Title

A reductionist paradigm for high-throughput behavioural fingerprinting in *Drosophila melanogaster*

Authors and affiliations

Hannah Jones[1], Jenny A Willis[2], Lucy C Firth[2], Carlo N G Giachello[2], and Giorgio F Gilestro[1#]

 # To whom correspondence should be addressed. E-mail: giorgio@gilest.ro

[1]Department of Life Sciences, Imperial College London, London, UK

[2]Syngenta, Jealott's Hill International Research Centre, Bracknell, UK

Abstract

Understanding how the brain encodes behaviour is the ultimate goal of neuroscience and the ability to objectively and reproducibly describe and quantify behaviour is a necessary milestone on this path. Recent progresses in machine learning and computational power have pushed the development and adoption of powerful systems leveraging on high-resolution videorecording to track pose and describe behaviour in all four dimensions, however the temporal and spatial resolution of these systems come as a compromise with their throughput and accessibility. Here we describe *coccinella*, an open source reductionist framework that allows for high-throughput analysis of behaviour using real-time tracking on a distributed mesh of microcomputers followed by resource-lean statistical learning. We show that a reductionist system outperforms state-of-the-art alternatives when exploring the pharmacobehaviour in *Drosophila melanogaster*.

Introduction

The nervous system integrates stimuli, internal states, expectations, and previous experience to regulate behavioural output. Describing, quantifying, and modulating behaviour are critical aspects of modern neuroscience and, ever since its inception, the field has spent considerable effort into building and sharing paradigms or tools aimed at objectively and reproducibly quantify behaviours in the most disparate animal models, to the point that this exercise is now recognised as an exciting subfield of neuroscience in its own right: ethomics[1,2]. As the portmanteau name itself suggests, ethomics is not just about describing behaviour (*etho-*) but also about doing so in a high-throughput fashion (*-omics*), collecting data simultaneously from a large number of individuals with no or limited human intervention. Irrespective of the behaviour or the animal model to be analysed, the first compromise a researcher will face when choosing a tool for behavioural quantification will always be between "throughput" and "resolution": a high-throughput analysis will allow for powerful experimental manipulations – such as genetics or pharmacological screenings – offering unbiased approaches in identifying neuronal circuits, genes, molecules underpinning behaviour; high-resolution analysis, on the other hand, promises to identify and discriminate even minuscule differences that may not be immediately visible to the human eye, and to label behaviours into identifiable classes (*e.g.:* "grooming", "courting", "trembling") that may be more relevant to researchers interested in modelling

disease or generally more interested in an anthropomorphic description. In the past years, the field has generally converged towards the adoption of high-resolution video recording of activity, in some cases adopting cameras that have millisecond temporal resolution or developing setups that provide depth information for three-dimensional reconstruction of motion or posture[3–7]. Given the recent evolution in machine learning and progresses in computational power, even these high-resolution analyses can be at least in part compatible with high-throughput approaches[8], especially when employed on small invertebrate animal models[9–12] or when aided by robotic handling[13]. These systems, however, can still be prohibitively expensive for most laboratories, and not easily compatible with throughput in the -omics scale. Moreover, besides the technical urge of removing entry barrier and make ethomics an accessible tool, an equally important underlying questions concerns what is the minimal amount of information that needs to be extracted to identify and classify behaviour. Do we always necessarily gain information from extracting micro-postural features or by analysing activity in three dimensions? To what extent this may actually add counterproductive biological noise to some assays?

Here we introduce *coccinella*: a new experimental framework that combines high-throughput, inexpensive, real-time ethomics[14] with state-of-the-art statistical learning[15,16] to characterise and discriminate complex behaviours using a reductionist approach based solely on one simple feature. *Coccinella* builds on ethoscopes, an accessible open source platform, to extract in real-time minimalist information from flies challenged in different behavioural paradigms. We show that this system outperforms the state-of-the-art alternatives in recognising the pharmacobehavioural space, providing better discernibility at a fraction of the cost, thus opening new doors to high-throughput ethomics.

