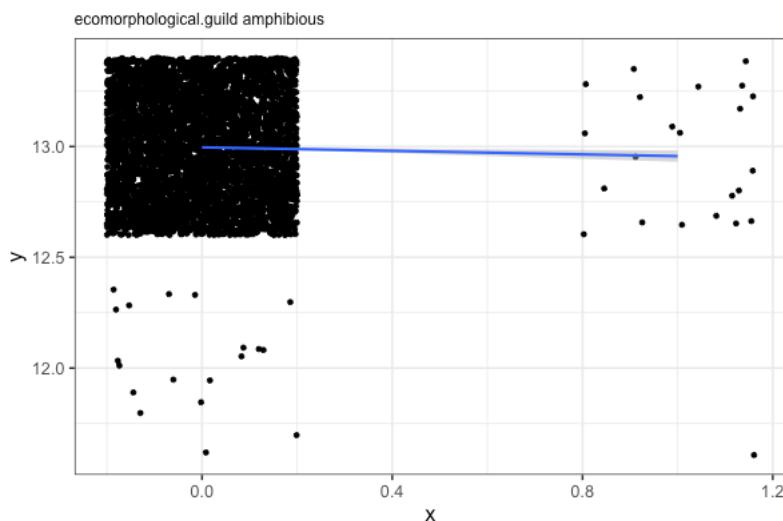


Supplementary Figure 6. Influence of explicit branch length accounting in the inference process for simulated data. p-values and true influence of predictor on response are plotted as in Supp Fig 1 and 2. The column facets for each frame now control the death parameter in the model generating synthetic phylogenetic trees. The frames correspond to PGLS with branch length information, no phylogenetic accounting, and PGLS with uniform branch lengths (topology only).

Comparative analyses with challenging (oDNA) data

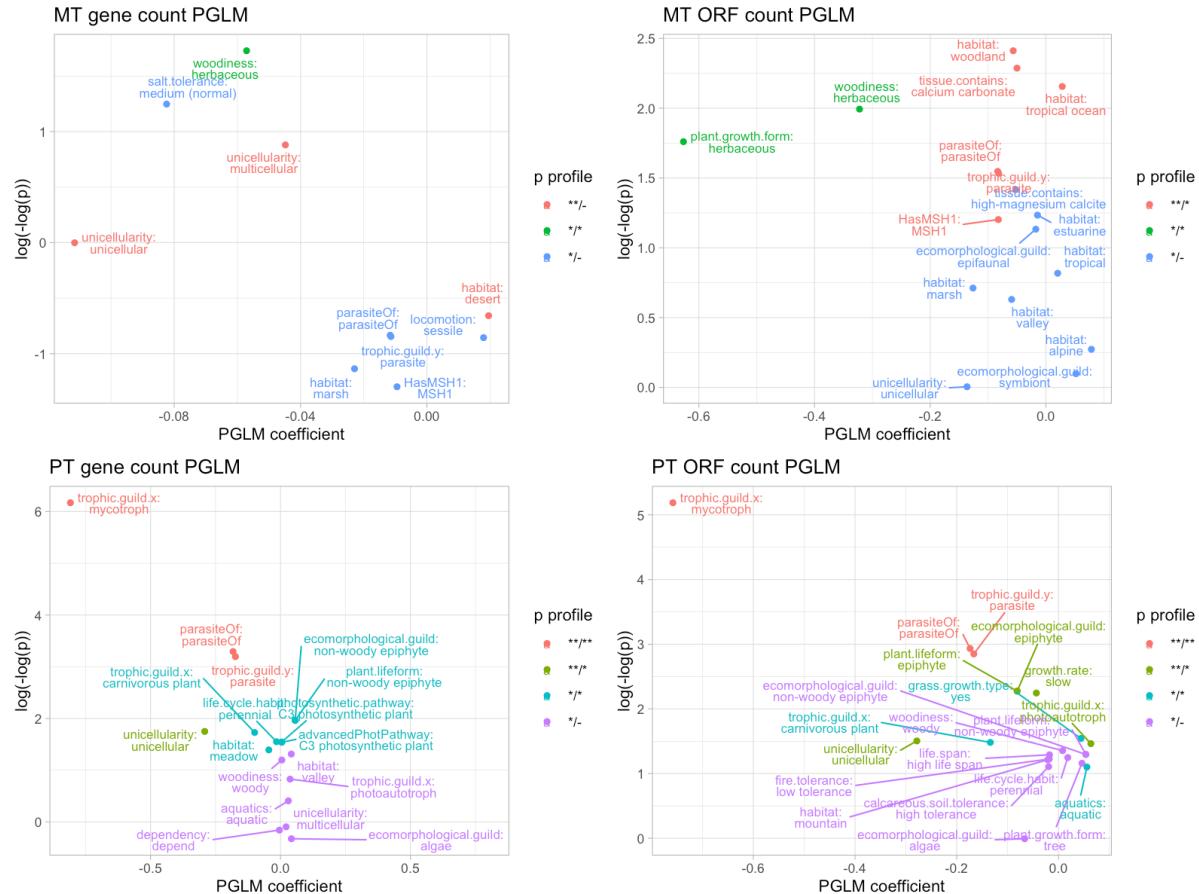


Supplementary Figure 7. Effect of different correlation models on PGLS performance in simulated evolution. Top, symmetric tree; bottom, birth-death trees with different death parameters.



