

## Reporting Summary

Nature Research 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 Research policies, see [Authors & Referees](#) 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

No commercial/custom code was used for data collection.

Data analysis

We list here the software URLs of all open-source tools used for data analysis.

R: <https://www.R-project.org/>

Arrow: <https://github.com/PacificBiosciences/GenomicConsensus>

Canu: <https://github.com/marbl/canu>

Flye: <https://github.com/fenderglass/Flye>

WTDBG <https://github.com/ruanjue/wtdbg2>

BUSCO: <https://gitlab.com/ezlab/busco>

Jellyfish: <https://www.cbcb.umd.edu/software/jellyfish/>

GenomeScope: <http://qb.cshl.edu/genomescope/>

Repeatmasker: <http://www.repeatmasker.org>

Repeatmodeler: <http://www.repeatmasker.org/RepeatModeler.html>

MAKER: <http://www.yandell-lab.org/software/maker.html>

SNAP: <http://snap.cs.berkeley.edu>

AUGUSTUS: <http://augustus.gobics.de/>

Exonerate: <https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate>

snpEff: <http://snpeff.sourceforge.net/>

Enrichr: <http://amp.pharm.mssm.edu/Enrichr>

GATK: <https://software.broadinstitute.org/gatk/download/>

BWA: <https://github.com/lh3/bwa/releases>

EdgeR: <https://bioconductor.org/packages/release/bioc/html/edgeR.html>

Synteny plots: [https://github.com/biopython/biopython/blob/master/Doc/examples/Proux\\_et\\_al\\_2002\\_Figure\\_6.py](https://github.com/biopython/biopython/blob/master/Doc/examples/Proux_et_al_2002_Figure_6.py)

OrthoFinder: <http://www.stevkellylab.com/software/orthofinder>

STAR: <https://github.com/alexdobin/STAR>

featureCounts: <http://bioinf.wehi.edu.au/featureCounts/>  
 Picard: <http://broadinstitute.github.io/picard/>  
 Figtree: <https://github.com/rambaut/figtree>  
 CoGe: <https://genomevolution.org/coge/>  
 Symap: <http://www.agcol.arizona.edu/software/symap/>  
 vt: <https://github.com/atks/vt>

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [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 list of figures that have associated raw data
- A description of any restrictions on data availability

All sequencing data that support this study have been deposited in NCBI under the BioProject accession # PRJNA527614 and can be accessed at "<https://dataview.ncbi.nlm.nih.gov/object/PRJNA527614>".

Mass spectrometry data collected in this study has been deposited at MassIVE and can be accessed via this link - "<ftp://massive.ucsd.edu/MSV000084564/>".

## 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](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size: A total of 6 animals was used in this study as described in Table S1a and methods section.

Data exclusions: No data was excluded in this study.

Replication: Biological replicates, where appropriate, have been used and described in the methods section and supplementary tables

Randomization: This is a de novo genome sequencing project and hence, randomization does not apply to this study.

Blinding: This is a de novo genome sequencing project and hence, blinding does not apply to this study.

## 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	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Animals and other organisms

Policy information about [studies involving animals](#); ARRIVE [guidelines](#) recommended for reporting animal research

Laboratory animals

We obtained horse blood from a horse (*Equus caballus*) that is a research mare at the Texas A&M Large Animal clinic.  
 Species: *Equus caballus*  
 Breed: American quarter horse

	Sex: female Age: 6 years old
Wild animals	Two animals used in this study from India were fresh road kills and were approved for use in the study as described in the methods section.
Field-collected samples	This study did not include field-collected samples
Ethics oversight	This study did not require ethics approval.

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

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

Sample preparation	Sample preparation and processing are described in the materials and methods section of the main text.
Instrument	BD Accuri™ C6 personal flow cytometer
Software	BD Accuri C6
Cell population abundance	Described in materials and methods section of main text
Gating strategy	Gating strategy is provided in Supplementary Fig. 1a

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.