Supplementary Information

Rate of molecular evolution and pleiotropy on the evolutionary trajectories of penguin genes.

Felipe Avello-Duarte, Luis R. Pertierra, Juliana A. Vianna, Pablo Razeto-Barry, Elie Poulin.

Felipe Avello-Duarte

felipe.avello@ug.uchile.cl

Supplementary Data

Pleiotropy: terms for the Biological Processes' ontology from Gene Ontology for each coding sequence and the pleiotropic level after redundancy reduction.

dNdS: mean dN/dS ratio for each coding sequence using the one ratio model from PAML.

Pairwise dNdS: mean dN/dS ratio for each pair of species and each coding sequence using the one ratio model from PAML (Afo: emperor penguin, Apa: king penguin, Pad: adelie penguin, Pan: chinstrap penguin, Ppe: papua penguin (Antarctic Peninsula), Ppp: papua penguin (Falkland Islands), Ppk: papua penguin (Kerguelen Islands), Ppt: papua penguin (Crozet Island), Emi: little penguin, Shu: Humboldt penguin, Sme: Galápagos penguin, Sma: magellanic penguin, Sde: african penguin, Ech: macaroni penguin, Epa: Fiordland penguin, Esc: erect-crested penguin, Emo: northern rockhopper penguin, Ecc: southern rockhopper penguin).

Branch dNdS: dN/dS ratio for each branch in the penguin phylogeny and each coding sequence using the free ratio model from PAML. Each branch is denoted by a number, following the topology of the phylogeny obtained by Vianna et al. (2020):

((((((18,19)17,(21,22)20)15,16)14,(((9,10)8,(12,13)11)7,6)5)4,((28,32)24,33)23)3,(1,2)0);

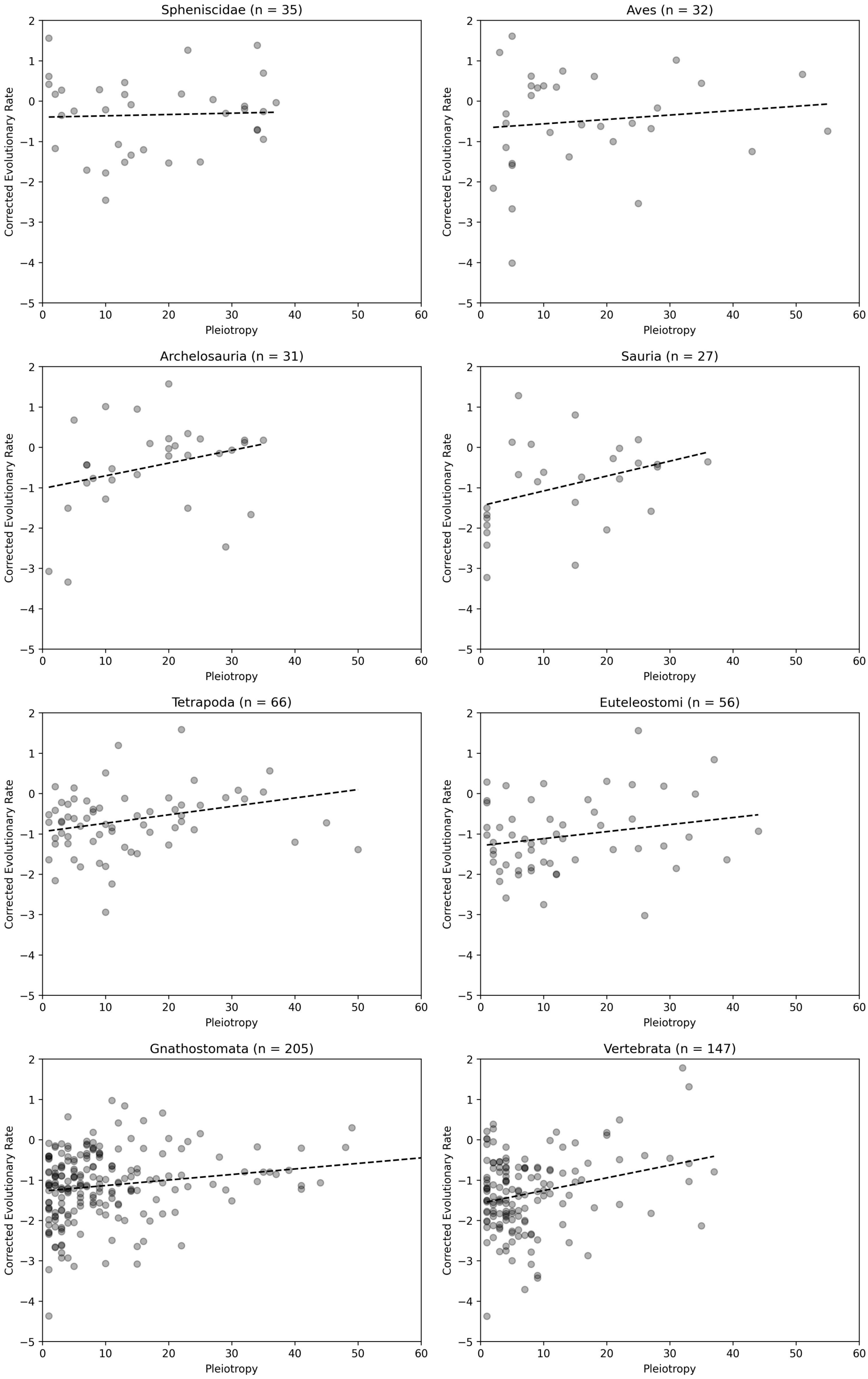
((((((Sme,Shu),(Sde,Sma)),Emi),(((Ecc,Emo),(Epa,Esc)),Ech)),((Ppe,Pan),Pad)),(Apa,Afo));