Results

Drosophila ethomics studies generally rely on image acquisition through so called industrial-cameras, able to collect videos with high temporal and spatial resolution and featuring mounts for a large selection of lenses. Some of the cameras commonly used for these purposes (*e.g.:* FLIR, Point Grey, Basler)[17] are expensive and are normally employed in close-up imaging that microscopically highlights the smallest anatomical features of the animal but at the same time greatly limits the number of experimental subjects that can be recorded by a single device. Normally, each camera would be connected to a dedicated powerful computer for acquisition and storage of the videos. The cost and physical footprint of these setups make them incompatible, at least for most laboratories, with high-

throughput simultaneous acquisition. To lower this barrier, we created a framework that employs the distributing computing power of ethoscopes[14], thus allowing for inexpensive analysis of activity in hundreds or thousands of flies at once. Each ethoscope is powered by a Raspberry Pi microcomputer and collects images through a Raspberry Pi NoIR camera, overlooking a bespoke 3D printed arena hosting 24 individually isolated flies. Digital images are acquired at a resolution of 1280x960 pixels with each animal being constricted in an ellipse measuring 25.8 ± 1.4 x 9.85 ±1.4 pixels (Figure 1a). In this setup, flies are free to move in a circular two-dimensional space with a diameter of 11.5mm designed to maintain the animal in the walking position[18] while venturing on solidified agar providing nutrients, or drugs. At this working resolution, ethoscopes would be capable to acquire videos at 30 fps and store them to be used for offline analysis at a later point, but we instead opted for online tracking and real-time parametric extraction, meaning that images were analysed by each raspberry Pi at the very moment they were acquired (Figure 1b). This latter modality limits the temporal resolution of information being processed (one frame every 444 ms ± 127 ms, equivalent to 2.2 fps on a Raspberry Pi3) but provides the most affordable and high-throughput solution, dispensing the researcher from organising video storage or asynchronous image processing. In previous analysis of activity and sleep[14], we found that the maximal velocity of the fly over a period of 10 seconds best described the basic motion features of the animal, allowing us to accurately differentiate between different activity patterns, such as walking, grooming, feeding[14]. We therefore adopted this reductionist measure for *coccinella* too, ultimately producing monodimensional time-series of a behavioural correlate, which were then digested using highly comparative time-series analysis (HCTSA)[15], a computational framework that effortlessly subjects time-series to more than 7700 literature-relevant statistical tests, looking for meaningful discriminative features. Features successfully extracted through HCTSA were then used to classify behaviour using a linear support vector machine (SVM_{linear})[19] (Figure 1b) and here presented and compared using confusion matrices (Supplementary Figure 1a). To test the ability of this system to discriminate behaviour, we started by exploring the pharmacobehavioral space of flies fed with a panel of known or putative neurotropic chemicals, comprising molecules previously described in the literature along with experimental ones being considered as potential insecticides (Figure 1c, 1d – Supplementary Table 1). Using an initial panel of 17 treatments (16 drugs and one solvent control) we were able to predict compounds with an accuracy of 71.4% (vs. 5.8% of a random classifier – Figure 1c). Some compounds induced behaviour with a particularly high predictive fingerprint: dieldrin, for instance, was predicted with an accuracy of 94% and flonicamid with an accuracy of 87%. For others,

our system fared more poorly (*e.g.* tetramethrin showed 41% accuracy). In all cases, however, the relative confusion was negligible, with all compounds being correctly identified as the first choice and with the first predicted compounds having, on average, a score that was 15-times greater compared to the second best choices (Figure 1c – min.: 3.3x; max.: 65x). To validate our framework and exclude that the system operates by overfitting biologically irrelevant information we follow two lines of control. First, we fed flies with lower concentrations of the same compounds (Figure 1c). Feeding flies with different concentrations of drugs unsurprisingly showed a different effect on short term lethality (Supplementary Figure 1b), with several compounds hitting a 25% lethality rate before the end of the experiment when fed at the highest concentration (1000ppm – Supplementary Figure 1b). In these conditions, the predictive accuracy of the system decreased from 71.4% (1000ppm) to 61.8% (100ppm), falling to 36.1% with the lowest concentrations (1ppm), indicating that the system does operate on pharmacologically induced, biologically meaningful behavioural correlates (Figure 1c). A similar drop in accuracy was observed using a smaller panel of 12 treatments (Supplementary Figure 2a). As a second line of work to test specificity, we obtained genetic mutants known to be resistant to specific pharmacological treatments: the $para^{L1029F}$ allele encodes for a version of the α-subunit of voltage-gated sodium channel conferring resistance to DDT and pyrethroids[20]; the Rdl^{A301S} allele encodes for a version of the ligand-gated chloride channel conferring resistance to dieldrin and fiproles[21]. Challenging these mutants with their respective compounds created confusion in the clustering algorithm for which finding the difference between drug treatment and solvent control became a harder task, especially in the case of DDT and $para^{L1029F}$ (Supplementary Figure 2a). The observed drop in accuracy suggests again that *coccinella* is working on biologically relevant behavioural signatures and the fact that some discrimination can still be observed with targets harbouring point mutations, which severely affecting compound efficacy, is indicative of very high sensitivity.

