

Title

The relationship between diet, plasma glucose, and cancer prevalence across vertebrates

Authors

Stefania E. Kapsetaki^{1,2,3,4,5,*}, Anthony J. Basile^{6,7,8}, Zachary T. Compton^{1,9,10}, Shawn M. Rupp^{1,2}, Elizabeth G. Duke^{1,11,12}, Amy M. Boddy^{1,12,13}, Tara M. Harrison^{1,11,12}, Karen L. Sweazea^{6,7,14}, Carlo C. Maley^{1,2,14}

These authors jointly supervised this work: Karen L. Sweazea, Carlo C. Maley.

* Corresponding author

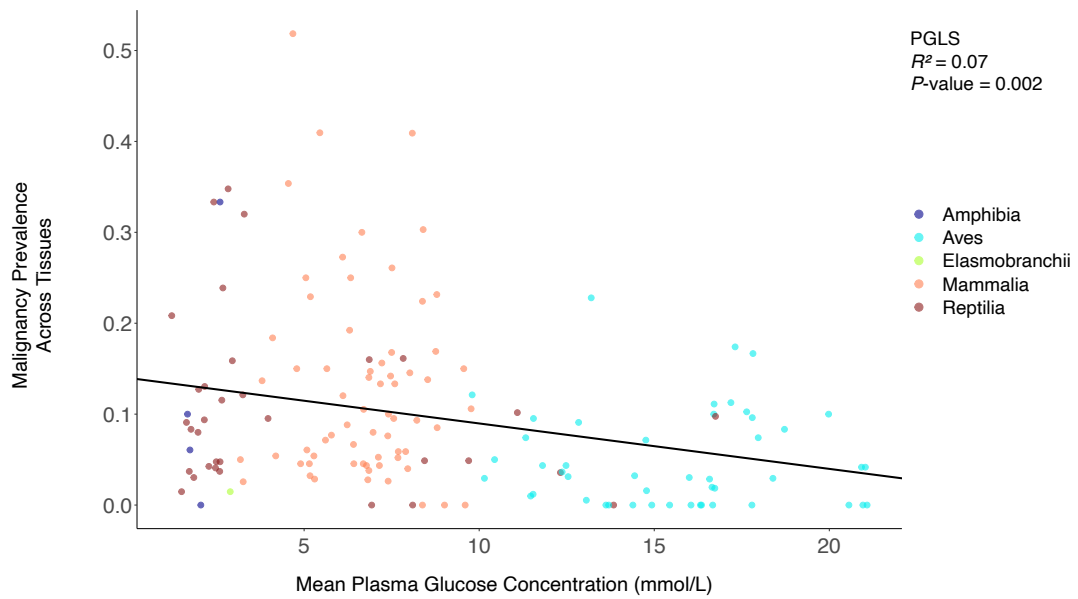
email: std534260@ac.eap.gr

Affiliations

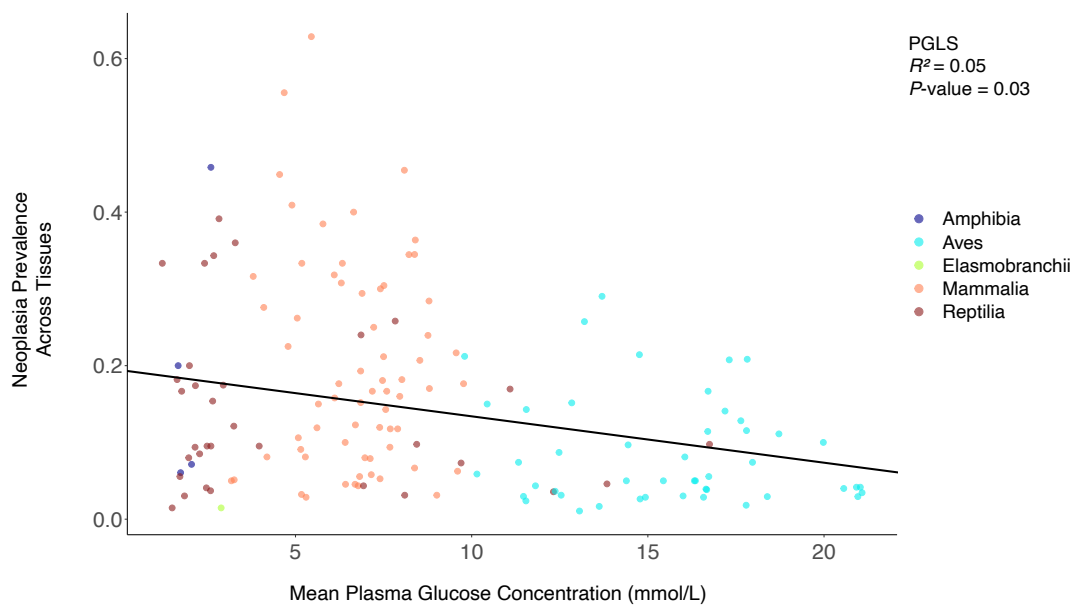
¹ Arizona Cancer Evolution Center, Arizona State University, Tempe, AZ, USA. ² Center for Biocomputing, Security and Society, Biodesign Institute, Arizona State University, Tempe, AZ, USA. ³ Tufts University, School of Arts and Sciences, Department of Biology, Medford, MA, USA. ⁴ Frederick University, School of Health Sciences, Department of Pharmacology, Nicosia, Cyprus. ⁵ Hellenic Open University, Patras, Greece. ⁶ Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA. ⁷ College of Health Solutions, Arizona State University, Phoenix, AZ, USA. ⁸ Department of Human Ecology, State University of New York (SUNY) at Oneonta, Oneonta, NY 13820 USA. ⁹ University of Arizona Cancer Center, Tucson, AZ, USA. ¹⁰ University of Arizona College of Medicine, Tucson, AZ, USA. ¹¹ Department of Clinical Sciences, North Carolina State University, Raleigh, NC, 27607 USA. ¹² Exotic Species Cancer Research Alliance, North Carolina State University, Raleigh, NC, 27607 USA. ¹³ Department of Anthropology, University of California Santa Barbara, CA, USA. ¹⁴ School of Life Sciences, 427 East Tyler Mall, Arizona State University, Tempe, Arizona, USA.

