



OPEN

DATA DESCRIPTOR

# Behavioral dataset for Long-Evans and its schizophrenia-like substrain through several generations

Gábor Kőrösi<sup>1</sup>, Oliver Czibalmos<sup>1</sup>, Gabriella Kekesi<sup>2</sup> & Gyongyi Horvath<sup>2,3,4</sup>✉

We present a high-throughput behavioral dataset acquired with Ambitus, an automated reward-based corridor system that records locomotor and exploratory activities and cognitive functions after minimal handling. The collection contains 91 raw and derived variables, each measured across four consecutive trials, for 1,342 Long-Evans rats, including a triple-hit schizophrenia-like substrain (Lisket) bred through 16 generations. All data files, detailed metadata and analysis scripts are openly available on Zenodo. This resource enables longitudinal and multivariate studies of behavioral phenotypes, trans-generational effects, and strain differences, and it provides a benchmark for machine-learning-based marker discovery in rodent models.

## Background & Summary

We present an openly available, comprehensive, multi-generation behavioral dataset for a schizophrenia-like rat model. The dataset comprises 1,342 Long-Evans and Lisket rats tested across 16 generations (2019–2025). Ninety-one parameters describing locomotion, exploration, reward collection and learning were recorded in four sequential trials per animal with the Ambitus corridor system, yielding >360,000 individual data points. All raw CSV, aggregated Parquet and metadata files, together with analysis notebooks, are available on Zenodo (<https://doi.org/10.5281/zenodo.16414790>)<sup>1</sup> under a CC-BY 4.0 licence.

Schizophrenia is a complex mental disorder characterized by several symptoms (e.g. positive, negative and cognitive). Modeling this disease using animals is very challenging because of the higher complexity of the human brain compared to the animals. Translational research depends on the relevance of animal models that replicate several signs of a human disorder. Rodents are valuable tools for studying human diseases, although existing animal models cannot fully replicate clinical conditions, especially neuropsychiatric disorders, including schizophrenia. Genetic and environmental factors both contribute to the aetiology of schizophrenia; therefore a multifactorial (“triple-hit”) model is warranted. A “triple-hit” rat substrain, termed Lisket has been developed from the Long-Evans (LE) rats with high behavioral activity and cognition function compared to other strains (e.g. Wistar)<sup>2–4</sup>, by combining post-weaning social isolation, subchronic treatment with the N-methyl-d-aspartate [NMDA] receptor antagonist, ketamine, and selective breeding based on behavioral phenotype manipulations through 16 generations (G0–G15).

Several methods are available to determine behavioral activity and learning ability of rodents during short time periods, including novel object recognition, hole-board and several types of mazes (e.g. Morris Water-maze, and radial mazes with different number of arms)<sup>5–12</sup>. Other, more sophisticated methods, such as tasks with touch screen or lever press require very intensive, long-lasting learning sessions to acquire the tasks<sup>13–17</sup>. To balance throughput and richness, we designed the Ambitus apparatus combining corridor and hole-board elements to record behavior automatically in a rectangular pathway with side-boxes<sup>18</sup>. The Ambitus detects automatically the behavior of animals during simple tasks. We hypothesized that food-deprived rats would collect food rewards from the first trial without prior acclimatization in the corridor system. Sequential trials with paradigm alterations were designed to detect behavioral changes within a single testing session. The

<sup>1</sup>Department of Computer Algorithms and Artificial Intelligence, Institute of Informatics, Faculty of Science and Informatics, University of Szeged, Szeged, Hungary. <sup>2</sup>Department of Physiology, Albert Szent-Györgyi Medical School, University of Szeged, Szeged, Hungary. <sup>3</sup>Present address: Department of Medical Physics and Informatics, University of Szeged, Szeged, Hungary. <sup>4</sup>Present address: Sztárai Institute, University of Tokaj, Sárospatak, Hungary. ✉e-mail: [horvath.gyongyi@med.u-szeged.hu](mailto:horvath.gyongyi@med.u-szeged.hu)

10<sup>th</sup> week of age – Ambitus test

Schematic of the arena layout

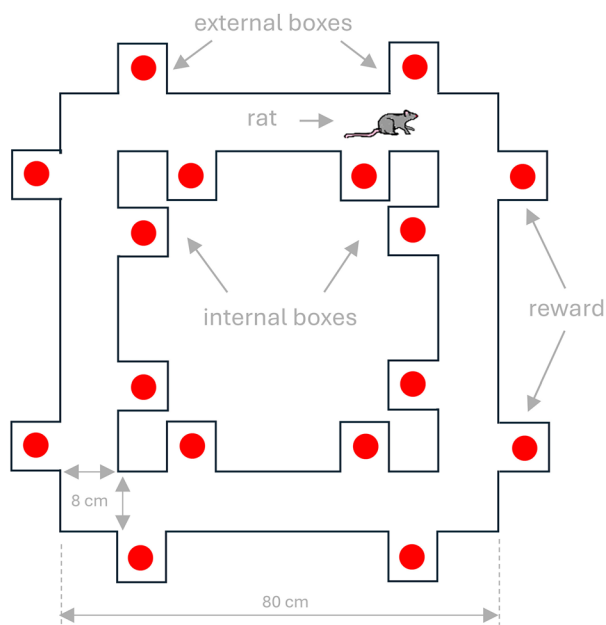
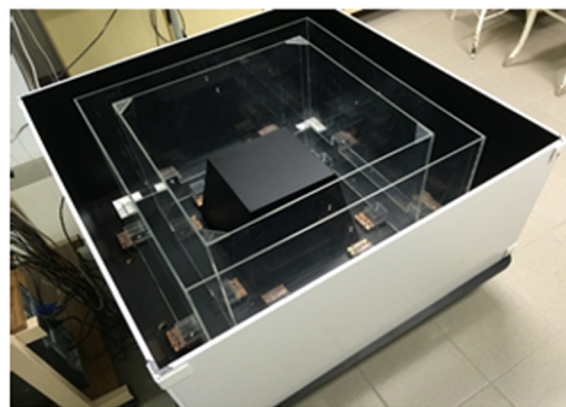