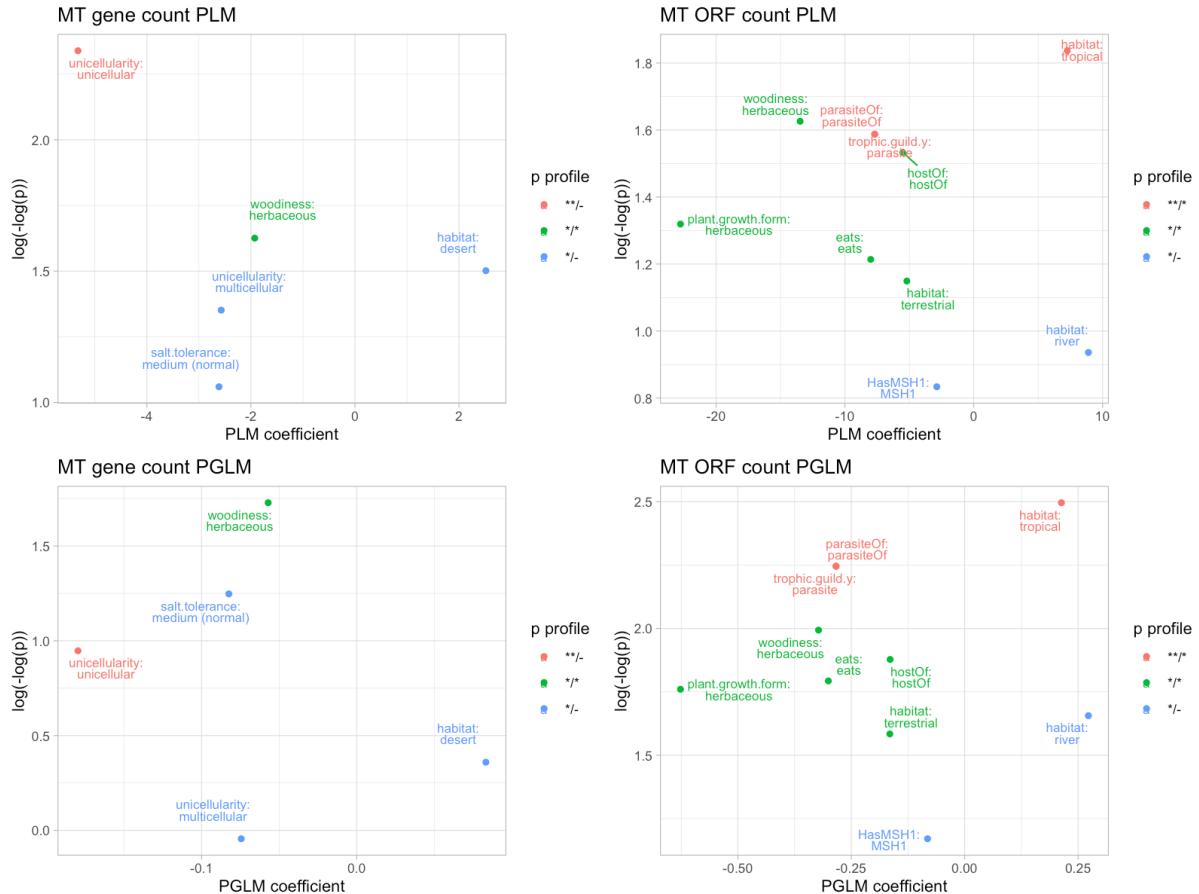
Supplementary Figure 8. **Illustration of outlier influence.** Horizontal axis is predictor feature (0, not amphibious; 1, amphibious); vertical axis is gene count (12 or 13); points are jittered to show density. The vast majority of non-amphibious samples have 13 genes; the presence of the single individual with 12 genes in the much smaller amphibious sample “significantly” changes the mean. But the effect completely depends on that single observation, and therefore can likely be regarded as a sampling artefact.

Comparative analyses with challenging (oDNA) data



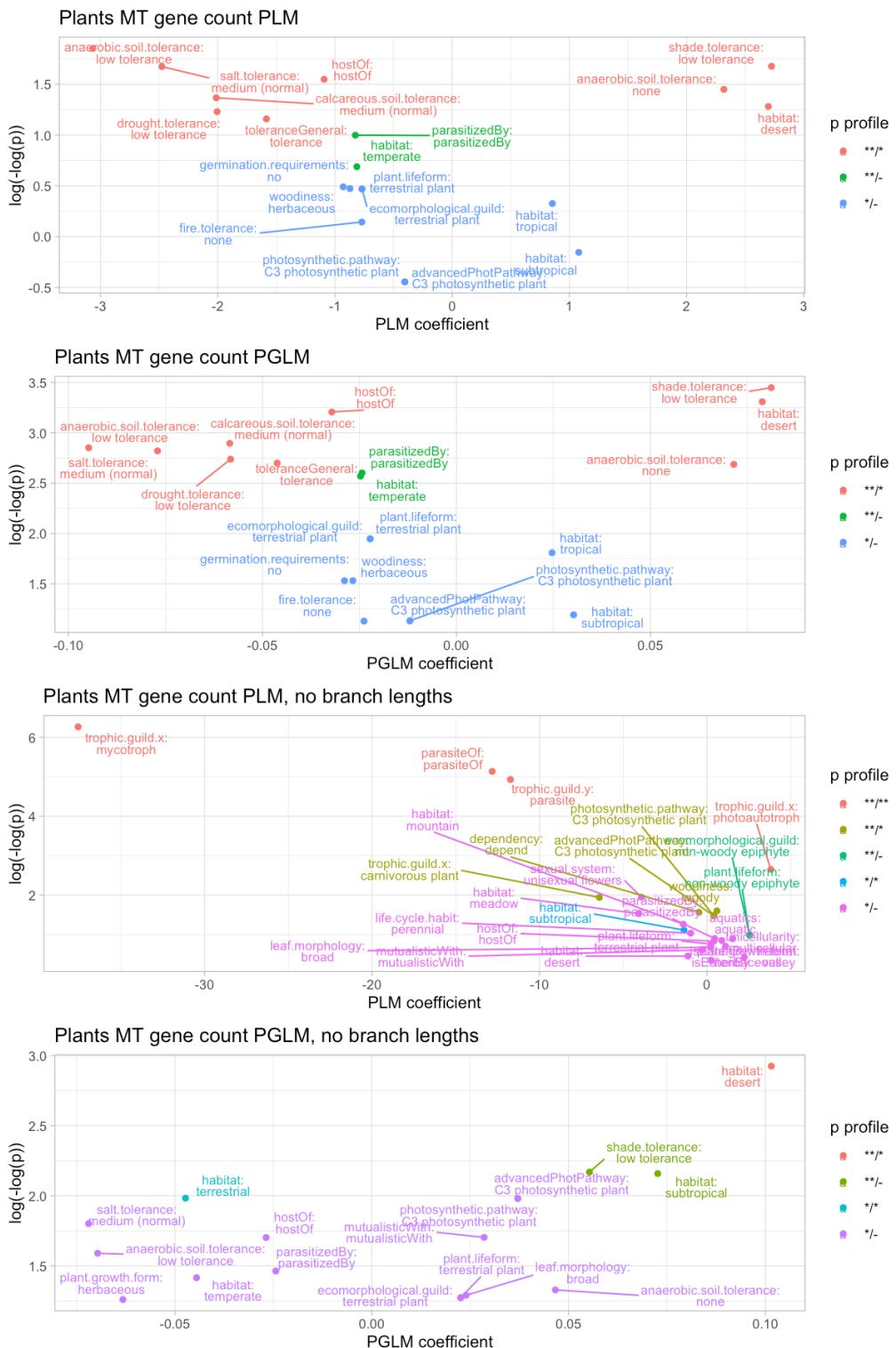
Supplementary Figure 9. Features correlated with oDNA gene counts using PGLM. PGLM coefficients (x-axis) and p-values (y-axis, double-logged and inverted) for relationships between different organismal traits and organelle DNA gene counts (mtDNA and ptDNA), counted as confirmed protein-coding genes or CDS regions. This analysis is applied to the cross-eukaryote dataset as described in the text. The figure shows statistics from the PGLM approach, but colours correspond to profiles of statistical significance using both PLM and PGLM approaches. ** denotes $p < 0.05$ after Bonferroni; * $p < 0.05$ without correction; - $p > 0.05$ (for example, **/* means one approach gave a Bonferroni-robust $p < 0.05$ and the other gave 0.05 not robust to Bonferroni). The PLM coefficient gives the average inferred change in gene count if an organism has a given property. The majority of traits give substantially higher p-values and lower-magnitude coefficients; plots are vertically truncated to focus on the more robust results.