Supplementary Information

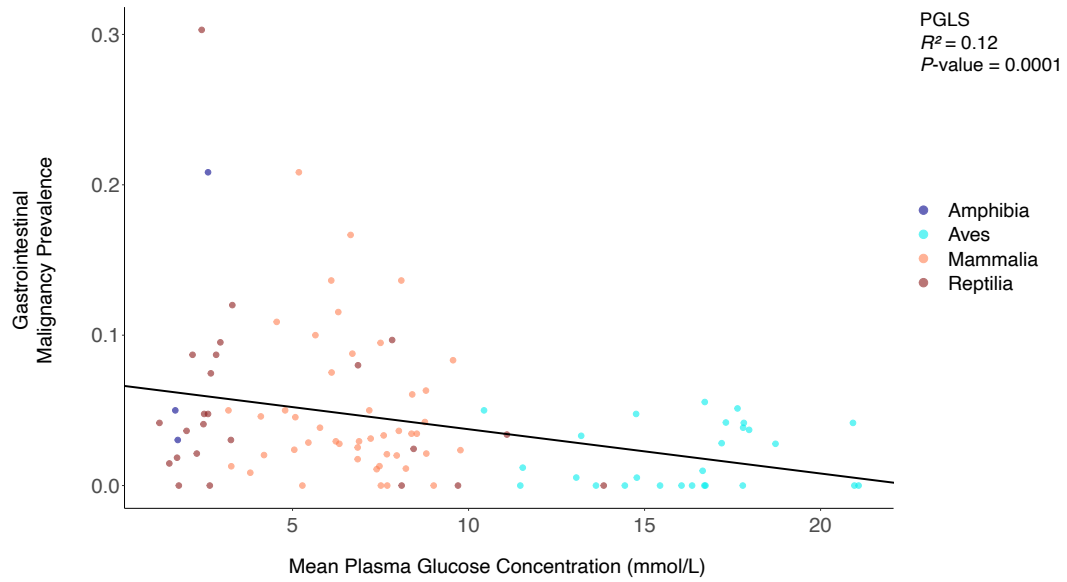
A



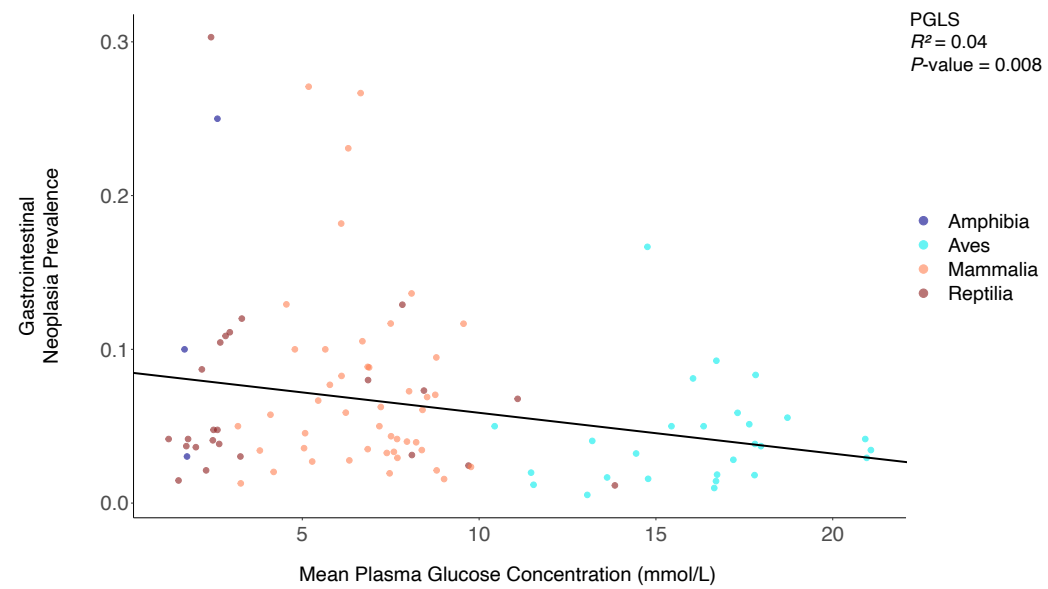
B



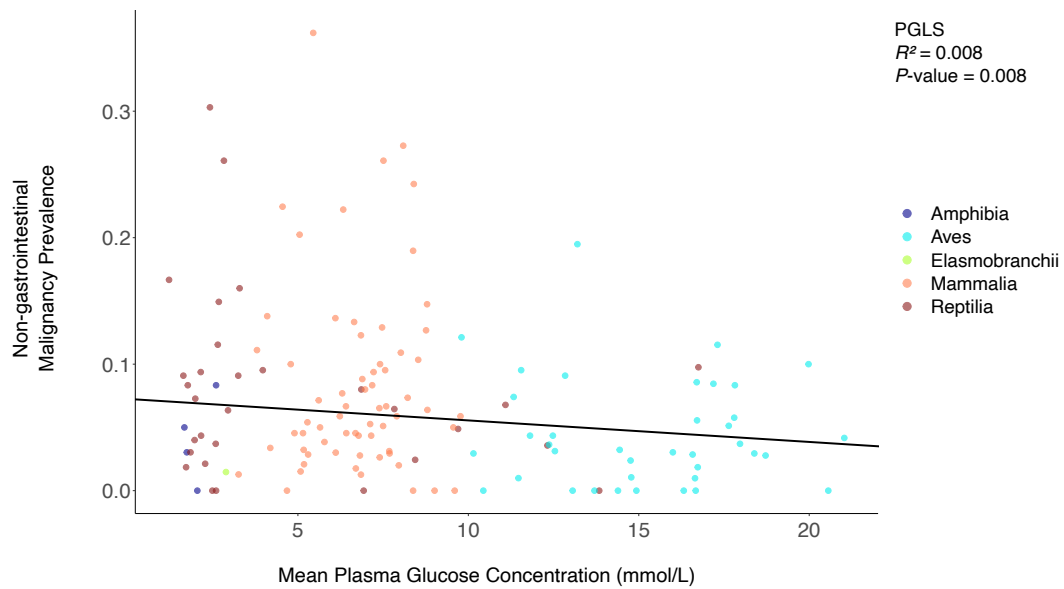
C



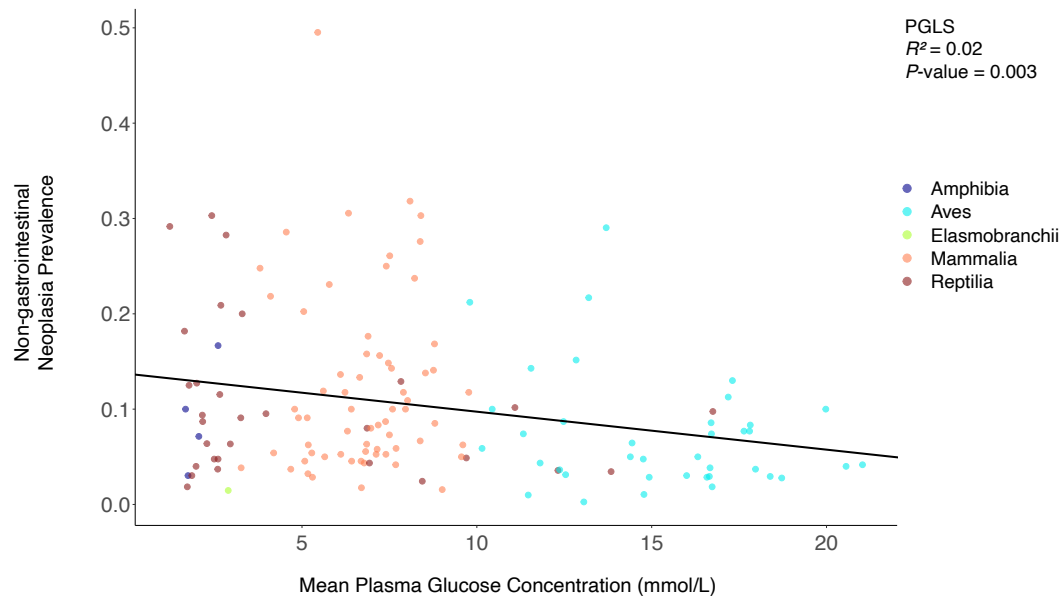
D



E

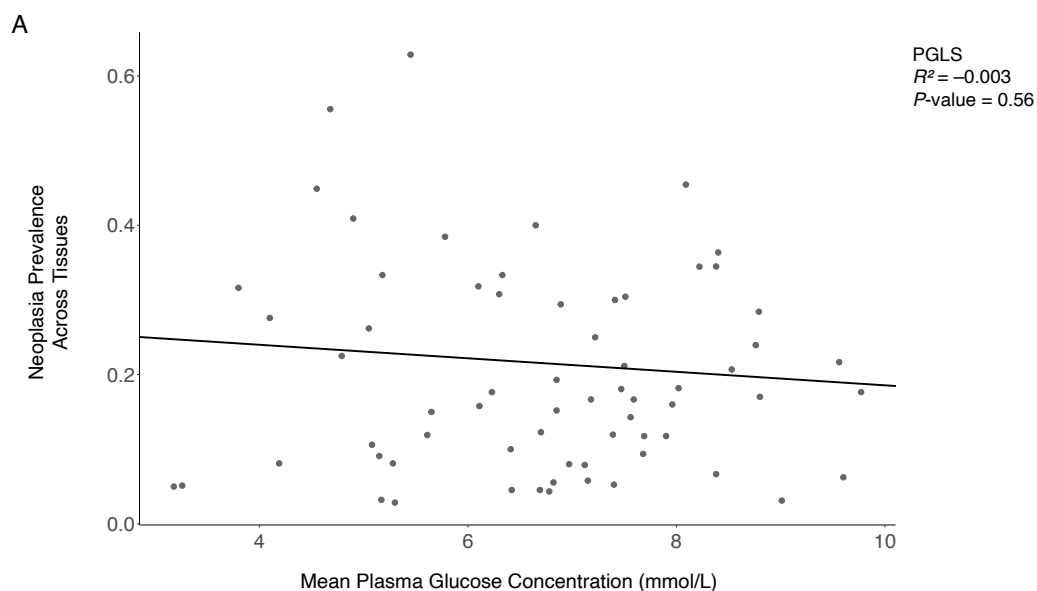


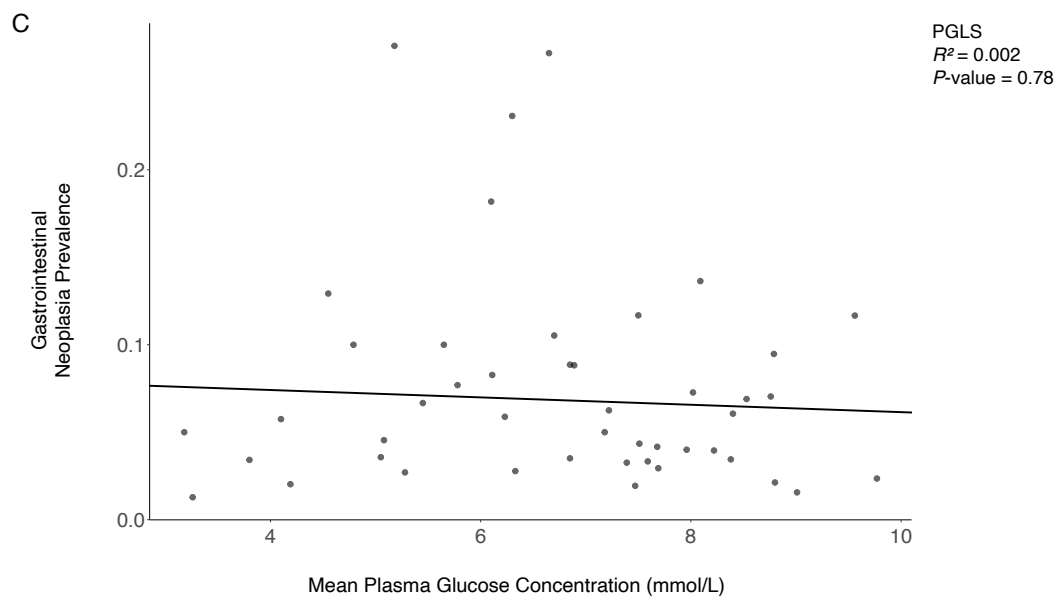
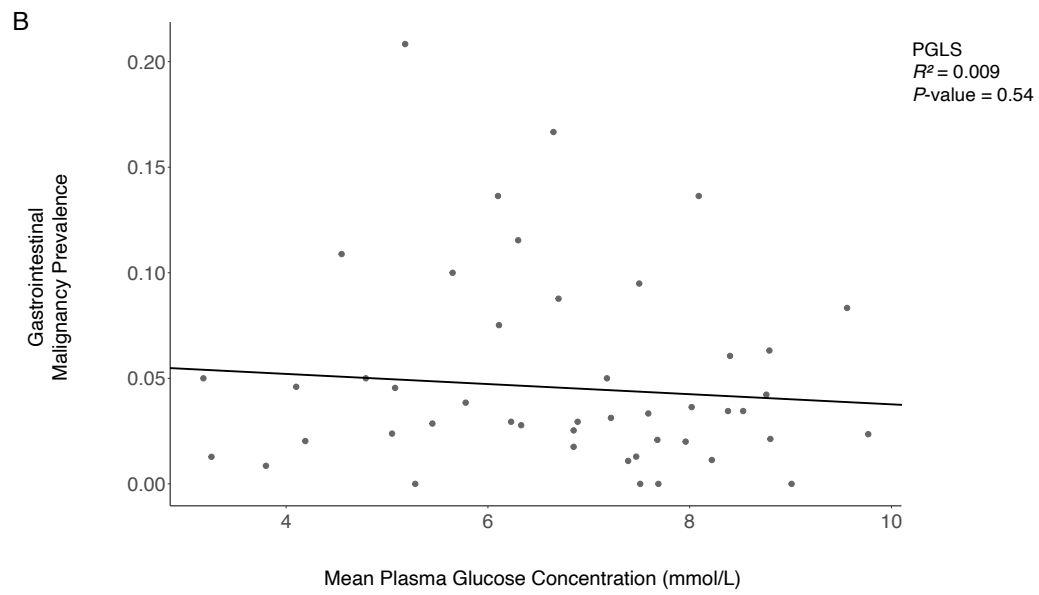
F