Having established the accuracy and sensitivity of the system, we next wanted to test its usefulness in a genuine high-throughput scenario. We subjected a total of 2192 flies to a panel of 40 treatments (Figure 1d), mostly featuring known compounds but also two unexplored molecules (Supplementary Table 1). Given that the 100ppm intermediate concentrations showed the best compromise between accuracy and lethality (Supplementary figure 1b), we performed this larger screen using compounds at 100ppm only. Even with such a large panel, the system was able to first guess 39 out of 40 of the tested treatments

(the only exception being the Syngenta Compound #5) with an overall accuracy of 44.5%, vs 2.5% of the random classifier.

A reductionist approach served us well so far, identifying with remarkable accuracy even subtle changes when we explored the pharmacobehavioural space of a large number of neuroactive compounds in wild type and mutant flies. But how does it compare to other more established paradigms? *Coccinella* is arguably to be preferred in terms of accessibility and throughput, but what is the amount of useful information that we are sacrificing? To quantify any possible loss in information content, we run a series of parallel experiments in which flies were fed with a selected panel of 12 treatments (11 drugs and a solvent control, Figure 2) and their behaviour analysed either using *coccinella* or using the state-of-the-art methods, which started from high-resolution imaging and employed supervised machine learning for pose estimation (DeepLabCut[17]) followed by unsupervised identification of behavioural grammar (B-SOiD[4]). To widen the range of comparisons, data were then either immediately clustered using different common clustering algorithms (K-nearest neighbours, random forest) or first processed through a smaller, selected subsample of the HCTSA array (Catch22[16]) before being clustered using SVM_{linear} (Figure 2a). In this challenge, *coccinella* unambiguously identified 10 out of 12 compounds, with poor performance only for two of them (flubendiamide and tetramethrin) and an overall accuracy of 42.4%, vs 8.3% of the random classifier (Figure 2b). Surprisingly, none of the state-of-the-art high-resolution paths did better than this. The combination of pose-estimation → grammar extraction → random forest classification scored as the second best, with an accuracy of 25.4% but with only three compounds being unambiguously identified (Figure 2c). The same experimental dataset clustered with even poorer performance when using K-nearest neighbours (Figure 2e) and even the application of HCTSA features extraction to the B-SoID output still could not match the accuracy observed with *coccinella*'s reductionist approach (Figure 2d). This analysis is not meant to be conclusive. We expect that some alternative combination of state-of-the-art approaches will probably manage to match or likely improve over *coccinella*'s performance, yet the fact we could obtain such an impressive result with a system that is arguably unmatched in terms of throughput and economic cost is, alone, an argument that gives new weight to this (and future) reductionist approaches.

Finally, to push the system to its limit, we asked *coccinella* to find qualitative differences in sleep rebound after different regimes of sleep deprivation. *Drosophila melanogaster* shows a strong,

conserved homeostatic regulation of sleep that forces flies to recover, at least in part, lost sleep (for instance after a night of forceful sleep deprivation)[22,23]. We previously showed that the extent of sleep rebound observed after sleep deprivation only loosely correlates with the amount of lost sleep[24] and it remains an open question, in the field, whether similar amounts of sleep rebound may in fact differ from each other in some inscrutable feature that would underpin a different "sleep depth"[25,26], similarly to what it is believed to happen in mammals. We therefore analysed a dataset of 727 flies that experienced different regimes of mechanically enforced sleep deprivation during the 12 hours of the night (Figure 3). Flies were housed in tubes that would rotate after a set time of inactivity ranging from 20 to 1000 seconds leading to different degrees of sleep restriction (Figure 3a, dataset sampled from[24]). In this experimental paradigm, all treatments led to a statistically significant rebound compared to the undisturbed control animals (Figure 3b). We then run *coccinella* on the two subsets of the panel: the baseline data, acquired the morning before the sleep deprivation (Figure 3c), and the rebound data, on the morning after (Figure 3d). Unsurprisingly, we could not detect any internal biological difference in the pre-sleep deprivation control set, featuring flies of identical genotype and age housed in different tubes before the sleep deprivation treatment. In these conditions, *coccinella* could not discern, and performed exactly as a random classifier would (9% vs 9% – Figure 3c). However, analysis of those same animals during rebound after sleep deprivation showed a clear clustering, segregating the samples in two subsets with separation around the 300 seconds inactivity trigger (Figure 3d). This result is important for two reasons: on one hand, it gives, for the third time, strong evidence that the system is not simply overfitting data of no biological significance given that it could not perform any better than a random classifier on the baseline control. On the other hand, *coccinella* could find biologically relevant differences on rebound data after different regimes of sleep deprivation. Interestingly enough, the 300 seconds threshold that *coccinella* identified has a deep intrinsic significance for the field, for it is considered to be the threshold beyond which flies lose arousal response to external stimuli, defining a "sleep quantum", the minimum amount of time required for transforming inactivity bouts into sleep bouts[22,23].