Photo of the Ambitus testing apparatus



## Experimental paradigm

Trial 1	2 min	Trial 2	3 hours	Trial 3	2 min	Trial 4
Task 1: 16 rewards (external and internal boxes)				Task 2: 8 rewards (only internal boxes)		

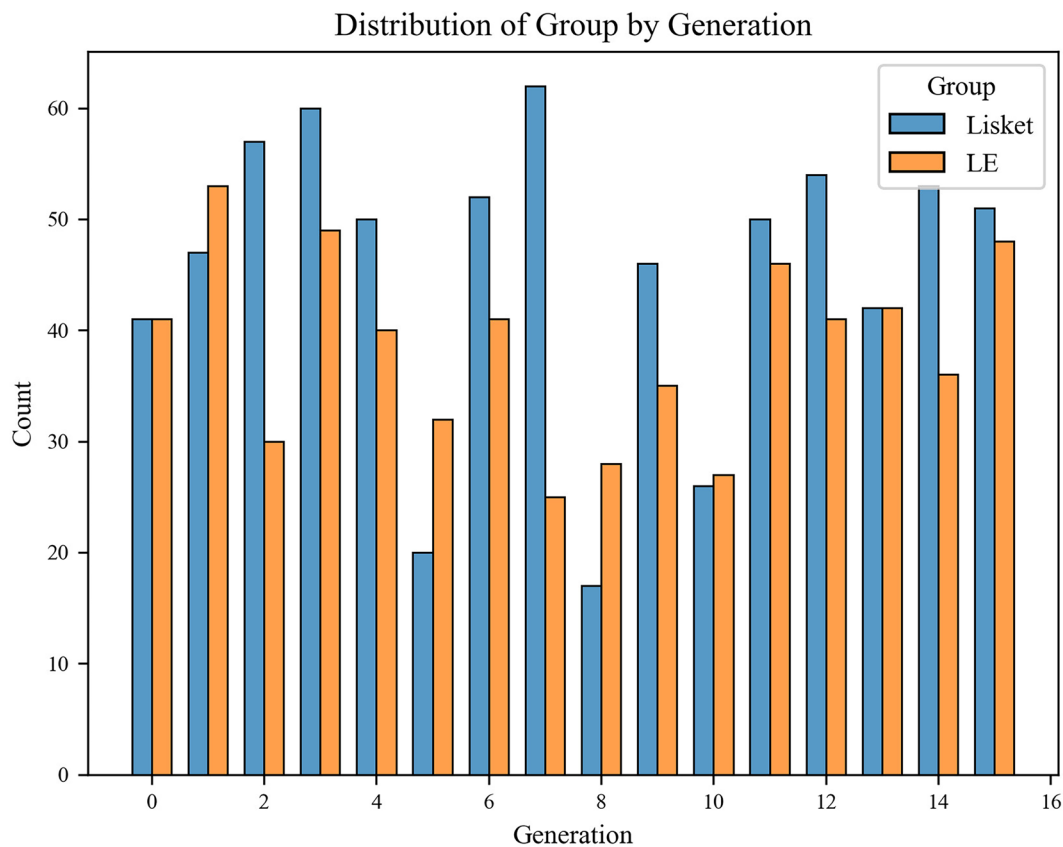
**Fig. 1** Ambitus apparatus: ground plan of the corridor with 16 side-boxes equipped with photobeams and experimental paradigm.

instrument can provide rapid and quantitative behavioral readouts within minutes data about the locomotor and exploratory activities and learning capacity of animals within a short time<sup>18</sup>.

While the reliability of Ambitus has been demonstrated in specific paradigms<sup>18</sup>, the present Data Descriptor releases the full seven-year dataset in a FAIR format. By providing comprehensive raw and processed files plus detailed metadata, the resource enables researchers to estimate heritability across generations, identify behavioral markers, validate translational models of schizophrenia, and benchmark machine-learning pipelines. The scale of the dataset further supports pharmacological or toxicological screening, as well as studies of age-related cognitive decline and neural plasticity. Usage examples, step-by-step loading instructions and baseline analysis notebooks accompany the archive to facilitate immediate reuse. Together, these materials furnish an open, extensible platform for behavioral neuroscience and data-driven discovery.

### Methods

**Animals and ethical approval.** Long-Evans, and Lisket rats in both sexes during a 7-year period were involved in the study. Animals were treated in accordance with the guidelines set by the Government of Hungary for the humane treatment of animals. All experiments were approved by the Hungarian Ethical Committee for Animal Research (RN: XIV/1248/2018 and XIV/1421/2023) and conducted in compliance with the EU Directive 2010/63/EU and the Animal Research: Reporting of In Vivo Experiments (ARRIVE) guidelines 2.0. Animals were kept with a 12h light/dark cycle under conditions of controlled temperature ( $22 \pm 1^\circ\text{C}$ ) and before the tests the animals were food deprived for two days, but water was always freely available. The experimental procedures were performed between 8:00 a.m. and 4:00 p.m. in a room with dim general light without external cues.



**Fig. 2** Distribution of animals across experimental groups for each generation. The dataset consists of 728 Lisket (model) and 614 LE (control) rats.