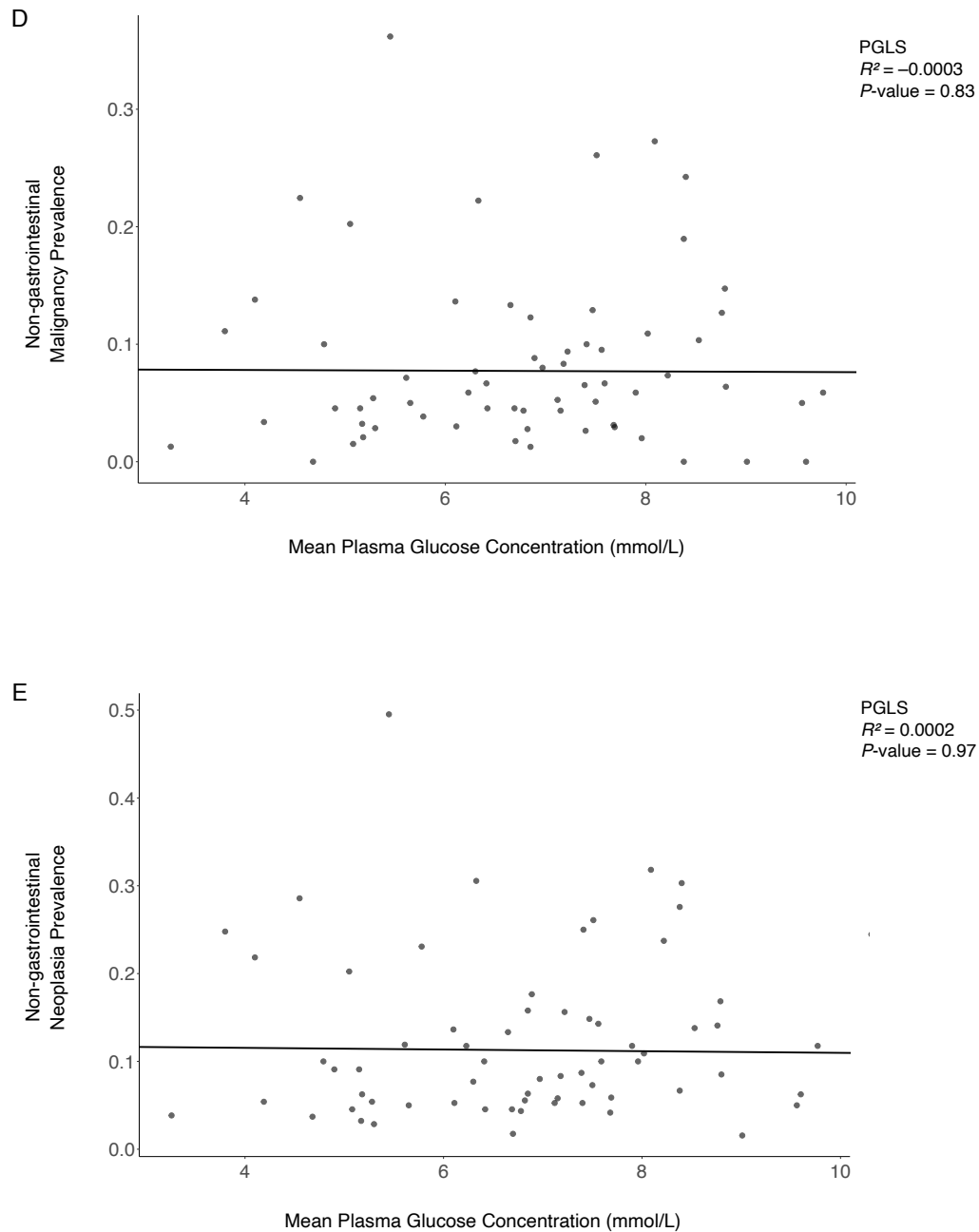


Supplementary Figure 1. Relationship between malignancy prevalence or neoplasia prevalence and mean plasma glucose concentration across vertebrates. (A) Mean plasma glucose concentration is negatively correlated with malignancy prevalence across tissues (151 species; PGLS: $P\text{-value} = 0.002$). (B) Mean plasma glucose concentration is not correlated with neoplasia prevalence across tissues (151 species; PGLS: $P\text{-value} = 0.03$). (C) Mean plasma glucose concentration is negatively correlated with gastrointestinal malignancy prevalence across 100 species (PGLS: $P\text{-value} = 0.0001$). (D) Mean plasma glucose concentration is negatively correlated with gastrointestinal neoplasia prevalence across 100 species (PGLS: $P\text{-value} = 0.008$). (E) Mean plasma glucose concentration is

negatively correlated with non-gastrointestinal malignancy prevalence across 138 species (PGLS: P -value = 0.008), and (F) non-gastrointestinal neoplasia prevalence across 138 species (PGLS: P -value = 0.003). All five negative correlations (A, C-F) remained when controlling for body mass, but did not remain when controlling for birds versus mammals versus ectotherms (Supplementary Data 2). Each dot represents the malignancy prevalence across tissues (A), the neoplasia prevalence across tissues (B), the gastrointestinal malignancy prevalence (C), the gastrointestinal neoplasia prevalence (D), the non-gastrointestinal malignancy prevalence (E), the non-gastrointestinal neoplasia prevalence (F), and the average plasma glucose concentration of one species; Amphibia: Dark Blue; Aves: Blue; Elasmobranchii: Green; Mammalia: Orange; Reptilia: Red. The outlier species (Rosner's test) with the highest malignancy or neoplasia prevalence depending on the panel are: *Didelphis virginiana*, *Mustela putorius*, and *Panthera onca* in panel A from top to bottom; *Mustela putorius* in panel B; *Lampropeltis getula*, *Leptodactylus fallax*, *Varecia variegata*, and *Petaurus breviceps* in panel C from top to bottom; *Lampropeltis getula*, *Varecia variegata*, *Petaurus breviceps*, *Leptodactylus fallax*, *Ursus maritimus*, and *Vulpes zerda* in panel D from top to bottom; *Mustela putorius*, *Lampropeltis getula*, *Panthera onca*, *Lampropeltis triangulum*, and *Puma concolor* in panel E from top to bottom; and *Mustela putorius* in panel F. The statistical tests were two-sided. P -values have not been adjusted for FDR corrections. The P -values of pgl analyses that passed the FDR correction are noted with an asterisk in Supplementary Data 2. PGLS: phylogenetic generalised least squares. For information on how to generate this figure, please see Source Data file.