Phylostrata: output file for the phylostratification analysis with assigned stratum for each coding sequence.

Phylostrata dating: date used to locate each phylostratum in time based on phylogenetic reconstructions. The date corresponds to the estimated date of divergence between the clade to which the emperor penguin belongs and the reference species.

ODP Sites: Ocean Drilling Program drilling sites and locations used to calibrate the paleotemperature models.

Supplementary Figures



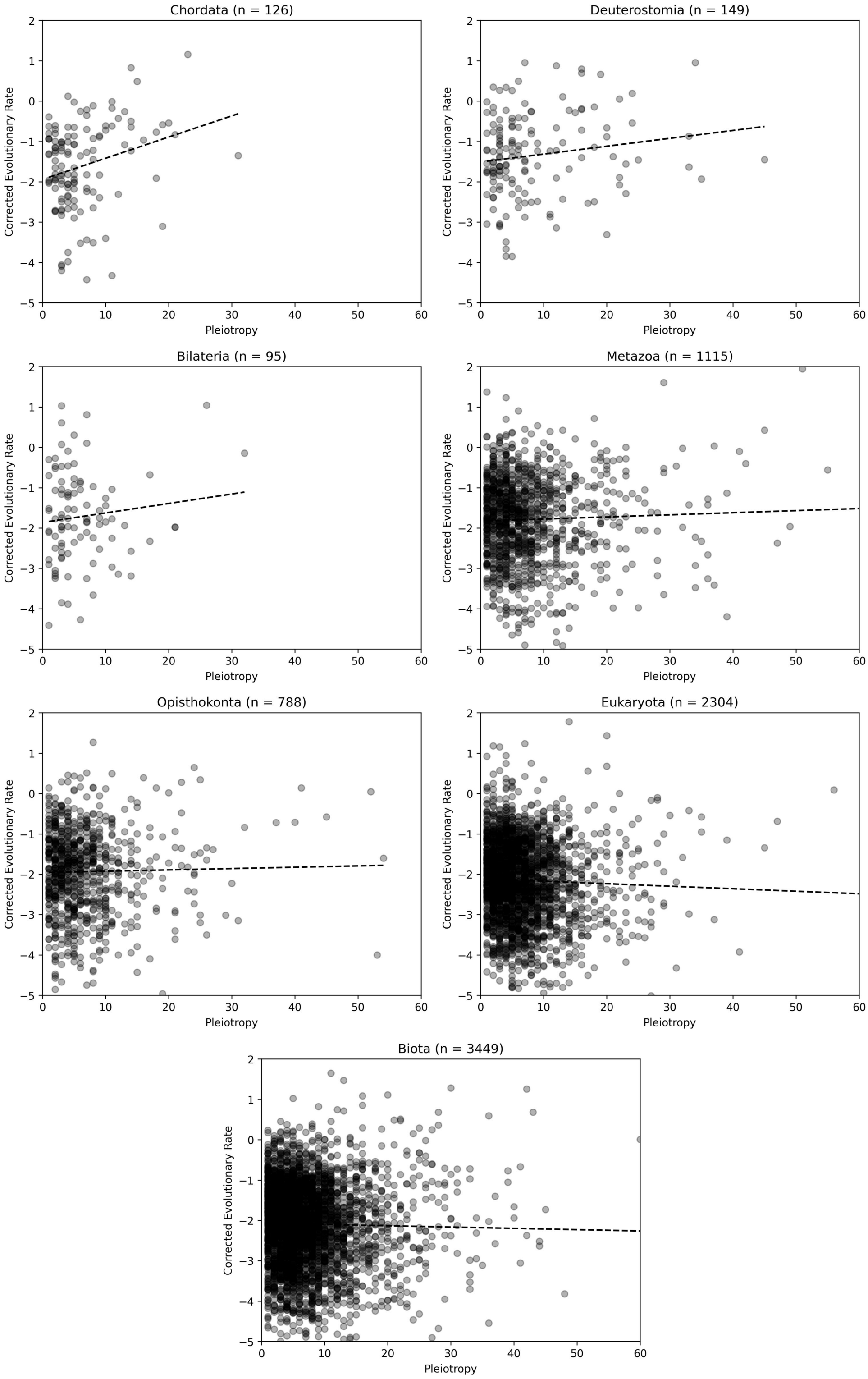


Fig S1. Molecular evolutionary rate as a function of pleiotropic level, detailed by phylostratum. d_N/d_S ratio is corrected by a Box-Cox transformation. The dashed line represents the linear regression between the two variables. The sample size is presented in the title of each plot.