Comparative analyses with challenging (oDNA) data



Supplementary Figure 10. Correlations with mtDNA gene count when metazoans are removed from the phylogeny. Plots organized and labelled as in Fig. 5. PLM and PGLM coefficients (x-axis) and p-values (y-axis, double-logged and inverted) for relationships between different organismal traits and mtDNA gene counts, counted as confirmed protein-coding genes or CDS regions. Colours correspond to profiles of statistical significance using PLM and PGLM approaches; ** denotes $p < 0.05$ after Bonferroni; * $p < 0.05$; - $p > 0.05$ (for example, **/* means one approach gave a Bonferroni-robust $p < 0.05$ and the other gave 0.05 not robust to Bonferroni). The PLM coefficient gives the average inferred change in gene count if an organism has a given property. The majority of traits give substantially higher p-values and lower-magnitude coefficients; plots are vertically truncated to focus on the more robust results.

Comparative analyses with challenging (oDNA) data



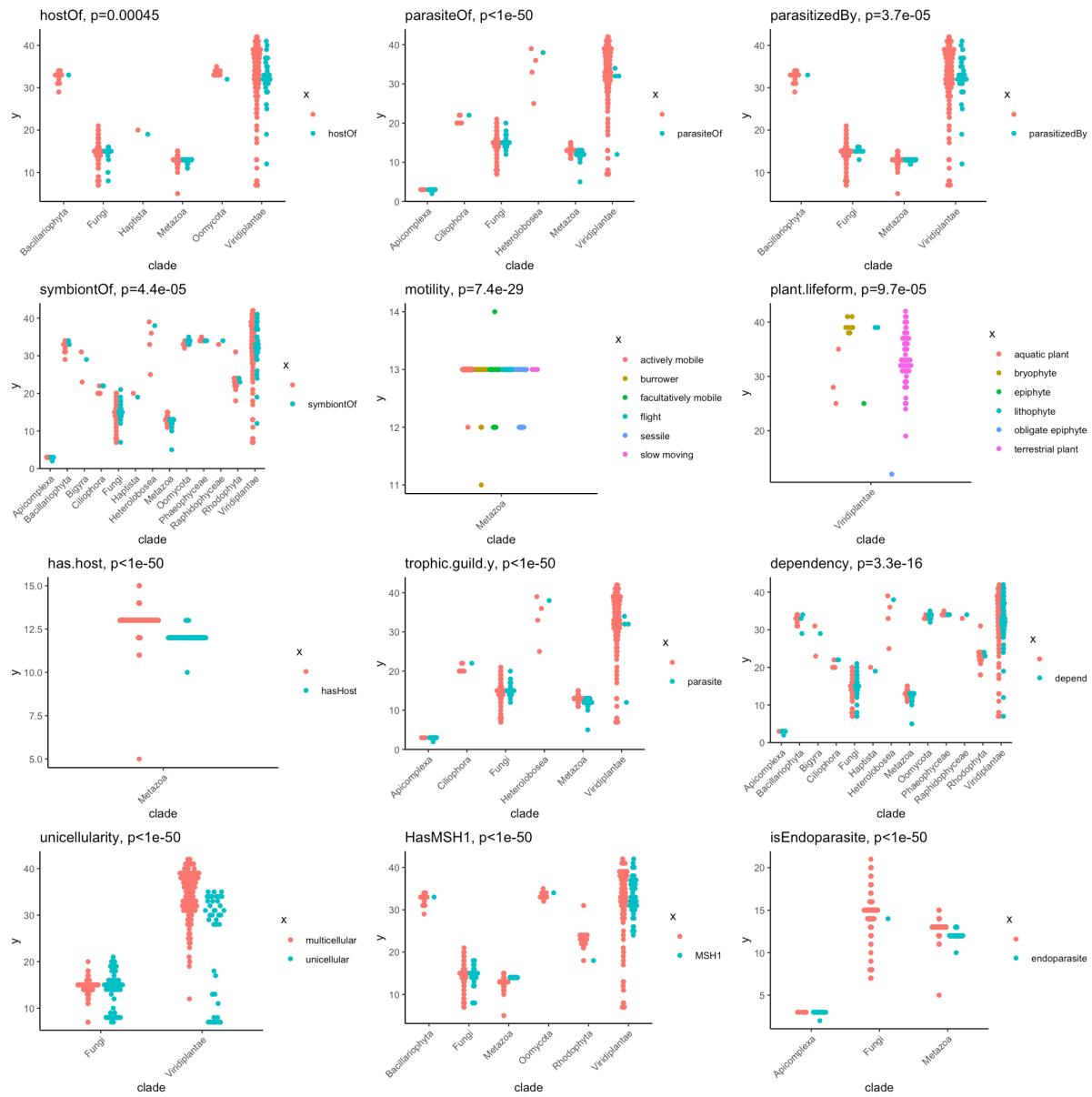
Supplementary Figure 11. mtDNA predictors in alternative plant phylogeny. Here, the plant phylogeny was adapted from the plant megaphylogeny found in (Jin and Qian 2022) comprised by phylogenies from (S. A. Smith and Brown 2018) and (Zanne et al. 2014). mtDNA correlates in plants-only phylogeny, using (top) explicit branch length estimates and (bottom) uniform branch lengths. Some connections differ from the cross-eukaryote case (although desert habitat and herbaceousness are retained), with several features connected with soil, shade, salt, and drought tolerance appearing in these data.