Discussion

System neuroscience is living a period of renaissance and large scale quantification of behaviour is driving this revolution, together with a plethora of new genetic tools that allow thermo- and opto-

genetic manipulations and – at least in *Drosophila* – a galore of genetic transformants for circuit tracking and manipulation. Progresses in machine learning and computer power have had a massive impact on the field of ethomics, especially in achieving levels of anatomical tracking that allow mapping of the tiniest movements on an experimental animal model with the highest temporal resolution and with little human supervision[6,17]. Most of these system, however, rely on relatively expensive setups and do not scale easily to high-throughput experimental paradigms. They are ideal to identify behavioural patterns and study fine motor control but may be undue for many other uses. Here we introduce a new framework, *coccinella*, that merges an open-source, economically accessible hardware platform (ethoscopes[14,27]) with a powerful toolbox for statistical analysis and clustering (HTCSA[15] / Catch22[16]). *Coccinella* is a reductionist tool not meant to replace the behavioural categorization that other tools can offer: it cannot apply anthropomorphic labels to behaviour ("courting", "lounging", "sipping", "jumping" *etc.*) but it can successfully identify large behavioural phenotypes and generate unbiased hypothesis on how behaviour – and a nervous system at large – can be influenced by chemicals, genetics, artificial manipulations in general. We provide evidence that *coccinella* can be used to successfully explore and compartmentalise the pharmacobehavioural space, possibly opening new avenues for drug discovery using *Drosophila melanogaster* as a model. The success of *Drosophila* as experimental model was built on the many genetic screens of the 1900s. We propose *coccinella* as an accessible, pivotal tool to restart this important line of work in any laboratory, without funding or access to technology being a discriminative factor.

Acknowledgements

We thank the Gilestro lab at Imperial College London and Robert Lind at Syngenta for useful discussions. HJ was supported by a BBSRC/CASE studentship in partnership with Syngenta (project reference BB/M011178/1/1958700). Stock obtained from the Bloomington Drosophila Stock Center (NIH P40OD018537) were used in this study.

Methods

Drosophila rearing

Fly lines were maintained on a 12-hour light: 12-hour dark (LD) cycle and raised on standard corn and yeast media (standard food). Canton-Special (CS) *Drosophila melanogaster* were used as the wild-type line for all experiments.

Drug resistant mutants

Rdl^{A301S} is derived from *Rdl*^{MDRR} (RRID:BDSC_1675), an *Rdl* allele isolated from a natural population in Maryland[28], and underwent isogenization and selection on dieldrin to eliminate the metabolic resistance and maintain the dieldrin target site resistance[29]. The *Rdl*^{MDRR} was obtained from the Bloomington Drosophila Stock Center. For *para:* the L1029F mutation, located in the voltage gated sodium channel paralytic, has been extensively reported to confer resistance to DDT and pyrethroids in many other insect species (called kdr, knockdown resistance, reviewed in[30]). In the *Drosophila* gene, kdr maps to L1029F and is equivalent to the often cited L1014F in other insects (e.g. *Musca domestica*[31]). The kdr L1029F mutation in *Drosophila* Para was introduced via CRISPR/Cas9-mediated genome editing (see Supplementary Materials and Methods). This genome edited generated mutation resulted in a similar resistance to DDT as previously reported[32].