**Generation protocol and breeding.** Based on earlier studies performed with Wistar and its schizophrenia-like substrain (Wisket)<sup>10,18,19</sup>, both LE (control) and Lisket (male and female) rats were weaned at 3 weeks of age. The Lisket rats were housed individually for 28 days and treated with ketamine (30 mg/kg, 4 mL/kg) intraperitoneally on 5 consecutive days on the 2nd week of isolation<sup>20</sup>. Subsequently, the animals were rehoused (2-3 animals/cage). LE rats were socially reared with no treatment. The estrus cycle of the female rats was allowed to vary randomly and was not measured or controlled. Three months old control and Lisket rats were selected for mating<sup>10</sup>. To minimize inbreeding sibling mating was avoided. Generally, 6 males were housed with 12 females in sex cages to let one male mate with two females for two weeks, after which the females were reared alone to ensure undisturbed delivery of the offspring. Generation 0 Lisket animals were exposed only to social isolation and ketamine treatment, while the subsequent generations were also selectively bred. In total 1,342 rats of 16 generations (G0-G15) during a 7-year period (year 0-6) were investigated.

**Ambitus apparatus.** The Ambitus apparatus, which is designed as a reward-based cognitive test, consists of a rectangular corridor constructed of clear Plexiglas on a black floor with an outer diameter of 80 cm, width of 8 cm and height of 50 cm (Fig. 1)<sup>18</sup>. The rats can move around the track between the walls in forward and backward directions. Each corridor contains four side boxes (5 × 5 × 5 cm; 2-2 at the inner and external sides of one corridor, altogether 16) for food rewards (puffed rice, 20 mg). Each box and the middle of each corridor are fitted with infrared LED (light emitting diodes) at one side and a photocells at the other side to measure nose-poking activity (as exploration) in the boxes, the collection of the rewards, and the locomotor activity, respectively, with 1 ms time resolution, and stores the information about the location of the animals and the duration and the number of visits. Thus, the Ambitus apparatus allowed the analysis of several behavioral parameters, whose definitions and calculation methods are described in detail in the accompanying data descriptor document available in Zenodo<sup>1</sup>. The apparatus was installed in a room where no other activities took place during the testing procedure. Before each trial the experimenter inserted the food rewards into the side boxes. Trials commenced by placing the rats into the same starting point within the corridor (Fig. 1); thereafter, the experimenter immediately left the room. The animals were allowed to explore the corridor and collect food rewards for 5 min (300s; cut-off time). The apparatus was cleaned with 70% alcohol after each rat.

**Behavioral test protocol.** At the age of 10 weeks the animals underwent the Ambitus test, as described previously (Fig. 1)<sup>18,21</sup>. Shortly, access to food was restricted 48 h prior to the test, to keep the animals motivated to perform the task. Two types of tasks were employed: Task 1 was performed (Trials 1-2, 2 min apart; all boxes were filled with reward) in the morning, and Task 2 (Trials 3-4, 2 min apart; only the inside boxes were baited) was performed 3 h later. Thus, for all the parameters, 4 values are available, which allow the determination of their

Feature	Percent missing
Expl_E_I_AFT_Calc	58.22
Expl_I_AFT_Calc	58.22
Expl_E_AFT_Calc	58.22
Expl_I_AFT_T_Calc	57.75
Expl_E_AFT_T_Calc	57.75
Expl_E_I_AFT_T_Calc	57.75
Expl_REP_E_AFT_Nr_Calc	57.73
Expl_REP_AFT_Nr_Calc	57.73
Expl_REP_I_AFT_Nr_Calc	57.73
EAT_E_Nr	50.00
EAT_E_%	50.00
Eff_Expl_E	50.00
L_C_E	50.00
EAT_E_T	50.00
Separation	0.60
Expl_E_BEF_Ratio	0.41
Expl_E_TOT_Ratio	0.41
Circle_Nr	0.41
Dir_Change_Nr	0.41
A_E	0.41
Expl_EI_BEF_Loco_ratio	0.34
Expl_E_I_TOT_Loco_ratio	0.34
Expl_E_BEF_Loco_ratio	0.34
Expl_I_BEF_Loco_ratio	0.34
Expl_E_I_BEF_T	0.30
Expl_I_BEF_T	0.30
Expl_E_BEF_T	0.30
Expl_I_TOT_Loco_ratio	0.30
Expl_E_TOT_Loco_ratio	0.30
Expl_E_AFT_Nr	0.28
Expl_I_AFT_Nr	0.28
Expl_E_AFT_T	0.19
Expl_I_AFT_T	0.19
Expl_E_I_AFT_T	0.19
Expl_I_BEF_Calc	0.02

