nature portfolio

Corresponding author(s): Felipe Avello-Duarte

Last updated by author(s): Jun 19, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The retrieval of Gene Ontology information was conducted using the Python library PyOMADBv2.0.0. The retrieval of environmental information was conducted using the R packages SDMpredictorsv0.2.15 and Leafletv2.2.2.
Data analysis	Data was processed with the Python libraries biopython/1.85, goatools/1.4.12, hdbscan/0.8.40, numpy/2.2.3, pandas/2.2.3, scikit-learn/1.6.1, and scipy/1.15.2, and the R package phylostratr.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Source datasets retrieved from other scientific articles are properly referenced in this paper. All generated data used in this manuscript are deposited in Zenodo (DOI:https://doi.org/10.5281/zenodo.15676841). The main generated datasets include pleiotropic level, evolutionary rate, and phylostratification information for each coding sequence tested, datation data, as well as Ocean Drilling Project sites' location.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.* *Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.* *Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The study explored the correlation between pleiotropic level and molecular evolutionary rate, and the change of this correlation as a function of time and environmental shifts. The pleiotropy was estimated using data from Gene Ontology and the molecular evolutionary rate was estimated based on the ratio between non-synonymous substitution rate and synonymous substitution rate. Additionally we measured the estimated time of origin of each coding sequence based on a phylostratigraphic analysis and the estimated change in environmental conditions as the mean sea surface temperature change by million year to account for the effects of time and environment in the evolutionary processes.
Research sample	We used as source dataset all the shared coding sequences predicted by genomic analysis of the Spheniscidae family, which corresponds to a total of 11,011 DNA sequences, retrieved from the Dryad repository (DOI:10.5061/dryad.pk0p2ngj2) related to the article Vianna et al. 2020. Genome-wide analyses reveal drivers of penguin diversification. PNAS. We were able of retrieving information of only 8,625 sequences, which corresponds to our main dataset used across the paper. The source dataset was chosen because of the novelty that represented a theoretical work on this group, the rich knowledge that has been achieved recently about their genomes thanks to the technological advancements on sequencing, and their unique evolutionary history and adaptations.
Sampling strategy	The sampling strategy aimed at collecting all available sequences shared by the extant penguin species that had at least one known associated Gene Ontology term (from the biological processes ontology) and that had an omega ratio greater than 0.0001 and lesser than 999, which are the minimum and maximum values raised by PAML when there is no non-synonymous substitution or no synonymous substitution, respectively.
Data collection	The genomic data was manually downloaded from the Dryad repository by Felipe Avello-Duarte. The environmental data was downloaded by Luis R. Pertierra using SDMpredictors and Leaflet packages.
Timing and spatial scale	The genomic data was downloaded on 2021.
Data exclusions	We excluded the data corresponding to subgroups of classical species of penguins to ensure a continuity with the results presented by Vianna et al. 2020. This include three populations of the Papua penguin (Antarctic Peninsula, Kerguelen Islands, and Crozet Islands), Eastern rockhopper penguin, and two populations of Macaroni penguin (Elephant Island and Macquirie Islands).

Reproducibility	IIn the case of the theoretical predictions, we performed multiple independent simulations with a wide range of parameter values, including different population sizes, environmental variability, and mutation effects. Every simulation showed the same pattern. For the empirical insights, we conducted random regrouping of genes between phylostrata to ensure that the curve of pleiotropic effect and time was robust.
Randomization	We did not explored the effect of sampling random genes from the penguin genome on the stated results.
Blinding	Blinding was not relevant, as the amount of data prevented the individualization of each coding sequence and the creation of expectations.

Did the study involve field work? ☐ Yes ☒ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*