Comparative analyses with challenging (oDNA) data



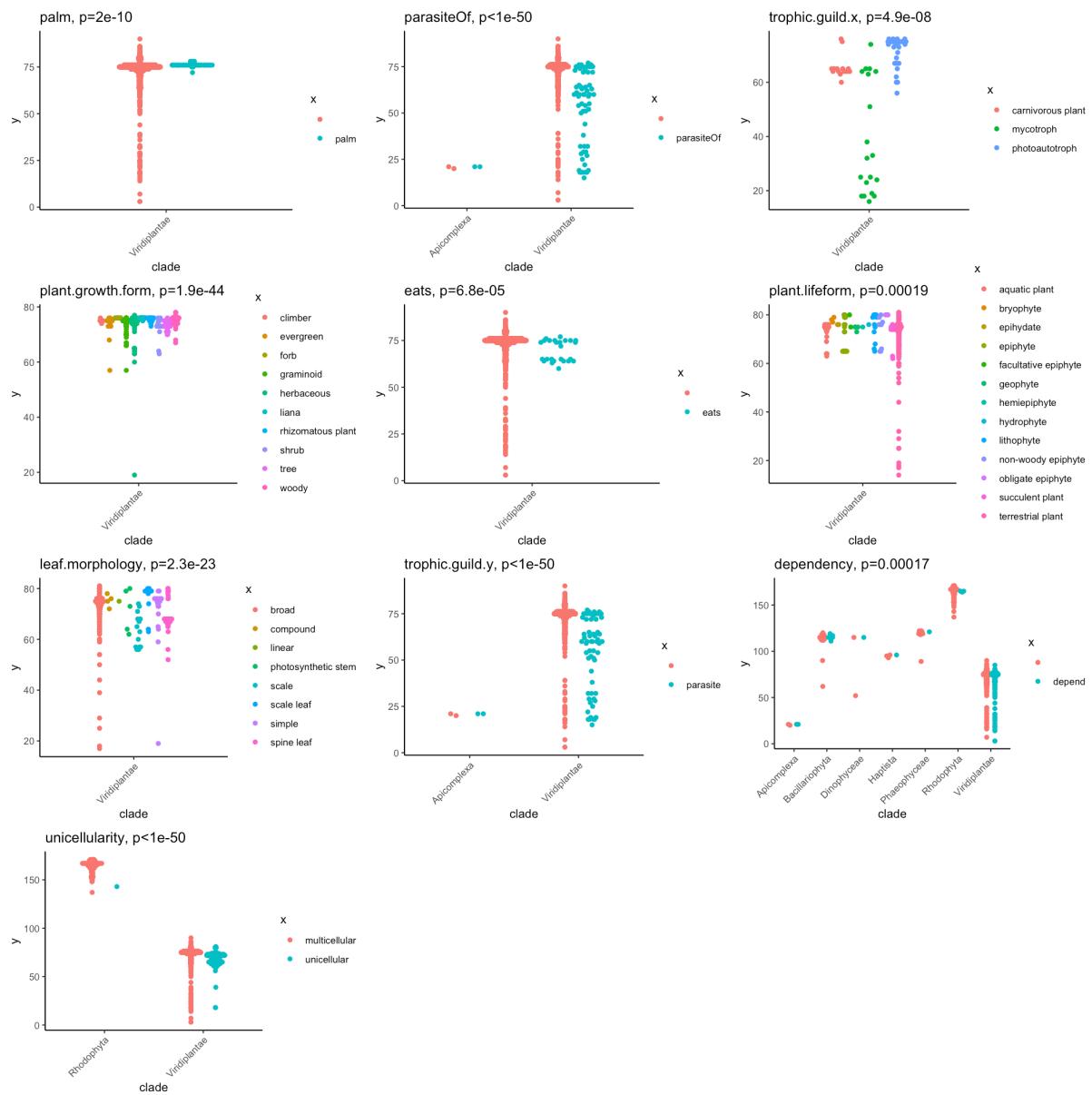
Supplementary Figure 12. ptDNA predictors in alternative plant phylogeny. Here, the plant phylogeny was adapted from the plant megaphylogeny found in (Jin and Qian 2022) comprised by phylogenies from (S. A. Smith and Brown 2018) and (Zanne et al. 2014). ptDNA correlates in plants-only phylogeny, using (top) explicit branch length estimates and (bottom) uniform branch lengths. Most connections are preserved from the cross-eukaryote case.