Handling and preparation of drugs

All insecticide compounds were supplied by Syngenta Ltd. from their in-house stock (see Supplementary table 1 for a full list of compounds used). Compounds were received in solid form and diluted in solvent containing 5% ethanol (VWR, 20821), 5% acetone (Sigma,179124) and 10% dimethylsulfoxide (DMSO) (D2650, Sigma) in distilled water to 1000ppm initially, then further diluted in the solvent mixture to 100ppm and 1ppm (where 1mL/L = 1000ppm). For insecticide assays, 0.5mL of 5% sucrose (Sigma, S0389), 1% agarose (Sigma, A6236) solution was pipetted into each well and allowed to set. Following this, 2µL of compound solution were placed on the surface and allowed to dry for 30 minutes or more. Male flies were then placed on the surface with a small glass cover slip placed on top (13mm circular cover slip, VWR631-0150). Flies were briefly anaesthetised (>1 minute) before being placed onto the surface of the plate. Once each well had been filled with a single male fly, arenas were placed into the ethoscope and recorded for a minimum of 2 days. All experiments were started between ZT0 and ZT1 and within 30 minutes of the flies being placed in the wells. For each compound in Figures 1c and Supplementary Figure 1c, 3 repeats were done at different time points. For Figure 1D, two repeats to three biological repeats per compound.

Data acquisition and processing

Ethoscope data were first processed in R using rethomics[27]. Each time series was exported (.csv) and converted using Python to individual time series in a file format (.dat) compatible with MatLab. A metadata (.txt) file served as a reference file of each individual time series with keywords outlining compound groups and concentrations for processing data using HCTSA.

HCTSA / Catch22

Following this process, HCTSA feature extraction was performed on the time-series data (for figures 1C, 1D, 3C, 3D, Suppfig1, Suppfig2A). After the features were extracted, outputs of error-producing operations were removed through a normalization process using a sigmoidal transformation. HCTSA inbuilt functions were then used to classify data using a linear SVM classifier and a confusion matrix comparing the time series was generated. For some of the time-series data (that in figures 2B, 2D, Suppfig2B), a smaller feature set of HCTSA, Catch-22 was used for feature extraction. Due to the smaller number of features used with this method, normalization was not required before using a linear SVM classifier to generate a confusion matrix comparing the results. A time series of 12 hours was used for the HCTSA analysis in Figure 1 and Supplementary Figure 1,2. A length of 3 hours was used for the HCTSA analysis in Figure 3. All video data in Figure 2 are from time series of 15 minutes.

Video generation for flies on insecticides

Custom 3D-printed squares were designed using the online CAD software Onshape and printed using Ultimaker 2+ 3D printers using PLA plastic. For insecticide assays, 0.5mL of 5% sucrose (Sigma, S0389), 1% agarose (Sigma, A6236) solution was pipetted into each well and allowed to set. Following this, 2µL of compound solution were placed on the surface and allowed to dry for 30 minutes or more before male flies were placed on the surface with a small glass cover slip placed on top (13mm circular cover slip, VWR631-0150). Flies were briefly anaesthetised (>1 minute) before being placed onto the surface of the square. Once each well had been filled with a single male fly, squares were placed in the arena and a video was recorded for a minimum of 12 hours using an ELP 8 megapixel camera with an IMX179 Sensor and 2.8-12 mm variable focus manual lens. All videos recordings were started between ZT0 and ZT1. Recordings of flies exposed to compounds were done in a randomised manner. Video data was then broken down into shorter segments of 15-minute videos for processing. The first 15-minutes following 1 hour of fly exposure to compound or control was used for pose extraction.

Deeplabcut / B-SoID

The use of DeepLabCut (version 2.1) followed the detailed protocol outlined by[5] . Briefly, frames for labelling were extracted from 3 representative videos using a K-means algorithm and frames were labelled with 22 unique body parts (head, left eye, right eye, thorax top, thorax bottom, abdomen top, abdomen middle, abdomen bottom, left wing tip, right wing tip, left foreleg tip, left foreleg middle, right foreleg tip, right foreleg middle, left middle leg tip, left middle leg middle, right middle leg tip, right middle leg middle, left back leg tip, left back leg middle, right back leg tip, right back leg middle). These frames were labelled locally with a DeepLabCut graphical user interface (GUI) before the project file was uploaded to Google Drive for training and video analysis to be done using Google Colab. The data was split into a 9:1 test:train dataset and training was run for more than 150,000 iterations before the average Euclidean error was computed between labels and predictions. The model at the best performing checkpoint was used to predict pose in novel videos. Following this, B-SOID was used to de-structure behaviour using the output of DeepLabCut and generate fly-specific time-series of behavioural grammar as in[4].