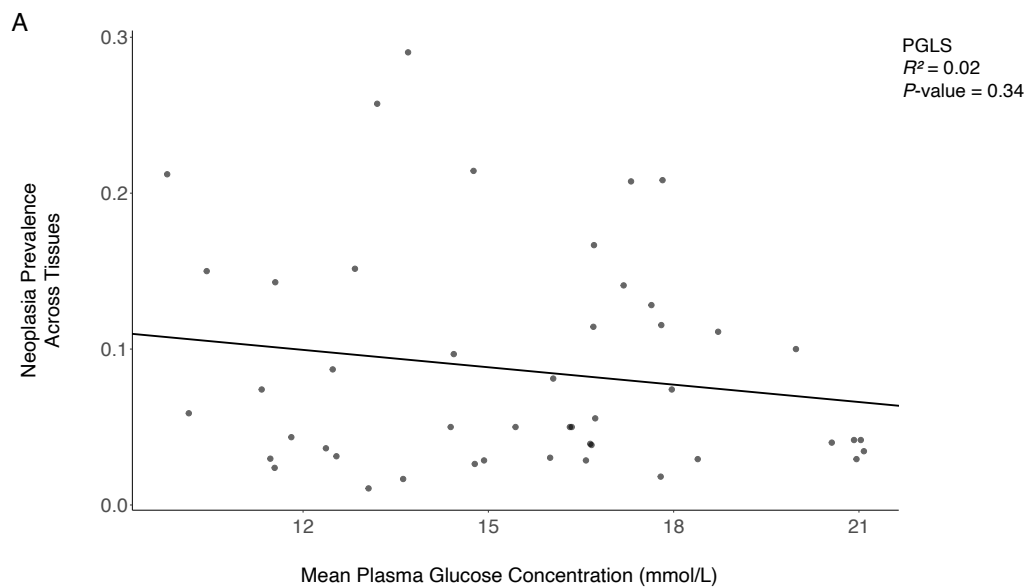


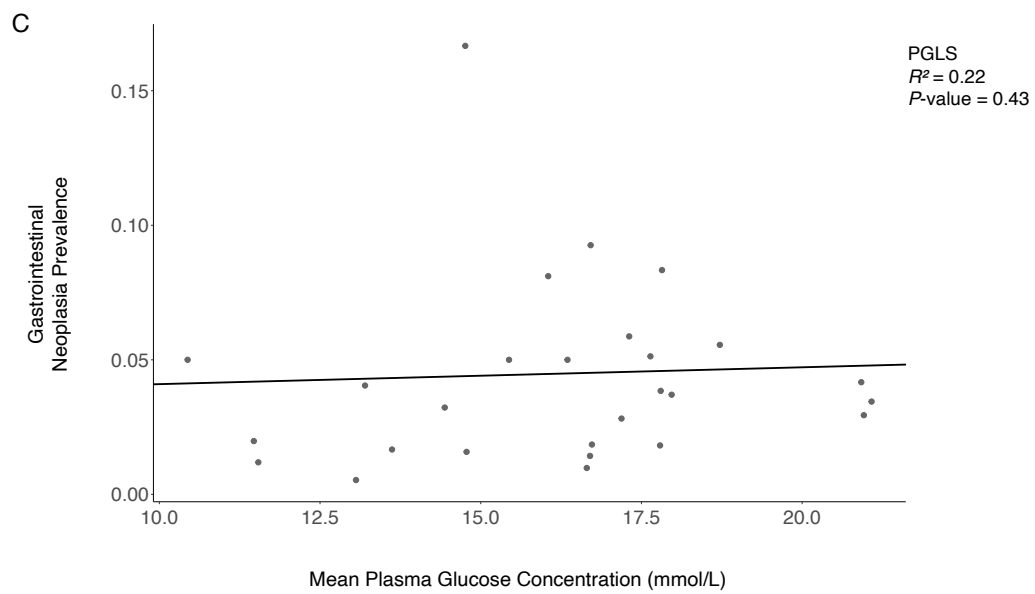
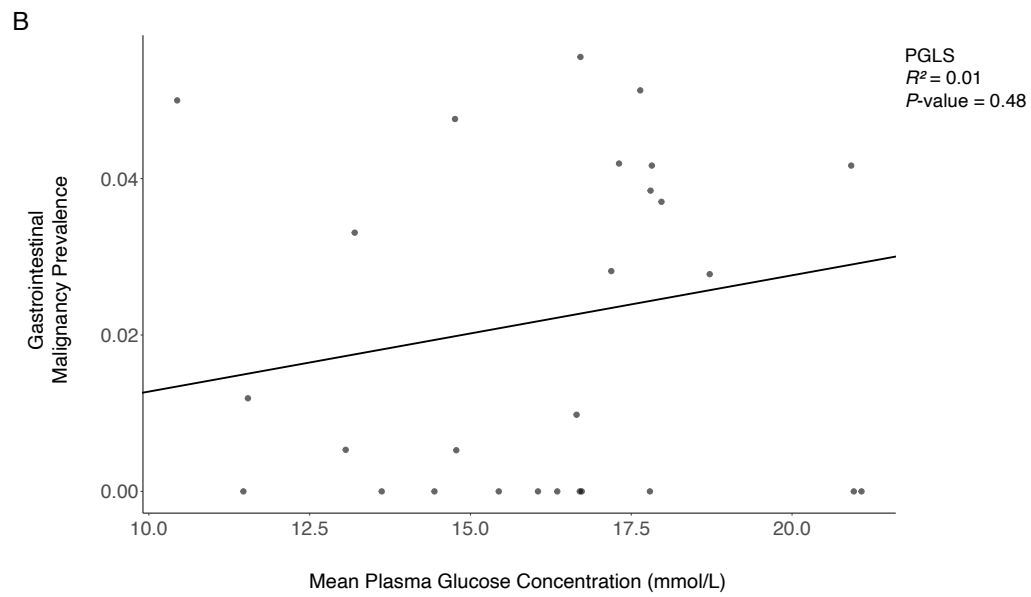


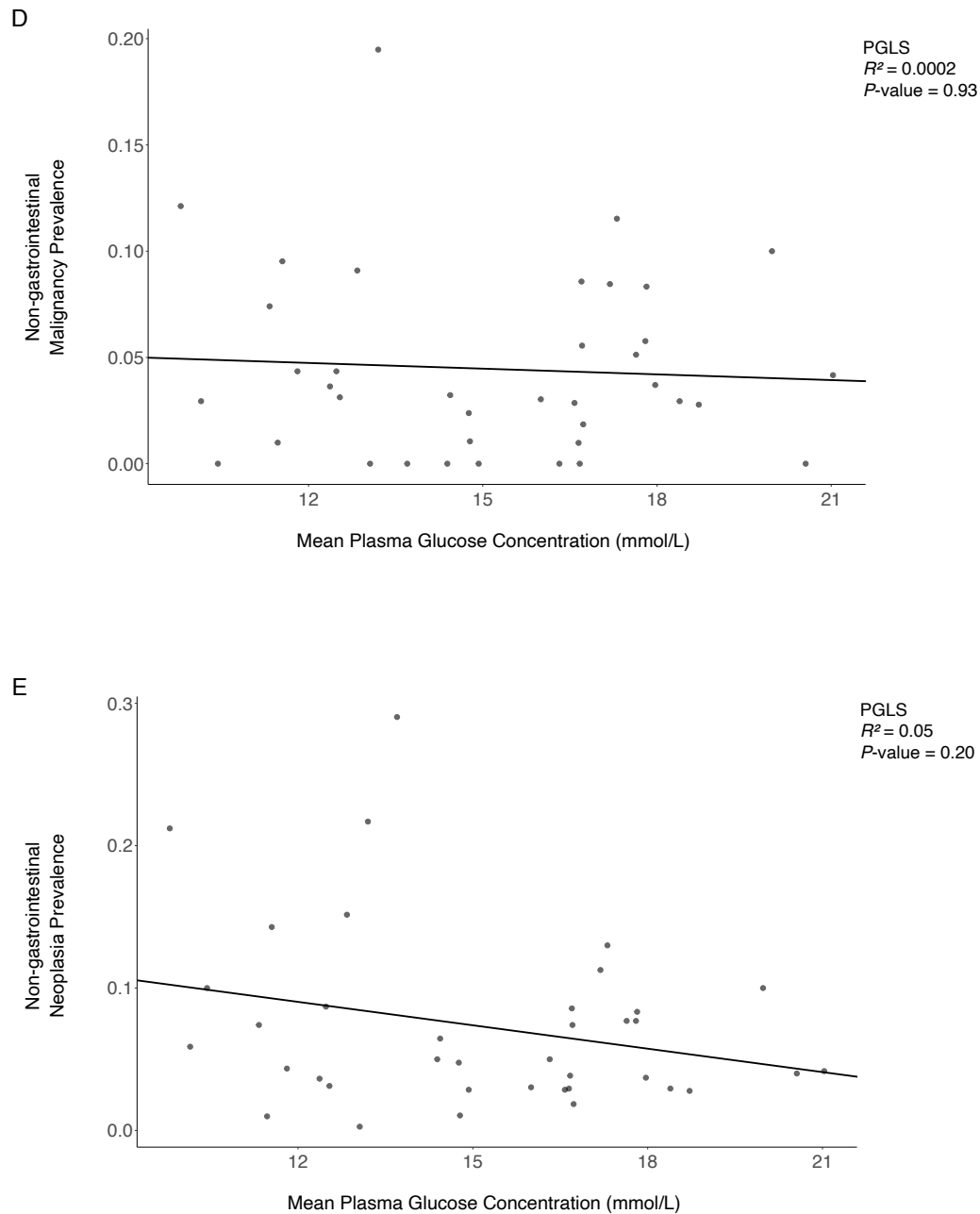


Supplementary Figure 2. No detectable correlation between cancer or neoplasia prevalence and mean plasma glucose concentration in mammals. There is no correlation between mean plasma glucose concentration and (A) neoplasia prevalence across tissues for 66 mammalian species, (B) gastrointestinal malignancy prevalence for 46 mammalian species, (C) gastrointestinal neoplasia prevalence for 46 mammalian species, (D) non-gastrointestinal malignancy prevalence for 65 mammalian species, and (E) non-gastrointestinal neoplasia prevalence for 65 mammalian species (PGLS: $P\text{-value} > 0.05$). Each dot shows the neoplasia prevalence across tissues (A), the

gastrointestinal malignancy prevalence (B), the gastrointestinal neoplasia prevalence (C), the non-gastrointestinal malignancy prevalence (D), the non-gastrointestinal neoplasia prevalence (E), and the average plasma glucose concentration of one species. The outlier species (Rosner's test) with the highest malignancy or neoplasia prevalence depending on the panel are: *Varecia variegata* in panel B; *Varecia variegata*, *Petaurus breviceps*, *Ursus maritimus*, and *Vulpes zerda* in panel C from top to bottom; and *Mustela putorius* in panels D and E. The statistical tests were two-sided. *P*-values have not been adjusted for FDR corrections. The *P*-values of pglis analyses that passed the FDR correction are noted with an asterisk in Supplementary Data 3. PGLS: phylogenetic generalised least squares. For information on how to generate this figure, please see Source Data file.