Comparative analyses with challenging (oDNA) data



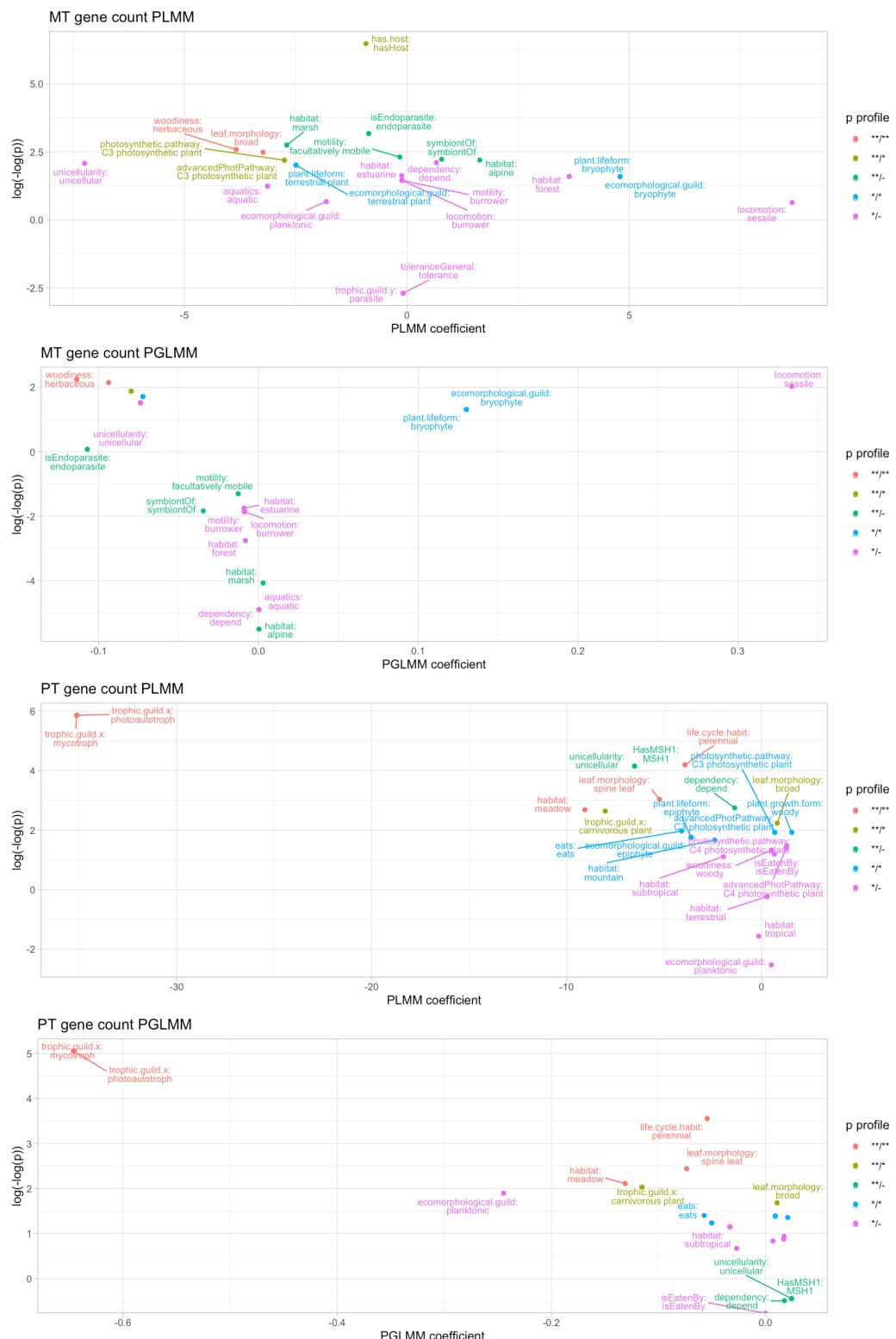
Supplementary Figure 13. mtDNA predictors using nonparametric, non-phylogenetic approach blocking eukaryotic clade. Horizontal axis organizes samples by eukaryotic clade; colour gives different levels of predictor factor; vertical axis gives gene count. Kruskal-Wallis or Scheirer-Ray-Hare tests were used to compare gene counts with different levels of the predictor factor, treating clade as a blocking factor where appropriate.

Comparative analyses with challenging (oDNA) data

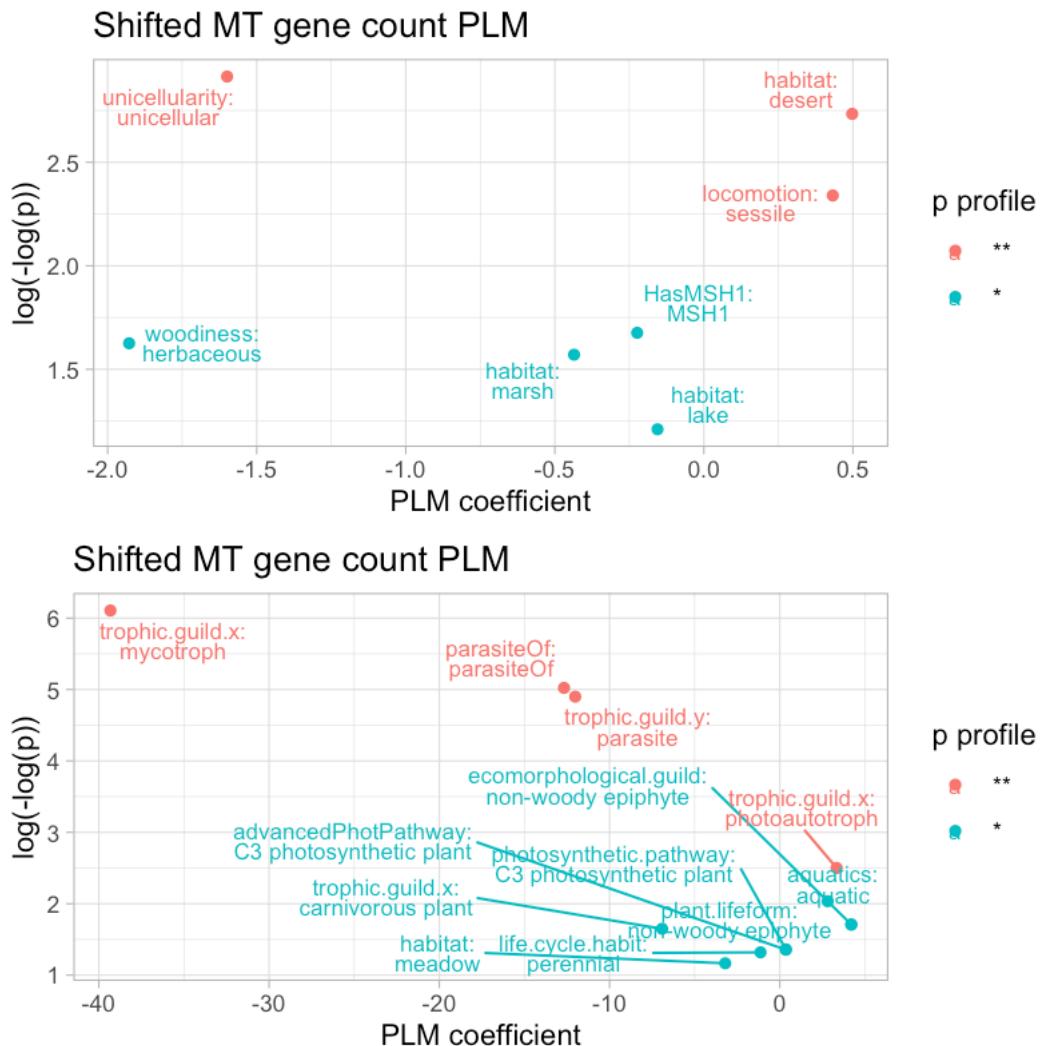


Supplementary Figure 14. ptDNA predictors using nonparametric, non-phylogenetic approach blocking eukaryotic clade. Horizontal axis organizes samples by eukaryotic clade; colour gives different levels of predictor factor; vertical axis gives gene count. Kruskal-Wallis or Scheirer-Ray-Hare tests were used to compare gene counts with different levels of the predictor factor, treating clade as a blocking factor where appropriate.