Sleep deprivation

Maximal velocity time series data generated from recording flies exposed to either control conditions (no SD) or incrementally increasing immobility-triggered SD conditions were taken from the dataset generated by Geissmann *et al*[24]. Only data for female flies were taken, and where groups consisted of more than 60 individuals, a sample of 60 random individual flies were chosen to take time series data from. Time series data were processed in the same way as described in section 4.2 using the raw ethoscope data.

Survival

Flies were fed with the specified compounds at the desired concentrations and concomitantly analysed in ethoscopes for 24 hours. Time of death was calculated as the last moment of detected motion.

Code and data availability

A notebook version of the source code used to generate all figures is available on the Zenodo public repository, along with all the metadata and the raw data collected in this study (DOIs: 10.5281/zenodo.7335575 and 10.5281/zenodo.7393689). Data were analysed using rethomics[33] and ethoscopy.

References

1. Brown, A. E. X. & Bivort, B. de. Ethology as a physical science. *Nat. Phys.* 1 (2018) doi:10.1038/s41567-018-0093-0.
2. Datta, S. R., Anderson, D. J., Branson, K., Perona, P. & Leifer, A. Computational Neuroethology: A Call to Action. *Neuron* **104**, 11–24 (2019).
3. Wiltschko, A. B. *et al.* Mapping Sub-Second Structure in Mouse Behavior. *Neuron* **88**, 1121–1135 (2015).
4. Hsu, A. I. & Yttri, E. A. B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors. *Nat. Commun.* **12**, 5188 (2021).
5. Nath, T. *et al.* Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat. Protoc.* **14**, 2152–2176 (2019).
6. Pereira, T. D. *et al.* Fast animal pose estimation using deep neural networks. *Nat. Methods* **16**, 117–125 (2019).
7. Gosztolai, A. *et al.* LiftPose3D, a deep learning-based approach for transforming two-dimensional to three-dimensional poses in laboratory animals. *Nat. Methods* **18**, 975–981 (2021).
8. Wiltschko, A. B. *et al.* Revealing the structure of pharmacobehavioral space through motion sequencing. *Nat. Neurosci.* **23**, 1433–1443 (2020).
9. McDermott-Rouse, A. *et al.* Behavioral fingerprints predict insecticide and anthelmintic mode of action. *Mol. Syst. Biol.* **17**, e10267 (2021).
10. Ayroles, J. F. *et al.* Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proc. Natl. Acad. Sci.* **112**, 6706–6711 (2015).
11. Branson, K., Robie, A. A., Bender, J., Perona, P. & Dickinson, M. H. High-throughput ethomics in large groups of Drosophila. *Nat. Methods* **6**, 451–457 (2009).
12. Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S. & Branson, K. JAABA: interactive machine learning for automatic annotation of animal behavior. *Nat. Methods* **10**, 64–67 (2013).

13. Alisch, T., Crall, J. D., Kao, A. B., Zucker, D. & de Bivort, B. L. MAPLE (modular automated platform for large-scale experiments), a robot for integrated organism-handling and phenotyping. *eLife* **7**, e37166 (2018).
14. Geissmann, Q. *et al.* Ethoscopes: An Open Platform For High-Throughput Ethomics. *PLOS Biol.* **15**, e2003026 (2017).
15. Fulcher, B. D. & Jones, N. S. hctsa: A Computational Framework for Automated Time-Series Phenotyping Using Massive Feature Extraction. *Cell Syst.* **5**, 527-531.e3 (2017).
16. Lubba, C. H. *et al.* catch22: CAnonical Time-series CHaracteristics. *Data Min. Knowl. Discov.* **33**, 1821–1852 (2019).
17. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
18. Simon, J. C. & Dickinson, M. H. A New Chamber for Studying the Behavior of Drosophila. *PLOS ONE* **5**, e8793 (2010).
19. Chatterjee, S., Dey, D. & Munshi, S. Chapter 4 - Feature selection and classification. in *Recent Trends in Computer-Aided Diagnostic Systems for Skin Diseases* (eds. Chatterjee, S., Dey, D. & Munshi, S.) 95–135 (Academic Press, 2022). doi:10.1016/B978-0-323-91211-2.00001-9.
20. Kaduskar, B. *et al.* Reversing insecticide resistance with allelic-drive in Drosophila melanogaster. *Nat. Commun.* **13**, 291 (2022).
21. Remnant, E. J. *et al.* The role of Rdl in resistance to phenylpyrazoles in Drosophila melanogaster. *Insect Biochem. Mol. Biol.* **54**, 11–21 (2014).
22. Shaw, P. J., Cirelli, C., Greenspan, R. J. & Tononi, G. Correlates of Sleep and Waking in Drosophila melanogaster. *Science* **287**, 1834–1837 (2000).
23. Hendricks, J. C. *et al.* Rest in Drosophila Is a Sleep-like State. *Neuron* **25**, 129–138 (2000).
24. Geissmann, Q., Beckwith, E. J. & Gilestro, G. F. Most sleep does not serve a vital function: Evidence from Drosophila melanogaster. *Sci. Adv.* **5**, eaau9253 (2019).