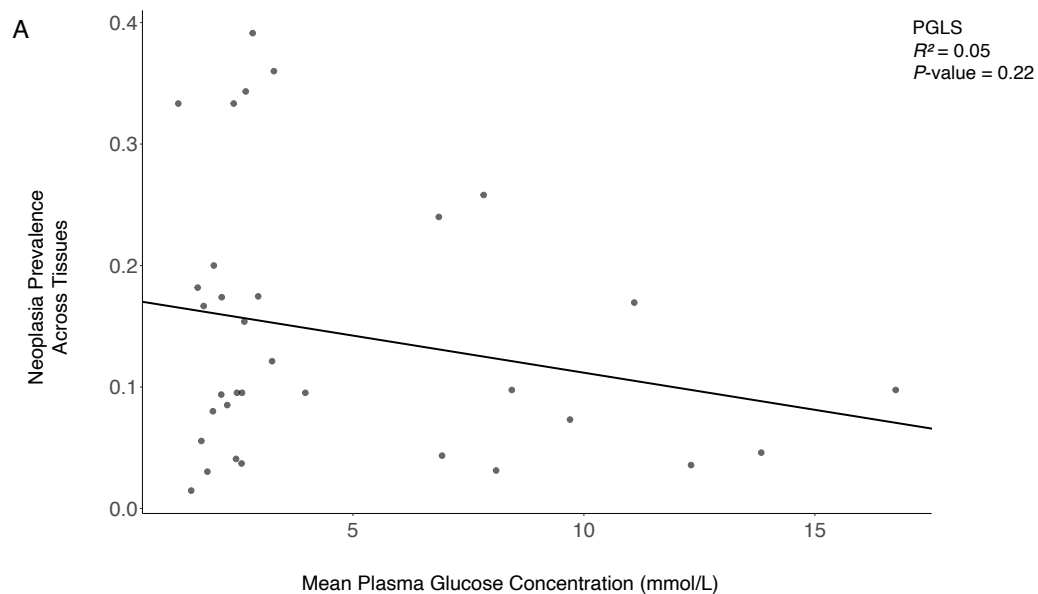


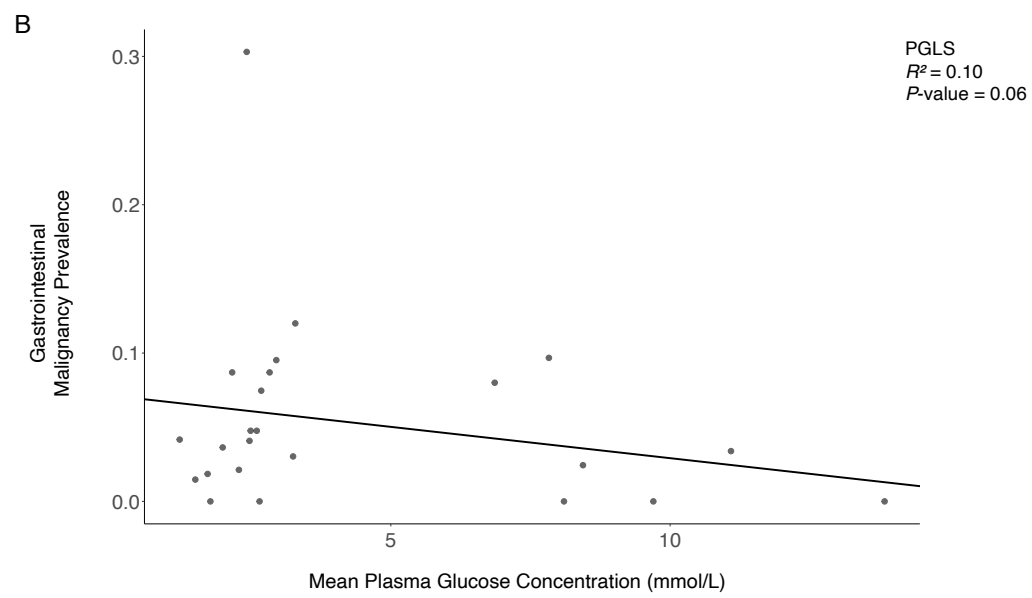


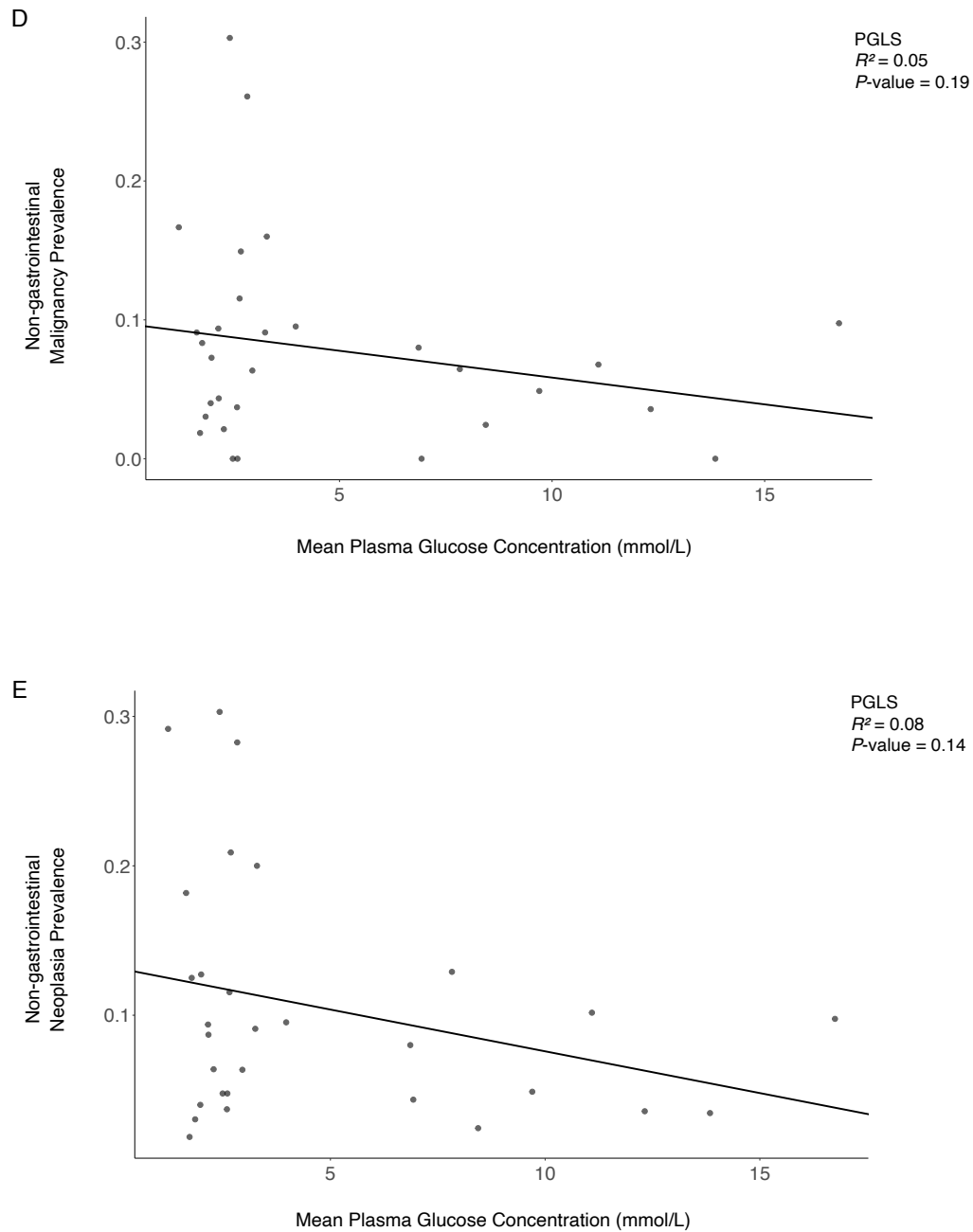


Supplementary Figure 3. No detectable correlation between cancer or neoplasia prevalence and mean plasma glucose concentration in birds. There is no correlation between mean plasma glucose concentration and (A) neoplasia prevalence across tissues for 47 bird species, (B) gastrointestinal malignancy prevalence for 27 bird species, (C) gastrointestinal neoplasia prevalence for 27 bird species, (D) non-gastrointestinal malignancy prevalence for 38 bird species, and (E) non-gastrointestinal neoplasia prevalence for 38 bird species (PGLS: $P\text{-value} > 0.05$). Each dot shows the neoplasia prevalence across tissues (A), the gastrointestinal malignancy prevalence (B), the

gastrointestinal neoplasia prevalence (C), the non-gastrointestinal malignancy prevalence (D), the non-gastrointestinal neoplasia prevalence (E), and the average plasma glucose concentration of one species. The outlier species (Rosner's test) with the highest malignancy or neoplasia prevalence depending on the panel are: *Ardeola speciosa* in panel C; *Gallus gallus* in panel D; and *Platalea ajaja*, *Gallus gallus*, and *Anas platyrhynchos* in panel E from top to bottom. The statistical tests were two-sided. *P*-values have not been adjusted for FDR corrections. The *P*-values of pgls analyses that passed the FDR correction are noted with an asterisk in Supplementary Data 4. PGLS: phylogenetic generalised least squares. For information on how to generate this figure, please see Source Data file.