Comparative analyses with challenging (oDNA) data



Supplementary Figure 15. Mixed-model approaches assigning random effects to clade and analysing the remaining relationships. (top) mtDNA and (bottom) ptDNA plots, as in Supplementary Figs. 11-12, of coefficient and p-value for linear mixed model (LMM) and Poisson generalized linear mixed model (GLMM) with random effects associated with clade, applied to either intercepts or intercepts and slopes (selected by AIC).



Supplementary Figure 16. Phylogenetic linear model results after normalizing clade oDNA gene counts. The mean oDNA gene count in each clade is subtracted from all the members of that clade, then PLM is run as before, with results for mtDNA (top) and ptDNA (bottom) reported as in Fig. 5.

> palm: palm (2 Arecaceae)
> hasEctoparasite: hasEctoparasite (413 Metazoa)
> hasEndoparasite: hasEndoparasite (1018 Metazoa)
> hostOf: hostOf (1975 Eukaryota)
> isEatenBy: isEatenBy (730 Eukaryota), isEaten (2 Ulva)
> mutualisticWith: mutualisticWith (1755 Eukaryota)
> parasiteOf: parasiteOf (397 Eukaryota)
> parasitizedBy: parasitizedBy (1993 Eukaryota)
> symbiontOf: symbiontOf (1889 Eukaryota)
> trophic.guild.x: photoautotroph (4 Chlorophyta)
> locomotion: sessile (83 Eukaryota), arboreal locomotion (51 Chordata), burrower (33 Metazoa), aerial (15 Insecta), saltation (3 Mammalia)
> shade.tolerance: high tolerance (3 Magnoliopsida), low tolerance (16 Streptophyta), medium (normal) (4 Streptophyta)
> plant.growth.form: woody (39 Streptophyta), herbaceous (11 Streptophyta), graminoid (11 Poaceae)
> grass.growth.type: no (21 Streptophyta)
> life.cycle.habit: perennial (8 Lamiales)
> Mating.System: Not Monogamous (534 Mammalia), Monogamous (48 Mammalia)
> calcareous.soil.tolerance: medium (normal) (13 Streptophyta), none (5 Magnoliopsida), low tolerance (4 Streptophyta)
> eats: eats (2538 Eukaryota), preys_on (52 Eukaryota)
> motility: facultatively mobile (45 Metazoa), actively mobile (514 Metazoa), sessile (46 Metazoa), burrower (33 Metazoa), flight (64 Chiroptera), slow moving (3 Metazoa)
> ectoparasite.of: ectoparasite of (19 Arthropoda)
> ecomorphological.guild: terrestrial plant (84 Streptophyta), planktonic (484 Eukaryota), surficial modifier (20 Metazoa), fossorial and_or ground dwelling (298 Mammalia), nektobenthic (20 Metazoa), epifaunal (47 Metazoa), nectonic (115 Metazoa), fossorial and_or ground dwelling only (107 Chordata), arboreal (175 Mammalia), amphibious (23 Chordata), biodiffusor (9 Metazoa), symbiont (32 Eukaryota), infaunal (11 Bivalvia), algae (5 Bacillariophyta), substrate-attached (12 Mollusca), burrow builder (18 Metazoa), fossorial (6 Squamata), tubicolous (26 Metazoa), bryophyte (8 Streptophyta), semifossorial (2 Mammalia), aquatic adapted (36 Chordata), deep infaunal (4 Bivalvia), low-level epifaunal (2 Pectinidae), semi-infaunal (5 Mollusca), lithophyte (2 Streptophyta), aquatic plant (3 Alismatales)
> salt.tolerance: low tolerance (6 Magnoliopsida), high tolerance (5 Eukaryota), none (7 Streptophyta), medium (normal) (7 Magnoliopsida)
> drought.tolerance: medium (normal) (8 Streptophyta), low tolerance (10 Streptophyta), none (2 Magnoliopsida), high tolerance (4 Magnoliopsida)
> tissue.contains: high-magnesium calcite (61 Metazoa), calcium phosphate (220 Decapoda), aragonite (62 Mollusca), calcium carbonate (86 Metazoa), biogenic silica (20 Demospongiae), chitin (16 Arthropoda), latex (2 Campanulaceae), no (3 Onchidiidae)
> germination.requirements: no (19 Magnoliopsida), yes (3 Streptophyta)
> fire.tolerance: high tolerance (5 Streptophyta), low tolerance (2 Magnoliopsida), none (12 Streptophyta), medium (normal) (2 Magnoliopsida)
> developmental.mode: altricial (100 Aves), determinate growth (9 Acropora), indeterminate growth (7 Acroporidae), precocial (45 Aves), paedomorphic (3 Ambystoma), semiprecocial (29 Aves), larval development (2 Anura)