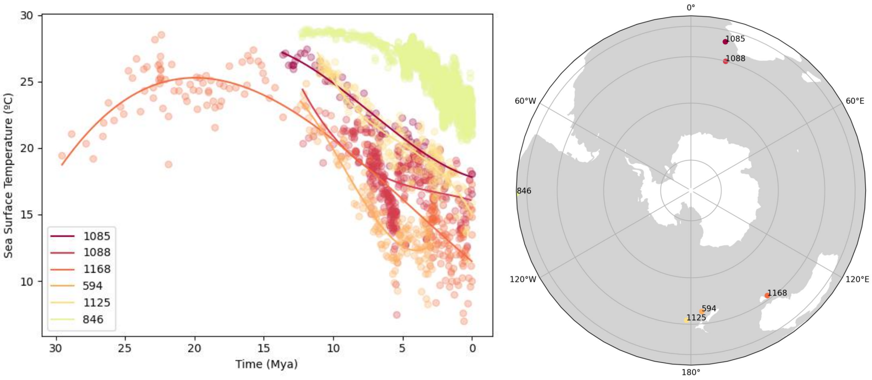


Fig S2. ODP drilling sites used for calibration. Dots show the paleotemperature estimations, and the curves correspond to polynomial functions fitted to the data. The map shows the locations of the ODP drilling sites.

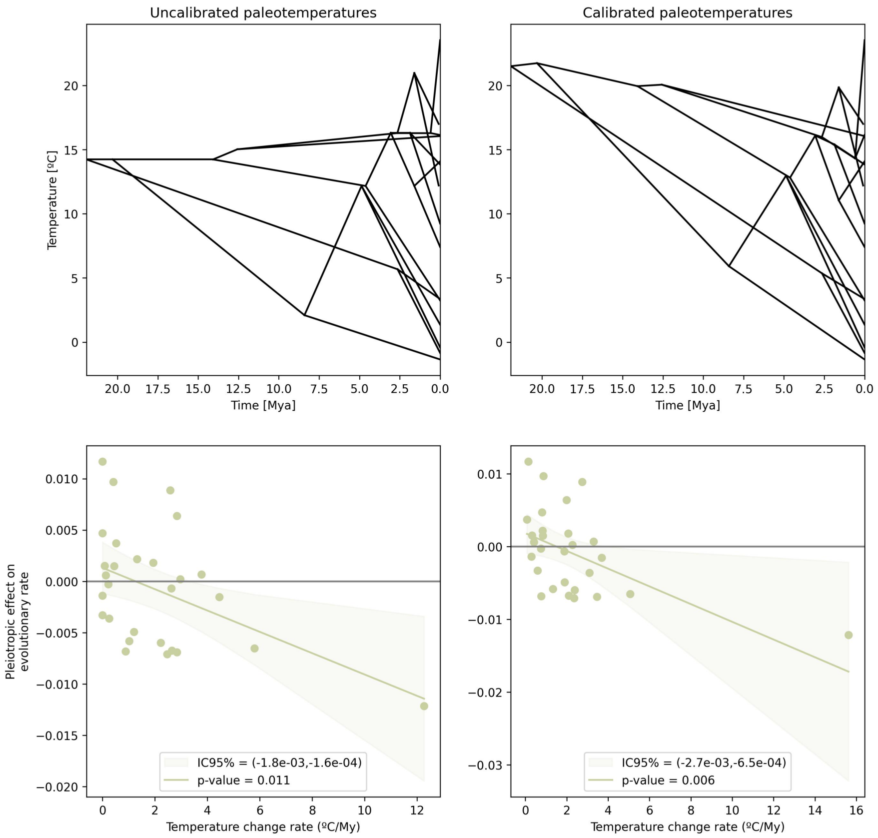


Fig S3. Comparison between estimations of paleotemperatures. The left panels show the results obtained with temperatures estimated directly from the paleotemperature models for the Miocene and the Pliocene. The right panels show the same data, but the paleotemperature is corrected using sea surface temperature estimations based on $U^{k'}_{37}$ ratios. The top panels display the phylogeny of penguins, with each node positioned at the corresponding date and temperature estimated for the reconstructed distributions. Bottom panels show the effect of temperature change on each branch of the phylogeny and its impact on the pleiotropic effect on molecular evolutionary rate. The bottom right panel displays the same data as Fig. 4a, including the data point that corresponds to the branch that connects the Galápagos penguin to its most recent common ancestor with the Humboldt penguin.