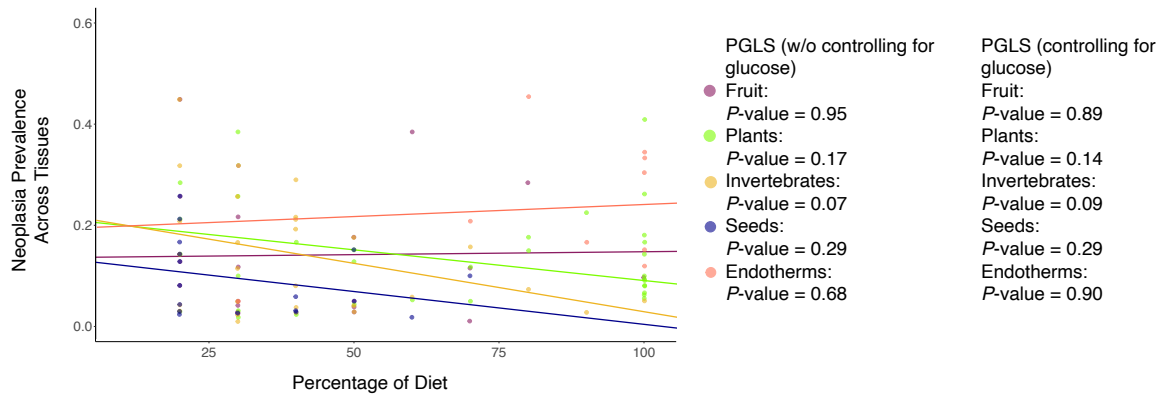




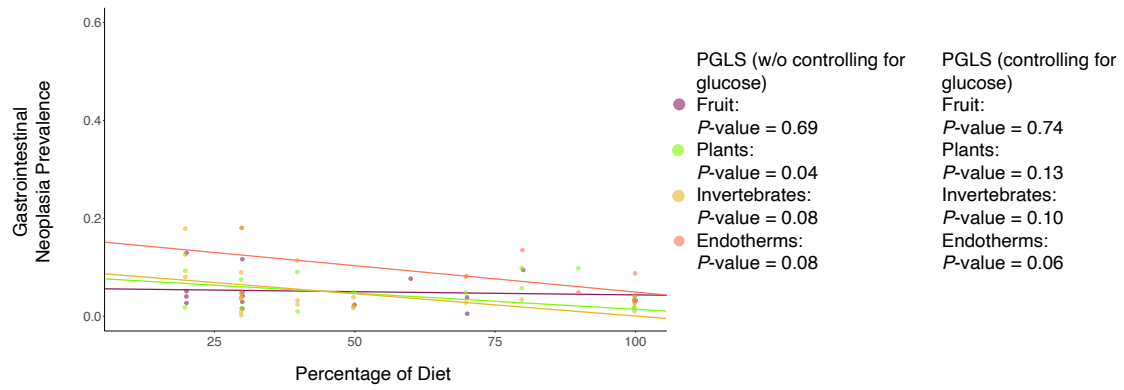
Supplementary Figure 4. No detectable correlation between cancer or neoplasia prevalence and mean plasma glucose concentration in reptiles. There is no correlation between mean plasma glucose concentration and (A) neoplasia prevalence across tissues for 33 reptile species, (B) gastrointestinal malignancy prevalence for 24 reptile species, (C) gastrointestinal neoplasia prevalence for 24 reptile species, (D) non-gastrointestinal malignancy prevalence for 30 reptile species, and (E) non-gastrointestinal neoplasia prevalence for 30 reptile species (PGLS: $P\text{-value} > 0.05$). Each dot shows the neoplasia prevalence across tissues (A), the gastrointestinal malignancy prevalence (B), the

gastrointestinal neoplasia prevalence (C), the non-gastrointestinal malignancy prevalence (D), the non-gastrointestinal neoplasia prevalence (E), and the average plasma glucose concentration of one species. The outlier species (Rosner's test) with the highest malignancy or neoplasia prevalence depending on the panel are: *Lampropeltis getula* in panels B and C; *Lampropeltis getula* and *Lampropeltis triangulum* in panel D from top to bottom; and *Lampropeltis getula*, *Python bivittatus*, and *Lampropeltis triangulum* in panel E from top to bottom. The statistical tests were two-sided. P-values have not been adjusted for FDR corrections. The P-values of pgls analyses that passed the FDR correction are noted with an asterisk in Supplementary Data 5. PGLS: phylogenetic generalised least squares. For information on how to generate this figure, please see Source Data file.