**Table 1.** Data completeness.

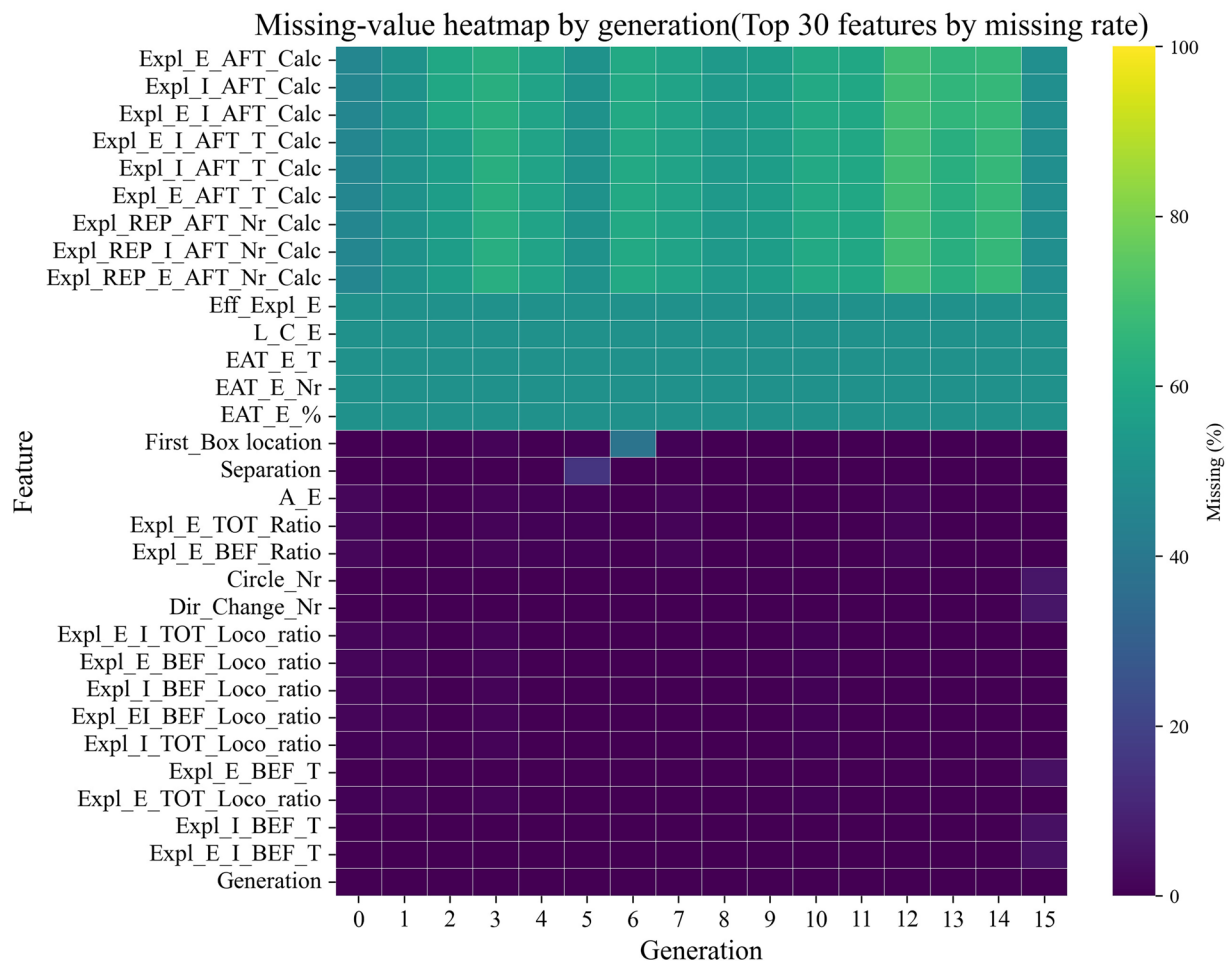
trial-dependent changes. For the purposes of the breeding program, selected behavioral variables were categorized using a quartile based scoring scheme adapted from Kékesi *et al.*<sup>19</sup>. Continuous measures including locomotor and exploratory activity, learning capacity and effective exploration were ranked across animals. Values in the lowest quartile were assigned a score of 0, values in the highest quartile a score of 2, and intermediate values a score of 1. The resulting composite scores were used to classify animals into low and high risk categories for a schizophrenia like phenotype.

## Data Records

The dataset is provided in two files: a separated value (csv) and one parquet file:

- **Raw dataset** (trial-level Parquet): This file contains raw trial-level data, where each row corresponds to a single trial for a specific animal. This allows detailed analysis of within-subject behavior over repeated sessions. The structure and variable definitions are also described in `ambitus_feature_descriptor.pdf` in Zenodo<sup>1</sup>. **File name:** `ambitus_0_15_log_04_08_2025.parquet`
- **Processed dataset** (ML-ready CSV): This file contains one row per animal, summarizing behavioral performance across four Ambitus trials. It does not include features that are used for calculations and the ones where the feature does not exist. Metadata such as group (LE or Lisket), generation, sex, and testing date are also included. A complete description of all variables is provided in the accompanying data descriptor document `ambitus_feature_descriptor.pdf` available in the same Zenodo repository<sup>1</sup>. **File name:** `ambitus_0_15_ml_ready_04_08_2025.csv`

Each animal has a unique identifier and is associated with 4 trials. Feature names are systematically constructed to indicate the source (e.g. trial index) or the applied statistical transformation (e.g., ratio, %, prefixes)



**Fig. 3** Missing-value heat-map by generation. Percent missing values (0–60%) for the 30 most affected features are displayed for generations 0 to 15. Colours range from dark purple (0%) to yellow (60%). Underlying code is available in the GitHub repository (vol 6.1).

A total of 1,342 animals were recorded, consisting of 728 Lisket and 614 Long-Evans (LE) rats (more details about the distribution see in Fig. 2). The group label distribution is nearly balanced across the dataset, which supports robust comparative analysis. These standardized identifiers facilitate cross-comparison and reproducibility.

All files are released under CC-BY 4.0 and are available at Zenodo (<https://doi.org/10.5281/zenodo.16414790>)<sup>1</sup>. The dataset is deposited on Zenodo as a versioned archive (v6.0, released 28 Oct 2025) containing two sub folders:

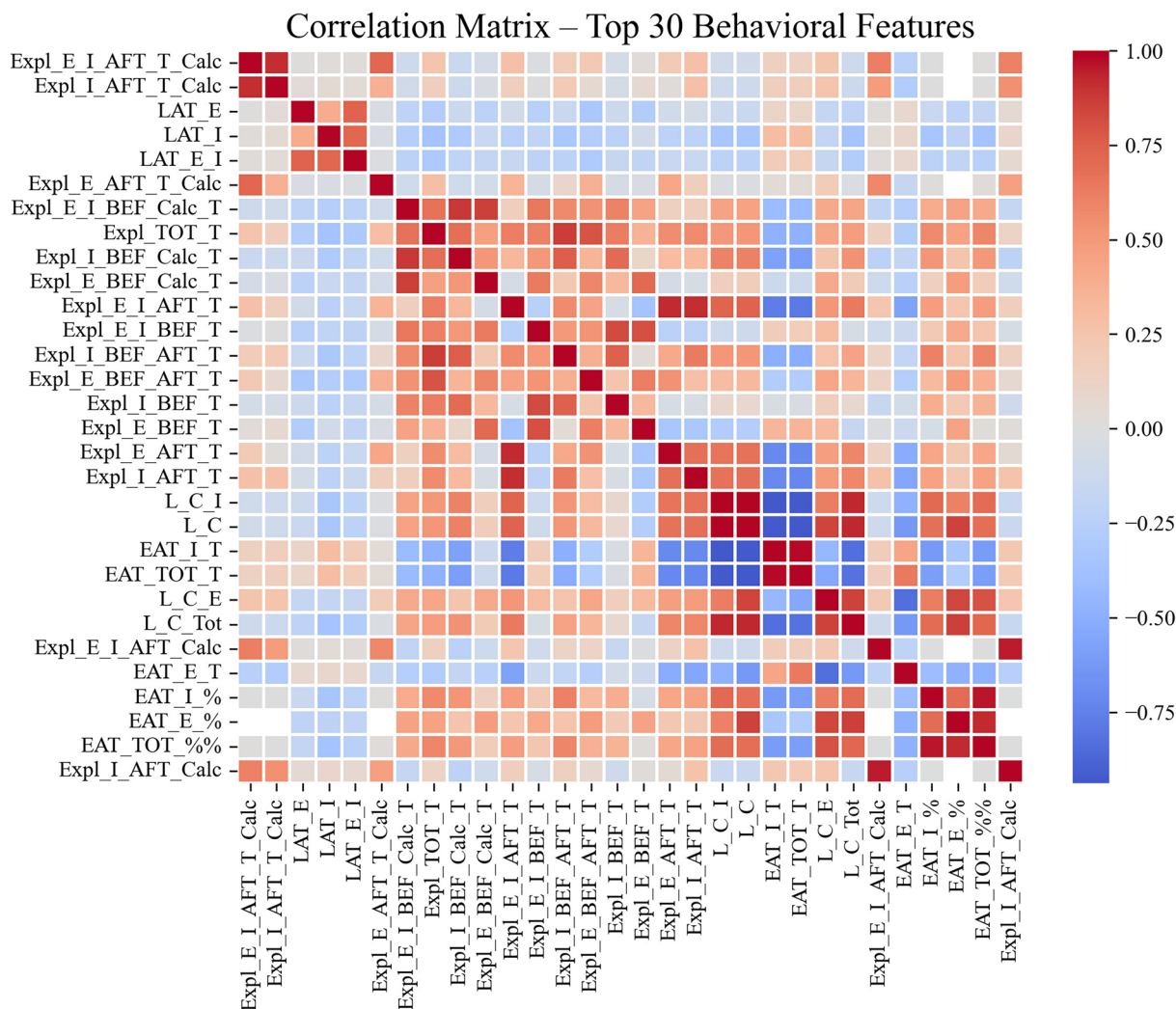
- raw/ambitus\_0\_15\_log\_04\_08\_2025.parquet
- processed/ambitus\_0\_15\_ml\_ready\_04\_08\_2025.csv

Missing values were treated as NA for the parquet and -1 in the CSV file for more detailed information about the files refer for Table 1.

The feature naming convention is followed as Task\_Location\_Measure (e.g. EAT\_E\_Nr, number of the collected rewards from the external boxes) the csv file uses the same convention extended with the “\_Trial” (e.g. EAT\_E\_Nr\_4, number of the collected rewards from the external boxes at the 4th trial).

### Technical Validation

All correlations were computed using Spearman rank correlation (two-tailed, Benjamini-Hochberg (BH)-adjusted). This non-parametric method was selected due to the non-normal distribution of several behavioral variables and the anticipated monotonic, potentially non-linear relationships among them. Spearman correlation provides more robust and reliable estimates in the presence of skewed distributions and outliers, which are common in behavioral datasets. All analyses and figures reported in this section can be fully reproduced with the notebook *technical\_validation.ipynb* (GitHub tag vol 6.1, Zenodo <https://doi.org/10.5281/zenodo.16414790>)<sup>1</sup>.

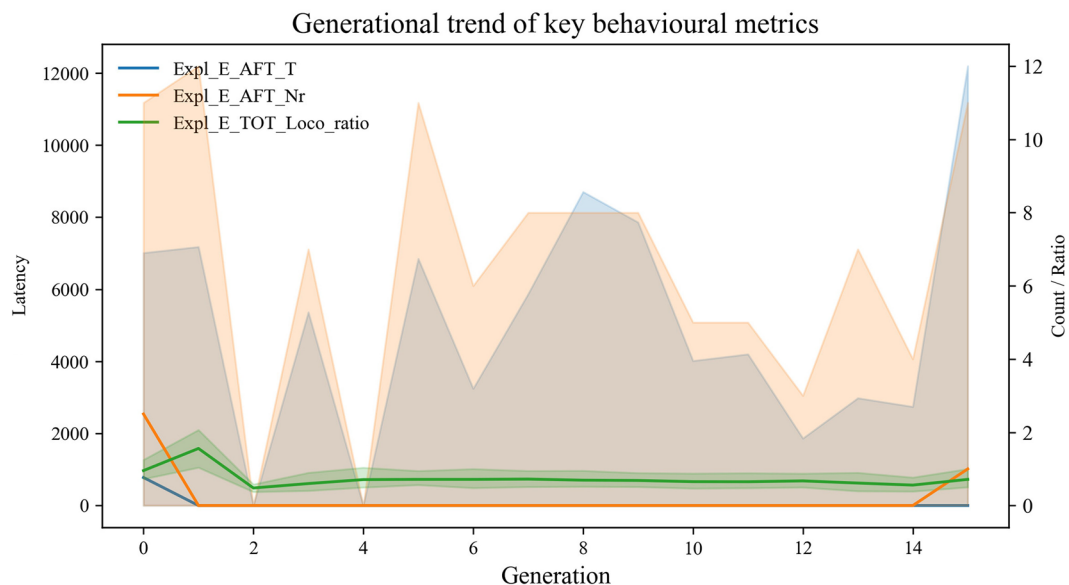


**Fig. 4** Correlation matrix of 30 behavioral features - Spearman correlation coefficients were calculated between the 30 most variable behavioral parameters derived from the Ambitus testing system. These features were selected based on highest standard deviation across all animals, and represent diverse cognitive, locomotor, and reward-related domains. Strong positive and negative correlations ( $|r| > 0.7$ ) indicate partially redundant measures or functionally related processes, such as coordinated reward retrieval and exploratory strategies. Underlying code is available in the GitHub repository (tag vol 6.1).