> Social.System: Social Group Living (248 Chordata), solitary (364 Chordata), No Harem (15 Mammalia), No Social Group Living (5 Mammalia), harem (4 Cercopithecidae)
> nitrogen.fixation: medium N fixation (2 Fabaceae)
> plant.lifeform: terrestrial plant (84 Streptophyta), bryophyte (8 Streptophyta), lithophyte (2 Streptophyta), aquatic plant (3 Alismatales)
> sexual.system: unisexual flowers (2 Cucurbitaceae)
> parasitoid.of: parasitoid of (3 Eukaryota), parasitoi of (13 Sarcophaga)
> leaf.morphology: broad (84 Streptophyta)
> photosynthetic.pathway: C3 photosynthetic plant (90 Streptophyta), C4 photosynthetic plant (4 Poaceae)
> anaerobic.soil.tolerance: none (8 Magnoliopsida), medium (normal) (6 Streptophyta), low tolerance (7 Magnoliopsida)
> trophic.level: omnivore (254 Metazoa), carnivore (261 Eukaryota), herbivore (207 Chordata)
> sexually.dimorphic: female larger (14 Aves), Sexes colored or patterned differently (15 Aves), male larger (26 Aves), male more colorful (15 Aves), no (34 Metazoa)
> woodiness: woody (39 Streptophyta), herbaceous (50 Streptophyta)
> has.host: hasHost (101 Metazoa)
> trophic.guild.y: parasite (399 Eukaryota)
> toleranceGeneral: tolerance (25 Eukaryota)
> aquatics: terrestrial (73 Chordata), aquatic (88 Eukaryota)
> dependency: depend (3171 Eukaryota)
> advancedPhotPathway: C3 photosynthetic plant (90 Streptophyta), C4 photosynthetic plant (4 Poaceae)
> unicellularity: multicellular (8959 Eukaryota), unicellular (162 Eukaryota)
> HasMSH1: MSH1 (142 Eukaryota)
> isEndoparasite: endoparasite (136 Eukaryota)
> isEctoparasite: ectoparasite (20 Arthropoda)
> water: None (73 Chordata), Freshwater (26 Chordata), Marine (3 Carnivora), Both (7 Salmonidae)
> migration: yes (17 Chordata), no (92 Chordata)
> thermalBiology: Endothermic (63 Chordata), Ectothermic (46 Chordata)
> hostAge: Adult (60 Chordata), Juvenile (2 Chordata)
> waveExposed: exposed (16 Scleractinia), sheltered (11 Scleractinia)
> habitat: sea sand (59 Metazoa), terrestrial (1024 Eukaryota), freshwater (708 Eukaryota), subtropical (347 Eukaryota), coastal (396 Eukaryota), marine (1128 Eukaryota), marine benthic (411 Eukaryota), hedge (3 Eukaryota), benthic (156 Metazoa), bathyal zone (26 Chordata), headwater (38 Eukaryota), tundra (7 Metazoa), forest (50 Eukaryota), tropical (564 Eukaryota), brackish water (80 Eukaryota), freshwater habitat (99 Eukaryota), desert (48 Eukaryota), oceanic benthopelagic zone (96 Actinopteri), lake (65 Eukaryota), pond (23 Eukaryota), savanna (22 Metazoa), montane forest (59 Eukaryota), valley (48 Eukaryota), mountain (153 Eukaryota), river (145 Eukaryota), tropical ocean (34 Metazoa), intertidal zone (31 Eukaryota), cave (35 Eukaryota), temperate (138 Eukaryota), outer continental shelf (69 Eukaryota), aquatic (17 Eukaryota), river valley (2 Aves), swamp (4 Metazoa), marsh (16 Eukaryota), pelagic zone (9 Metazoa), prairie (17 Eukaryota), marine reef (67 Metazoa), marine pelagic (46 Eukaryota), grassland (55 Eukaryota), estuarine (39 Metazoa), beach (27 Eukaryota), oceanic bathypelagic zone (97 Chordata), arid (49 Eukaryota), temperate ocean (15 Metazoa), lagoon (23

Metazoa), coral reef (6 Metazoa), woodland (10 Metazoa), river bank (32 Eukaryota), steppe (9 Chordata), cliff (4 Metazoa), inner continental shelf (12 Metazoa), coastal plain (4 Metazoa), broadleaf forest (3 Aves), wadi (4 Metazoa), saline water (6 Metazoa), rhizosphere (2 Eukaryota), anthropogenic (4 Eukaryota), hypersaline water (24 Metazoa), loam (6 Eukaryota), saline marsh (2 Metazoa), cold seep (2 Metazoa), ice (2 Metazoa), beach sand (6 Metazoa), riparian wetland (22 Chordata), alpine (15 Eukaryota), juniper woodland (6 Eukaryota), pasture (9 Eukaryota), tropical moist broadleaf forest (5 Eukaryota), arboreal habitat (4 Sciuridae), heath (5 Metazoa), scree (2 Chordata), alluvial soil (3 Metazoa), freshwater benthic (6 Eukaryota), sandy soil (5 Eukaryota), tropical humid forests (4 Eukaryota), kelp forest (2 Eukaryota), meadow (6 Metazoa), oak woodland (2 Eukaryota), fossorial and/or ground dwelling only (4 Mammalia), scrubland (8 Metazoa), karst (5 Eukaryota), mixed forest (2 Passeriformes), riffle (2 Chordata), crater lake (3 Chordata), littoral zone (3 Eukaryota), arboreal (locomotion) (2 Rodentia), continental slope (6 Chordata), mountain range (2 Chordata), oceanic pelagic zone (5 Actinopteri), forest soil (3 Eukaryota), hot spring (2 Eukaryota)
> intertidal: intertidal (71 Eukaryota)

Supplementary Table 1. Mitochondrial ecological predictors in our compiled database. Each row gives the name of the ecological factor and each non-empty level it takes in our database of species with mitochondrial genome sequences. Each factor level is accompanied by the number of positive matches in the database and the lowest taxonomic level beneath which these matches occur.