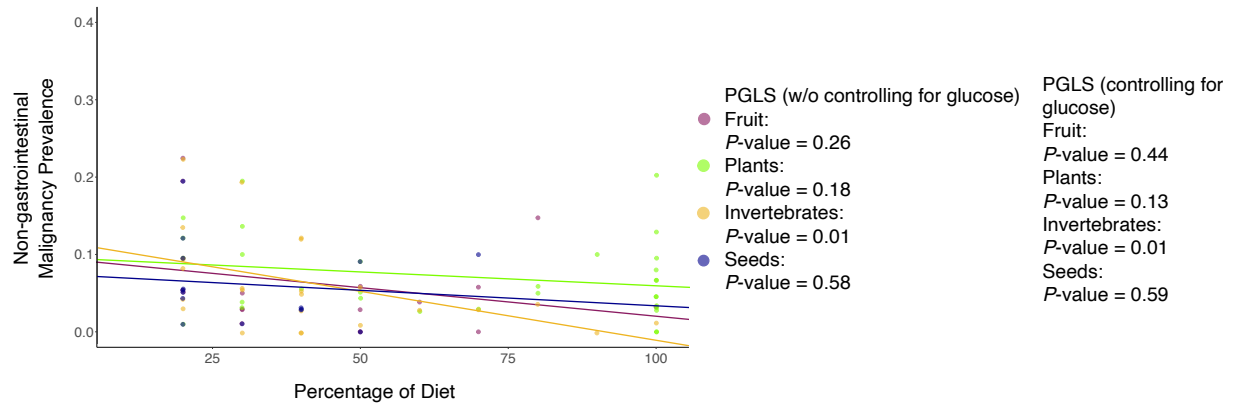
A



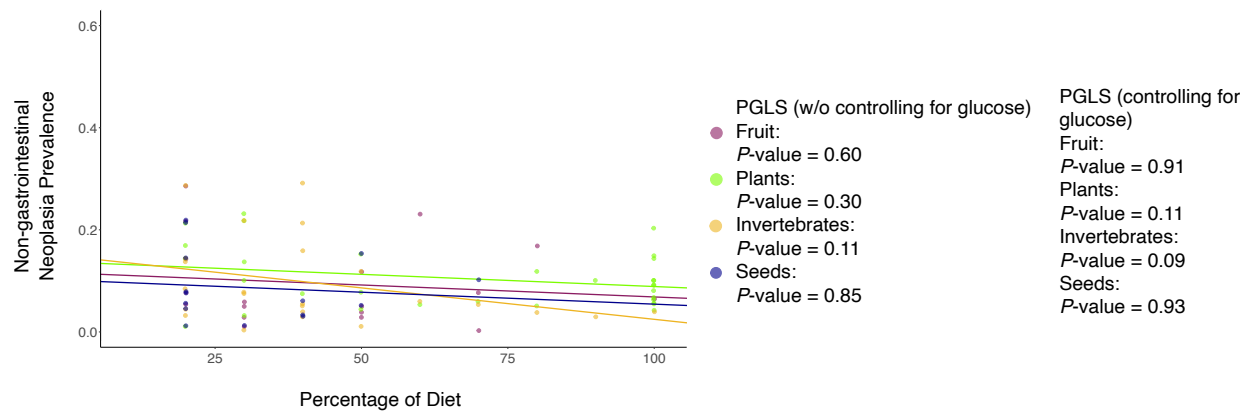
B



C

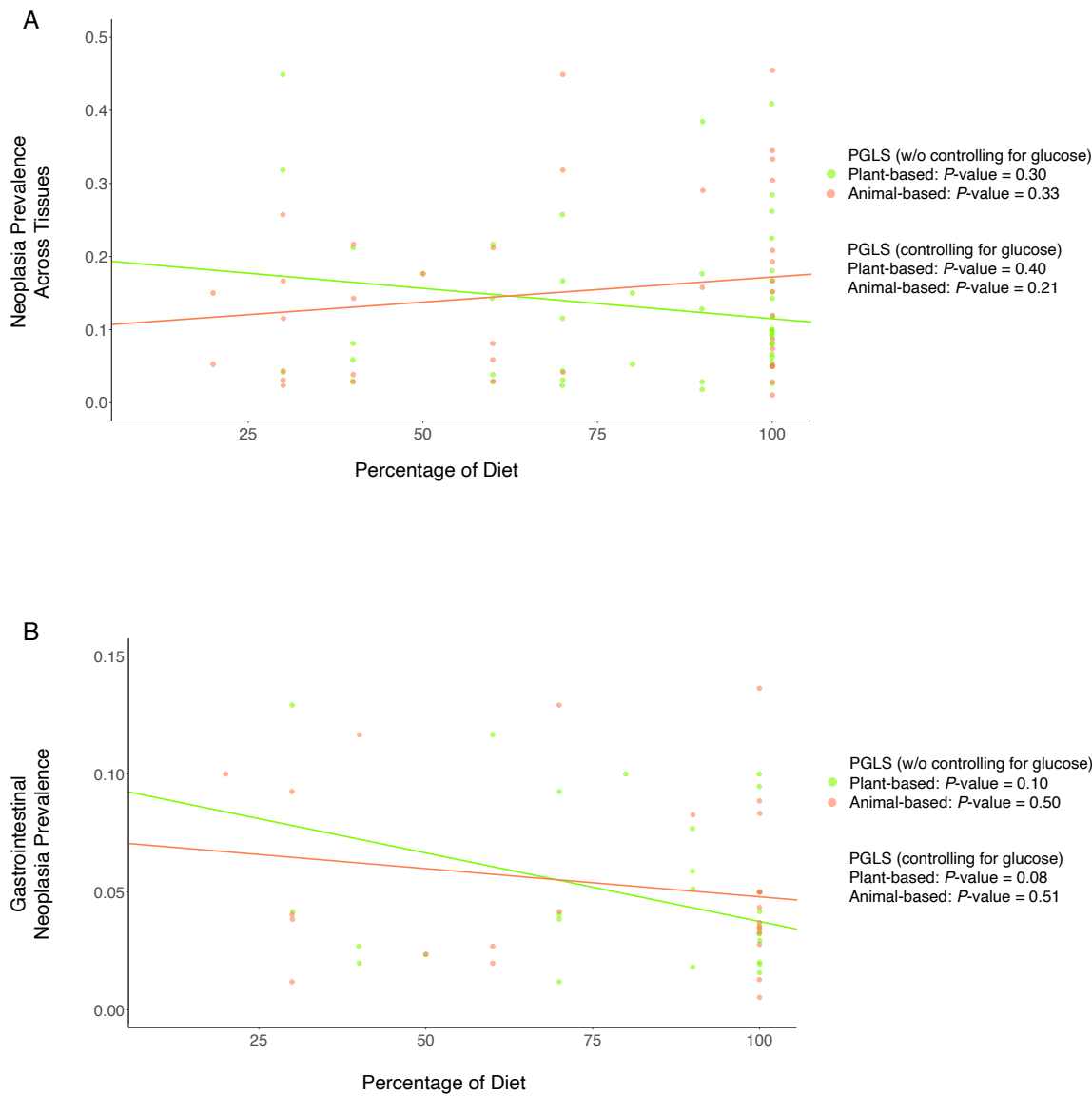


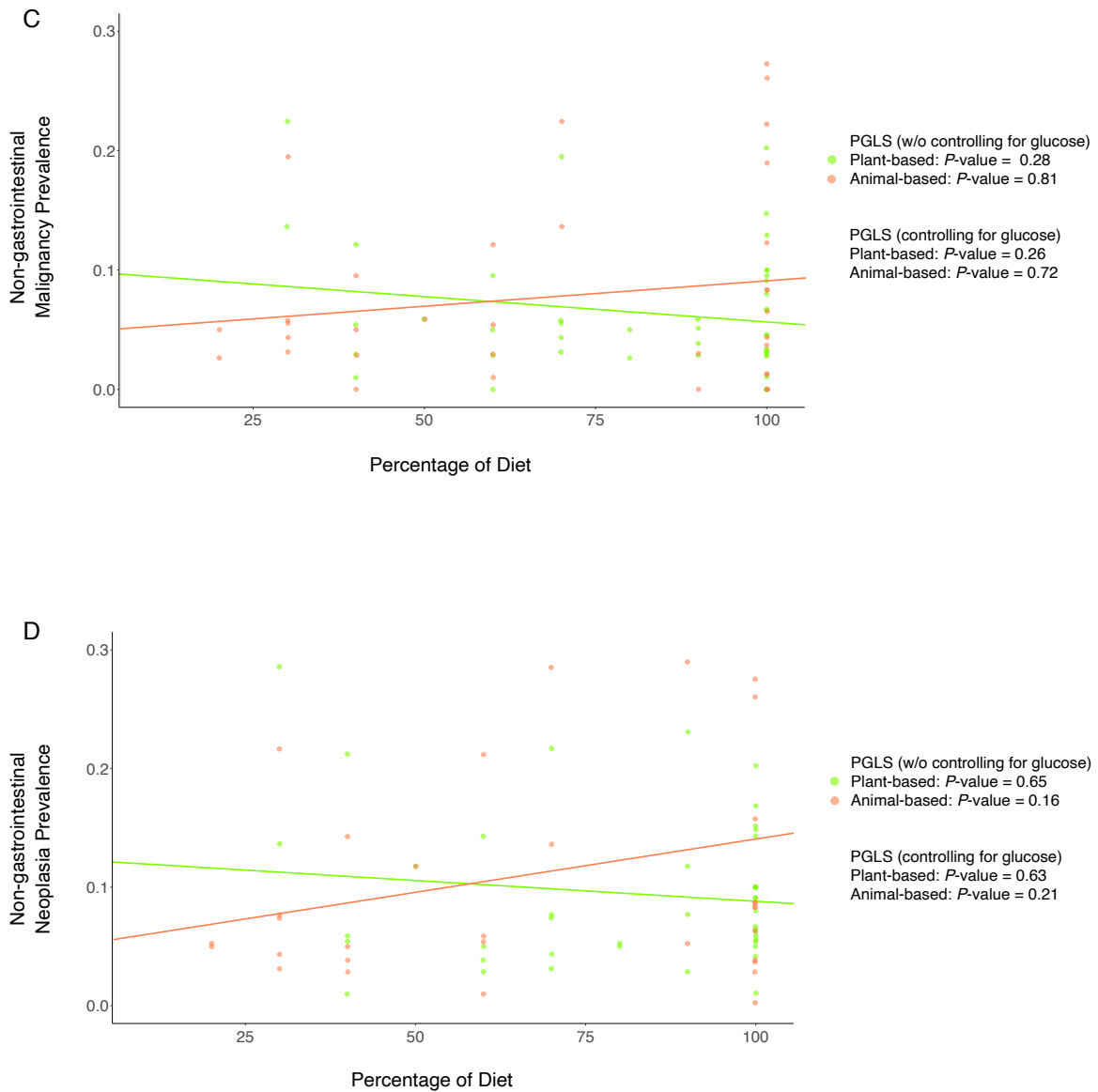
D



Supplementary Figure 5. No detectable correlation between cancer or neoplasia prevalence and the percentage of a specific diet in a species after correcting for multiple testing. (A) There is no correlation between the percentage of fruit, plants, invertebrates, seeds, or endothermic vertebrates (Endotherms) in a species' diet and neoplasia prevalence across tissues for 20, 34, 27, 17 or 10 species, respectively (PGLS: P -value > 0.05 ; Supplementary Data 7). (B) There is no correlation between the percentage of fruit, plants, invertebrates, or endothermic vertebrates (Endotherms) in a species' diet and gastrointestinal neoplasia prevalence for 15, 19, 18, and 10 species, respectively, after multiple testing corrections (Supplementary Data 7). (C) There is no correlation between the percentage of fruit, plants, invertebrates, or seeds in a species' diet and non-gastrointestinal malignancy prevalence for 18, 31, 24, or 15 species, respectively, after corrections for multiple testing (Supplementary Data 7). (D) There is no correlation between the percentage of fruit, plants, invertebrates, or seeds, in a species' diet and non-gastrointestinal neoplasia prevalence for 18, 31, 24, or 15 species, respectively (PGLS: P -value > 0.05). Each dot shows the neoplasia prevalence across tissues (A), the gastrointestinal neoplasia prevalence (B), the non-gastrointestinal malignancy prevalence (C), the non-gastrointestinal neoplasia prevalence (D), and the diet percentage of one species. We added minimal jitter in the plots in order to better visualize individual data points. The statistical tests were two-sided. P -values have not been adjusted for FDR corrections. The P -values of pgl analyses that passed the FDR correction are noted with an asterisk in

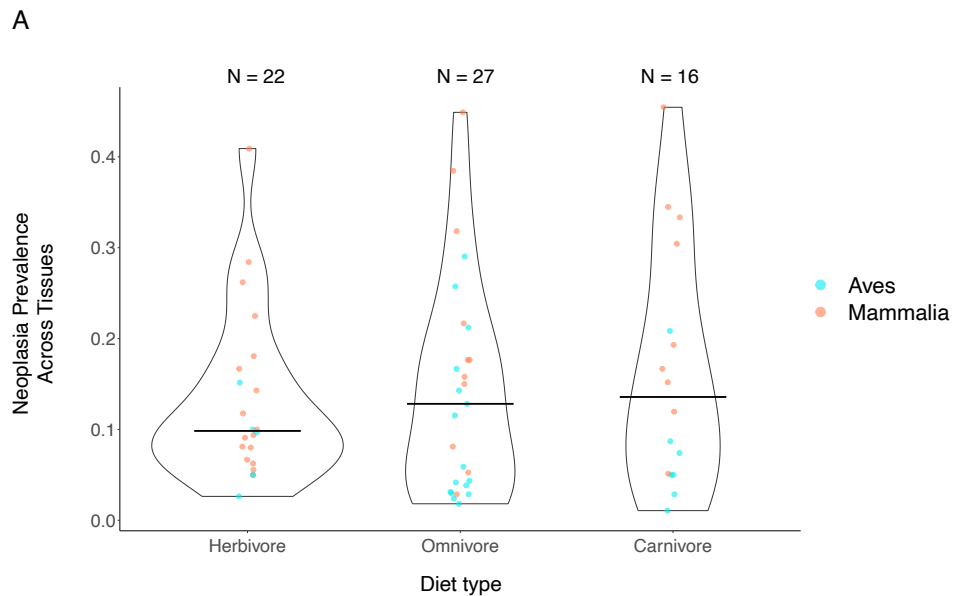
Supplementary Data 7. PGLS: phylogenetic generalised least squares. w/o: without. For information on how to generate this figure, please see Source Data file.