### Data completeness

We quantified the amount and structure of missing data in the processed table that contains one row per animal (1,342 rows  $\times$  91 features; trial-wise variables are aggregated). For each variable we calculated the proportion of missing entries (pandas 2.3, pair-wise deletion) and summarised the results in Table 1. Missing values primarily arise from low behavioural activity, which resulted in incomplete reward collection and consequently the absence of “after” measurements. This effect was more pronounced in the Lisket strain, leading to a higher frequency of “non-visit” trials. The prevalence of such trials represents an informative behavioural characteristic that can be quantified and analysed in downstream studies.

- **Overall coverage:** Eighty-three of the 91 variables show  $<1\%$  missing values; 55 of these are fully complete. The remaining eight variables belong to the Exploratory-after-task latency group (Expl\_E/I\_AFT\_Calc derivatives) and exhibit the highest loss (*maximum*  $\approx 60\%$ ; Fig. 3). These metrics are absent whenever an animal failed to visit any side-box during the post-task interval—an expected, protocol-dependent limitation rather than technical failure.
- **Generation consistency:** A heat-map of the 30 most affected features (Fig. 3) indicates that the missing-value pattern is remarkably uniform across the 16 generations: the small rise seen in G14-G15 mirrors the slightly shorter average trial length in these cohorts, not a sensor outage. This supports the assumption that no generation-specific dropout or batch effect is present.
- **Encoding and handling:** In the released Parquet table missing entries are stored as NaN; in the legacy CSV logs the same positions are coded as  $-1$  and converted to NaN during preprocessing. Users may safely apply



**Fig. 5** Generational trend of three representative exploratory metrics (median  $\pm$  IQR). Lines show the generation-wise median; shaded areas indicate the inter-quartile range. All three variables - Expl\_E\_AFT\_T (blue), Expl\_E\_AFT\_Nr (orange) and Expl\_E\_TOT\_Loco\_ratio (green) - display small, non-systematic fluctuations, supporting the conclusion of negligible generational drift. (latency = ms, count = events, ratio = unitless). Underlying code is available in the GitHub repository (tag vol 6.1).

list-wise or multiple imputation without violating data-integrity assumptions; an example SimpleImputer workflow is provided in the GitHub notebook *technical\_validation.ipynb* (tag vol 6.1)<sup>1</sup>.

Together, these analyses confirm that the dataset is >99% complete for the vast majority of variables and that the limited missingness is protocol-driven, well documented, and evenly distributed—satisfying the FAIR “quality and reusability” criterion.

Missing values in the legacy parquet file (raw/ambitus\_0\_15\_log\_04\_08\_2025.parquet) arise only when a parameter is undefined for a given task variant (e.g., latency-to-internal-box metrics in Task 2) or when an entire trial failed. In the ML-ready CSV table these cells are encoded as -1 and automatically converted to NaN during preprocessing. Because 83 of 91 features show <1% missingness and the remaining gaps are protocol-dependent (Table 1, Fig. 2), standard list-wise deletion or multiple imputation can be applied without biasing downstream analyses.

### Feature Correlation Analysis

To evaluate internal consistency and redundancy among the 91 behavioral metrics, we calculated pairwise Spearman rank correlation coefficients (pairwise deletion,  $n = 1,342$ ) using pandas 2.3 and visualised the 30 most variable features (selected by standard deviation) as a heat-map (Fig. 4; notebook *technical\_validation.ipynb* code in GitHub tag vol 6.1). The calculation does not distinguish between trials.

The resulting heatmap (Fig. 4) revealed several interpretable correlation clusters:

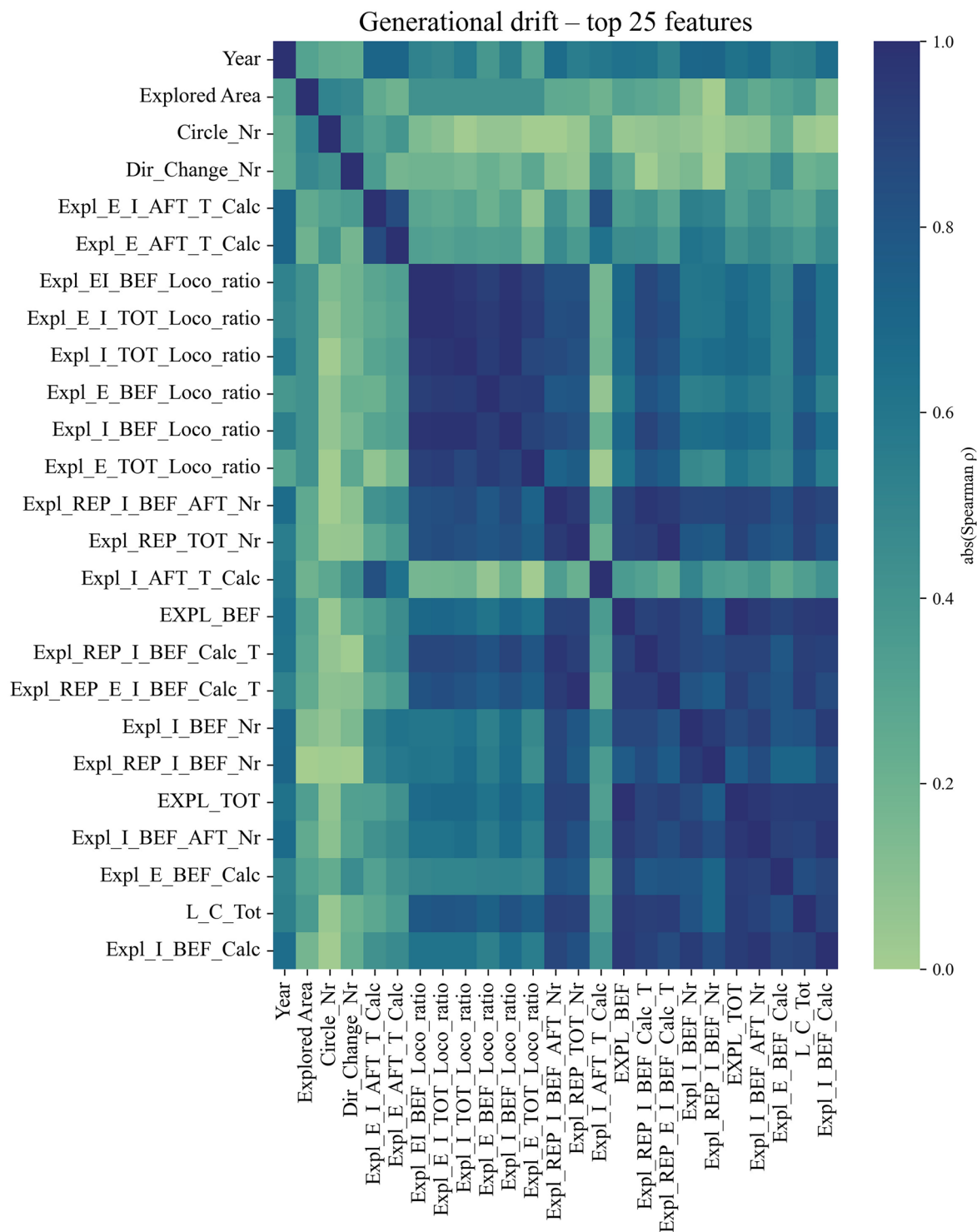
- Features derived from similar behavioral constructs (e.g., total exploration time and box visit counts) showed strong positive correlations, indicating potential redundancy or shared underlying processes.
- Measures associated with learning capacity and efficient task performance (e.g., reward collection efficiency, latency to first reward) formed distinct but moderately correlated groups.
- Some parameters (e.g., ratios or normalized variables) exhibited reduced correlation with absolute raw metrics, reflecting their design to capture relative performance patterns.

This analysis highlights both the internal consistency of feature groups and the presence of partially overlapping information, supporting the rationale for dimensionality reduction or feature selection in downstream modeling tasks.

### Generational drift

To visualise the absence of systematic cohort effects on well-populated variables, we replaced the two latency metrics that carry the highest proportion of missing values with three representative features that (i) rank among the 20 strongest raw  $\rho$  values yet (ii) are >98% complete:

- Expl\_E\_AFT\_T - elapsed time (sec) spent in exploratory behavior after task
- Expl\_E\_AFT\_Nr - count of exploratory events after task



**Fig. 6** Generational drift heat-map (top-25 features). Absolute Spearman correlations (0-0.25) between generation number and the 25 variables with the strongest drift. The muted colour range and lack of block structure indicate minimal lineage-specific change. Underlying code is available in the GitHub repository (tag vol 6.1).

- Expl\_E\_TOT\_Loco\_ratio - ratio of exploration to locomotion during the entire session

Median  $\pm$  IQR curves for these metrics (Fig. 5) remain essentially flat across the 16 generations, with overlapping inter-quartile bands and no monotonic trend, mirroring the low absolute Spearman coefficients observed. Together with the global heat-map (Fig. 6) this confirms that any residual drift is minimal, protocol-related and unlikely to bias downstream analyses.

Task	median_rho	IQR_rho	n_vars	n_reliable
Task1	0.44	0.32–0.59	91	17
Task2	0.40	0.29–0.51	86	8

**Table 2.** Task-level Spearman summary.

Together these findings indicate that the behavioral phenotypes are qualitatively conserved over the seven-year, 16-generation span, satisfying the dataset's internal-consistency requirement and supporting its reuse for downstream modelling.

### Within-task concordance

Because Trials 1–2 (Task 1) and Trials 3–4 (Task 2) were separated by three hours and differed in bait layout, we evaluated reproducibility with non-parametric Spearman rank correlations coefficients rather than intraclass correlation coefficient (ICC). Median  $\rho$  was 0.44 for Task 1 and 0.40 for Task 2 (IQR 0.32–0.59 and 0.29–0.51, respectively; Table 2). Locomotor activity and reward-collection counts displayed high concordance ( $\rho \geq 0.80$ ), confirming sensor stability, whereas exploration-latency metrics showed low agreement ( $\rho \leq 0.25$ ), consistent with rapid habituation and satiety rather than measurement error. The accompanying notebook *technical\_validation.ipynb* (GitHub tag vol 6.1)<sup>1</sup> reproduces the analysis.

### Usage notes

This dataset is suitable for supervised and unsupervised behavioral classification, longitudinal modeling of phenotypic drift, and generational tracking of cognitive function. It should be considered that cognitive functions are influenced by several factors, including attention, motivation, and locomotor activity; thus, the learning process cannot be measured in isolation. However, further correlation analyses between the parameters of behavioral activities and cognitive functions may reveal their associations. Users should be aware that certain behavioral metrics may be sensitive to environmental variables such as handling and lighting, particularly in early generations. Trials were conducted under standardized protocols, but raw trial-level data should be normalized before direct comparison across animals. All accompanying Jupyter notebooks (GitHub tag vol 6.1, Zenodo <https://doi.org/10.5281/zenodo.16414790>)<sup>1</sup> include clearly marked “playground” cells, allowing users to tweak preprocessing steps, model hyper-parameters and validation schemes.