> palm: palm (41 Arecaceae)
> hasEndoparasite: hasEndoparasite (2 Semenovia)
> hostOf: hostOf (309 Eukaryota)
> isEatenBy: isEatenBy (490 Eukaryota)
> mutualisticWith: mutualisticWith (464 Eukaryota)
> parasiteOf: parasiteOf (64 Eukaryota)
> parasitizedBy: parasitizedBy (269 Eukaryota)
> symbiontOf: symbiontOf (390 Eukaryota)
> trophic.guild.x: mycotroph (19 Magnoliopsida), photoautotroph (36 Eukaryota), carnivorous plant (14 Magnoliopsida)
> life.span: high life span (72 Streptophyta), moderate life span (96 Streptophyta), low life span (53 Streptophyta)
> shade.tolerance: high tolerance (47 Streptophyta), medium (normal) (73 Streptophyta), low tolerance (131 Streptophyta)
> plant.growth.form: herbaceous (157 Streptophyta), evergreen (10 Streptophyta), graminoid (492 Poales), shrub (12 Magnoliopsida), woody (174 Magnoliopsida), tree (73 Magnoliopsida), rhizomatous plant (26 Magnoliopsida), climber (4 Magnoliopsida), liana (36 Magnoliopsida), forb (17 Magnoliopsida)
> grass.growth.type: no (226 Streptophyta), yes (25 Magnoliopsida)
> life.cycle.habit: perennial (340 Streptophyta), annual (4 Magnoliopsida)
> calcareous.soil.tolerance: medium (normal) (92 Streptophyta), low tolerance (71 Streptophyta), none (31 Streptophyta), high tolerance (48 Streptophyta)
> growth.rate: slow (41 Streptophyta), fast (129 Streptophyta), moderate (80 Streptophyta)
> eats: eats (25 Magnoliopsida)
> ecomorphological.guild: terrestrial plant (1529 Streptophyta), symbiont (3 Magnoliopsida), planktonic (33 Eukaryota), succulent plant (12 Magnoliopsida), aquatic plant (25 Magnoliopsida), non-woody epiphyte (9 Streptophyta), lithophyte (11 Streptophyta), epiphyte (26 Streptophyta), mangrove biome (14 Magnoliopsida), algae (14 Eukaryota), geophyte (4 Streptophyta), facultative epiphyte (2 Streptophyta), hydrophyte (2 Streptophyta), epiphydate (2 Nymphaea), bryophyte (3 Streptophyta), obligate epiphyte (3 Polypodiopsida)
> salt.tolerance: high tolerance (24 Eukaryota), medium (normal) (58 Streptophyta), none (103 Streptophyta), low tolerance (69 Streptophyta)
> drought.tolerance: high tolerance (73 Streptophyta), none (26 Streptophyta), low tolerance (92 Streptophyta), medium (normal) (86 Streptophyta)
> tissue.contains: latex (109 Magnoliopsida)
> germination.requirements: yes (84 Streptophyta), no (163 Streptophyta)
> fire.tolerance: high tolerance (69 Streptophyta), low tolerance (57 Streptophyta), medium (normal) (60 Streptophyta), none (47 Streptophyta)
> nitrogen.fixation: medium N fixation (11 Magnoliopsida), high N fixation (8 Magnoliopsida), low N fixation (6 Magnoliopsida)
> plant.lifeform: terrestrial plant (1549 Streptophyta), succulent plant (12 Magnoliopsida), aquatic plant (25 Magnoliopsida), non-woody epiphyte (9 Streptophyta), lithophyte (11 Streptophyta), epiphyte (26 Streptophyta), geophyte (4 Streptophyta), facultative epiphyte (2 Streptophyta), epiphydate (2 Nymphaea), bryophyte (3 Streptophyta), obligate epiphyte (3 Polypodiopsida)
> sexual.system: unisexual flowers (22 Magnoliopsida), dioecious (18 Streptophyta), monoecious (7 Magnoliopsida), duodichogamous (3 Sapindaceae)

- > leaf.morphology: broad (1513 Streptophyta), simple (31 Streptophyta), spine leaf (53 Streptophyta), scale (14 Selaginella), photosynthetic stem (5 Streptophyta), scale leaf (16 Streptophyta), compound (4 Magnoliopsida)
- > photosynthetic.pathway: C3 photosynthetic plant (1421 Streptophyta), CAM photosynthetic plant (16 Streptophyta), C4 photosynthetic plant (144 Magnoliopsida)
- > propagule: fruit (14 Streptophyta), seed (10 Streptophyta)
- > anaerobic.soil.tolerance: high tolerance (43 Streptophyta), none (115 Streptophyta), medium (normal) (34 Streptophyta), low tolerance (55 Streptophyta)
- > woodiness: herbaceous (726 Streptophyta), woody (1039 Streptophyta), semi-woody (16 Magnoliopsida)
- > trophic.guild.y: parasite (63 Eukaryota)
- > toleranceGeneral: tolerance (287 Eukaryota)
- > aquatics: aquatic (48 Eukaryota)
- > dependency: depend (727 Eukaryota)
- > advancedPhotPathway: C3 photosynthetic plant (1418 Streptophyta), CAM photosynthetic plant (11 Streptophyta), C3_and_CAM (7 Streptophyta), C4 photosynthetic plant (139 Magnoliopsida), C3_and_C4 (9 Magnoliopsida)
- > unicellularity: multicellular (4047 Eukaryota), unicellular (171 Eukaryota)
- > HasMSH1: MSH1 (151 Eukaryota)
- > habitat: subtropical (42 Viridiplantae), mountain (53 Streptophyta), tropical (110 Eukaryota), terrestrial (366 Eukaryota), desert (23 Streptophyta), meadow (12 Streptophyta), temperate (93 Streptophyta), river (16 Streptophyta), coastal (46 Viridiplantae), valley (8 Streptophyta), marine (31 Eukaryota), outer continental shelf (4 Ulva), freshwater habitat (13 Eukaryota), anthropogenic (3 Magnoliopsida), lake (10 Magnoliopsida), freshwater benthic (3 Eukaryota), freshwater (17 Eukaryota), tropical moist broadleaf forest (6 Viridiplantae), forest (4 Streptophyta), headwater (6 Streptophyta), coastal plain (3 Streptophyta), intertidal zone (2 Magnoliopsida), arid (21 Streptophyta), montane forest (6 Streptophyta), montane grassland (5 Streptophyta), marine benthic (7 Eukaryota), river bank (4 Magnoliopsida), terra firme (5 Eukaryota), marsh (4 Magnoliopsida), grassland (12 Magnoliopsida), prairie (45 Magnoliopsida), tropical humid forests (2 Magnoliopsida), steppe (2 Magnoliopsida), savanna (4 Streptophyta), tropical semideciduous forest (3 Magnoliopsida), wetland (5 Magnoliopsida), pasture (3 Magnoliopsida), saline water (2 Streptophyta), loam (2 Magnoliopsida), tropical and subtropical moist broadleaf forests (2 Melastomataceae), alpine (8 Magnoliopsida), humus (15 Magnoliopsida), aquatic (2 Eukaryota), hedge (5 Magnoliopsida), heath (4 Magnoliopsida), aquatic vegetation (2 Poaceae), alluvial soil (2 Magnoliopsida), pond (4 Streptophyta), beach (3 Eukaryota), marine pelagic (2 Rhizosolenia), rocky grassland (2 Delesseriaceae)
- > intertidal: intertidal (6 Eukaryota)

Supplementary Table 2. Plastid ecological predictors in our compiled database. Each row gives the name of the ecological factor and each non-empty level it takes in our database of species with plastid genome sequences. Each factor level is accompanied by the number of positive matches in the database and the lowest taxonomic level beneath which these matches occur.