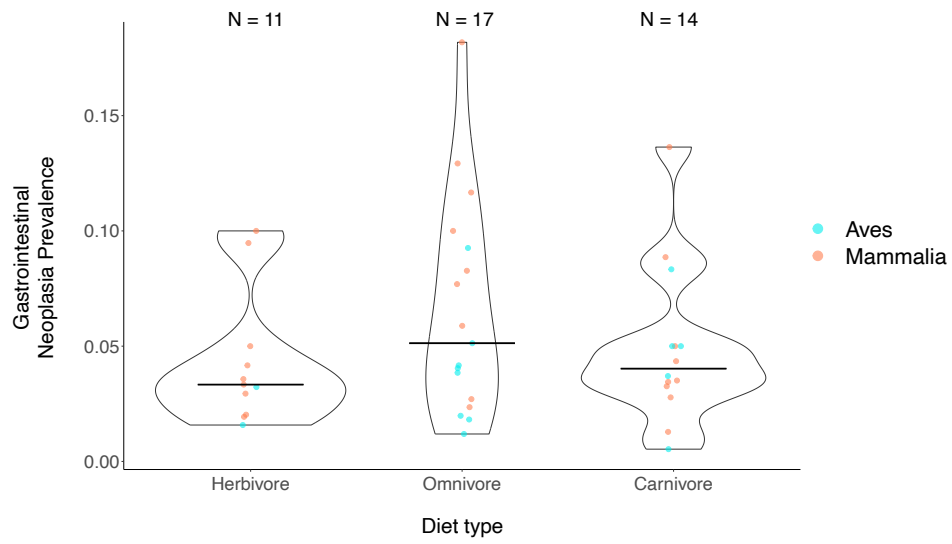


Supplementary Figure 6. No detectable correlation between malignancy prevalence or neoplasia prevalence and percentage of plant-based or animal-based food in species' diet after correcting for multiple testing. (A) There is no correlation between the percentage of plant-based or animal-based food in a species' diet and neoplasia prevalence across tissues for 44 or 38 species, respectively (PGLS: P -value > 0.05; Supplementary Data 7). (B) There is no correlation between the percentage of plant-based or animal-based food in a species' diet and gastrointestinal neoplasia prevalence for 27 species (PGLS: P -value > 0.05; Supplementary Data 7). (C) The percentage of plant-based or animal-based food in a

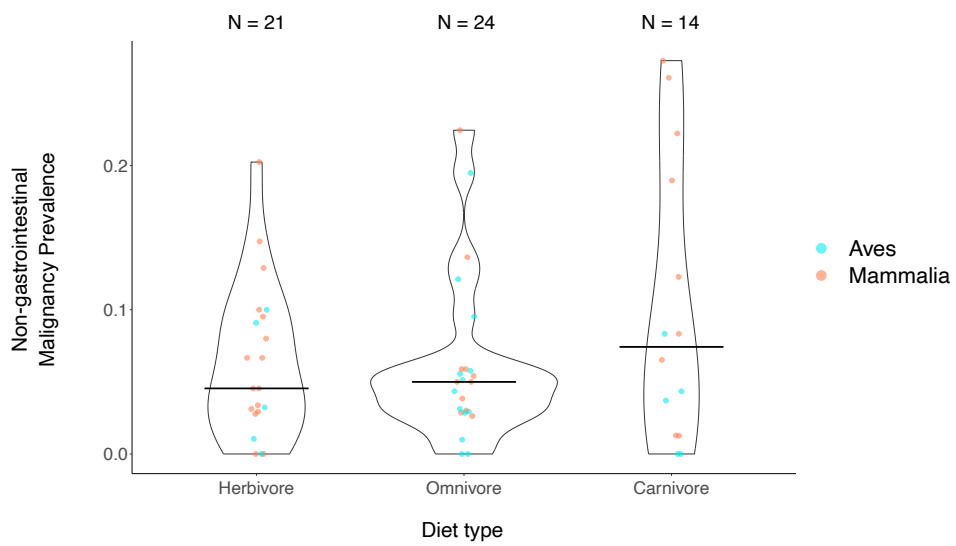
species' diet is not correlated with non-gastrointestinal malignancy prevalence for 43 or 34 species, respectively (PGLS: P -value > 0.05). (D) The percentage of plant-based or animal-based food in a species' diet is not correlated with non-gastrointestinal neoplasia prevalence for 43 or 34 species, respectively (PGLS: P -value > 0.05). Each dot shows the neoplasia prevalence across tissues (A), the gastrointestinal neoplasia prevalence (B), the non-gastrointestinal malignancy prevalence (C), the non-gastrointestinal neoplasia prevalence (D), and the percentage of plant-based or meta-based food in the diet of one species. We added minimal jitter in the plots in order to better visualize individual data points. The statistical tests were two-sided. P -values have not been adjusted for FDR corrections. The P -values of pglis analyses that passed the FDR correction are noted with an asterisk in Supplementary Data 7. PGLS: phylogenetic generalised least squares. w/o: without. For information on how to generate this figure, please see Source Data file.

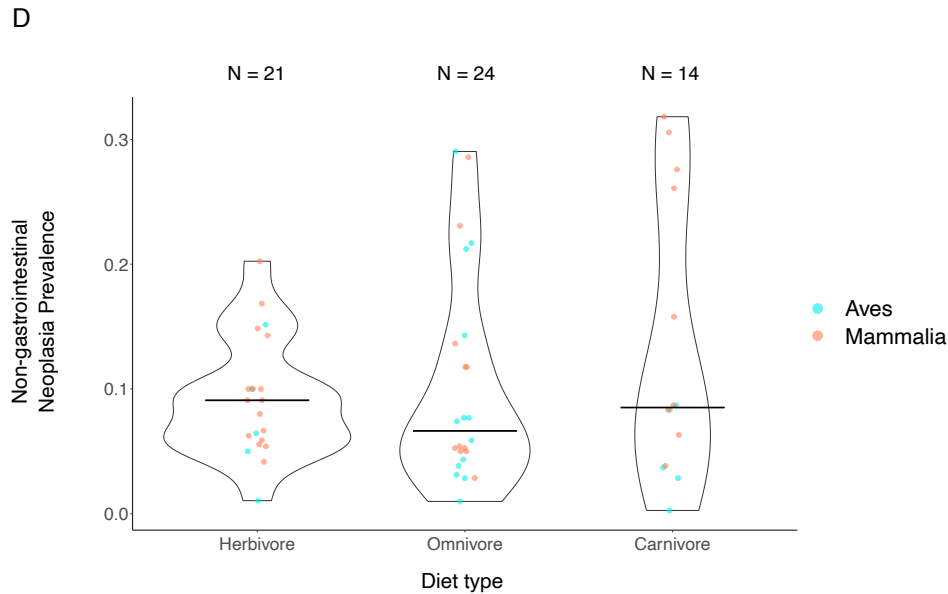


B



C





Supplementary Figure 7. No detectable correlation between malignancy prevalence or neoplasia prevalence and diet type. Diet type is not correlated with neoplasia across tissues (A), gastrointestinal neoplasia prevalence (B), non-gastrointestinal malignancy prevalence (C), or non-gastrointestinal neoplasia prevalence (D) (PGLS: P -value > 0.05 ; Supplementary Data 7). The statistical tests were two-sided. P -values have not been adjusted for FDR corrections. The P -values of pgl analyses that passed the FDR correction are noted with an asterisk in Supplementary Data 7. Each dot shows the neoplasia prevalence across tissues (A), the gastrointestinal neoplasia prevalence (B), the non-gastrointestinal malignancy prevalence (C), the non-gastrointestinal neoplasia prevalence (D), and the diet type of one species. N shows the number of species per diet category. Herbivores refer to animals that have a 100% plant diet, omnivores refer to animals that have $>0\%$ plant and meat diet, and carnivores refer to animals that have a 100% meat (invertebrate or vertebrate) diet. We added minimal jitter in the plots in order to better visualize individual data points. The horizontal black line in each diet category shows the median neoplasia prevalence across tissues (A), the median gastrointestinal neoplasia prevalence (B), the median non-gastrointestinal malignancy prevalence (C), or the median non-gastrointestinal neoplasia prevalence (D), in that diet category. For information on how to generate this figure, please see Source Data file.