### Data availability

All raw behavioural logs, aggregated Parquet/CSV tables, metadata files and folder structure described in this Data Descriptor are openly available on Zenodo under the accession <https://doi.org/10.5281/zenodo.16414790><sup>1</sup>. The dataset is released under a CC-BY 4.0 licence. All files required to reproduce the figures, analyses and supplementary materials can be accessed and downloaded without restriction.

### Code availability

All Python scripts and Jupyter notebooks used for feature extraction, preprocessing and illustrative analyses are archived at GitHub (<https://github.com/szegedai/ambitus-rat-behaviour-dataset>, release vol 6.1) and Zenodo (<https://doi.org/10.5281/zenodo.16414790>)<sup>1</sup>. The repository is distributed under the MIT licence and includes a requirements.txt file to facilitate full reproducibility.

Received: 11 August 2025; Accepted: 27 January 2026;

Published online: 09 February 2026

### References

- Körösi, G., Czibalmos, O., Kekesi, G. & Horváth, G. Behavioral dataset for Long-Evans and its schizophrenia-like substrain through several generations. Zenodo <https://doi.org/10.5281/zenodo.16414790> (2025).
- Holahan, M. R. *et al.* Spatial learning induces presynaptic structural remodeling. *Hippocampus* **16**, 560–570, <https://doi.org/10.1002/hipo.20185> (2006).
- Kumar, G. *et al.* Strain-dependent effects in a rat touchscreen battery of cognition. *Physiol. Behav.* **144**, 26–36, <https://doi.org/10.1016/j.physbeh.2015.03.001> (2015).
- Peinado, A. & Abrams, C. K. Spontaneous local network activity in cerebral cortex. *PLoS One* **10**, e0131259, <https://doi.org/10.1371/journal.pone.0131259> (2015).
- Casarrubea, M. *et al.* Temporal patterns analysis of rat behavior in hole-board. *Behav. Brain Res.* **208**, 124–131 (2010).
- Chambon, C. *et al.* A new automated method to assess the rat recognition memory: validation of the method. *Behav. Brain Res.* **222**, 151–157 (2011).
- Fitzgerald, R. E. *et al.* Validation of a radial maze test for assessing learning and memory in rats. *Toxicology* **49**, 425–432 (1988).
- Kawabe, K. *et al.* Long-term deficits of radial-arm maze learning. *Brain Res.* **1169**, 77–86 (2007).
- Nagel, J. A. *et al.* Facilitation of tunnel maze performance. *Peptides* **14**, 85–95 (1993).
- Petrovski, Z. *et al.* Gene-environment interactions in rats. *Behav. Brain Res.* **240**, 134–145, <https://doi.org/10.1016/j.bbr.2012.11.022> (2013).
- van der Staay, F. J. *et al.* The appetitively motivated cognitive holeboard. *Neurosci. Biobehav. Rev.* **36**, 379–403 (2012).
- Xu, R. X. *et al.* Development of hexagonal maze for memory evaluation. *Biomed. Rep.* **1**, 134–138 (2013).
- Brown, P. L. *et al.* Altered spatial learning, cortical plasticity and hippocampal anatomy in a neurodevelopmental model of schizophrenia-related endophenotypes. *Eur. J. Neurosci.* **36**, 2773–2781 (2012).
- Bruinsma, B. *et al.* An automated home-cage-based 5-choice serial reaction time task for rapid assessment of attention and impulsivity in rats. *Psychopharmacology* **236**, 2015–2026, <https://doi.org/10.1007/s00213-019-05189-0> (2019).
- Bussey, T. J. *et al.* The touchscreen cognitive testing method for rodents: How to get the best out of your rat. *Learn. Mem.* **15**, 516–523 (2008).

16. Dunlap, A. S. & Stephens, D. W. Experimental evolution of prepared learning. *Proc. Natl Acad. Sci. USA* **111**, 11750–11755 (2014).
17. Dunnett, S. B. *et al.* A novel extended sequence learning task (ESLeT) for rodents. *Brain Res. Bull.* **88**, 237–250 (2012).
18. Horvath, G. *et al.* Characterization of exploratory activity in AMBITUS. *Physiol. Behav.* **169**, 155–164, <https://doi.org/10.1016/j.physbeh.2016.11.039> (2017).
19. Kékesi, G., Petrovzski, Z., Benedek, G. & Horváth, G. Sex specific alterations in behavioral and cognitive functions in a three hit animal model of schizophrenia. *Behav. Brain Res.* **284**, 85–93 (2015).
20. Horvath, G. *et al.* Interaction of clozapine with metformin. *Sci. Rep.* **11**, 16862, <https://doi.org/10.1038/s41598-021-96478-2> (2021).
21. Horvath, G. *et al.* Cognitive training improves the disturbed behavioral architecture. *Physiol. Behav.* **201**, 70–82, <https://doi.org/10.1016/j.physbeh.2018.12.011> (2019).

### Acknowledgements

Project no. 2024-1.2.3-HU-RIZONT-2024-00031 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the 2024-1.2.3-HU-RIZONT funding scheme. This work was supported by University of Szeged Open Access Fund 7966.

### Author contributions

O.C. and G.Kőrösi performed data preprocessing, feature extraction, and implemented machine learning algorithms for behavioral classification. G.H. and G.Kekesi designed and executed the experiments and supervised the selective breeding program.

### Funding

Open access funding provided by University of Szeged.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to G.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026