25. French, A. S., Geissmann, Q., Beckwith, E. J. & Gilestro, G. F. Sensory processing during sleep in Drosophila melanogaster. *Nature* (tbc) doi:10.1038/s41586-021-03954-w.
26. Wiggin, T. D. *et al.* Covert sleep-related biological processes are revealed by probabilistic analysis in Drosophila. *Proc. Natl. Acad. Sci.* **117**, 10024–10034 (2020).
27. Geissmann, Q., Rodriguez, L. G., Beckwith, E. J. & Gilestro, G. F. Rethomics: an R framework to analyse high-throughput behavioural data. *bioRxiv* 305664 (2018) doi:10.1101/305664.
28. Ffrench-Constant, R. H., Roush, R. T., Mortlock, D. & Dively, G. P. Isolation of Dieldrin Resistance from Field Populations of Drosophila melanogaster (Diptera: Drosophilidae). *J. Econ. Entomol.* **83**, 1733–1737 (1990).
29. Blythe, J. *et al.* The mode of action of isocycloseram: A novel isoxazoline insecticide. *Pestic. Biochem. Physiol.* **187**, 105217 (2022).
30. Arena, J. P., Liu, K. K., Paress, P. S., Schaeffer, J. M. & Cully, D. F. Expression of a glutamate-activated chloride current in Xenopus oocytes injected with Caenorhabditis elegans RNA: evidence for modulation by avermectin. *Mol. Brain Res.* **15**, 339–348 (1992).
31. Dong, K. Insect sodium channels and insecticide resistance. *Invert. Neurosci.* **7**, 17 (2007).
32. Samantsidis, G.-R. *et al.* 'What I cannot create, I do not understand': functionally validated synergism of metabolic and target site insecticide resistance. *Proc. R. Soc. B Biol. Sci.* **287**, 20200838 (2020).
33. Geissmann, Q., Rodriguez, L. G., Beckwith, E. J. & Gilestro, G. F. Rethomics: An R framework to analyse high-throughput behavioural data. *PLOS ONE* **14**, e0209331 (2019).

Figure 1 – *Coccinella* successfully classifies pharmacobehaviours. a) Bespoke 3D printed arena can house 24 individual flies in a two-dimensional circular space. Each arena measures 11.5 mm in diameter and allows for back illumination either by visible or infrared light sources embedded in the ethoscope base. The red ellipse shows the data being extracted by the ethoscope in real time. w, *h*: width and height of the ellipse inscribing the animal. *φ*: the angle of the ellipse in reference to the

region of interest. *max. vel.*: maximal velocity over the last 10 seconds. **b)** Flowchart describing the analytical pipeline and the tools that compose *coccinella*. **c)** Confusion matrices for treatments with 16 compounds and a solvent control, with drugs used at concentrations of 1000ppm (left), 100ppm (center), 1ppm (right). The target icon indicates calculated accuracy while the rolling die indicate the accuracy of a random classifier. **d)** Confusion matrix for the largest panel of 40 treatments at 100ppm.

Figure 2 – Comparison between *coccinella* and the state-of-the-art. a) Experimental pipeline illustrating the four experimental analysis. **b)** Confusion matrix for 12 treatments (11 drugs at 1000ppm and one solvent control) analysed using *coccinella*. **c)** Same experimental treatments as in b, analysed using the DeepLabCut → B-SoID → Random forest pipeline starting from high resolution images. The Random forest classifier was trained on a 4:1 training:testing ratio. **d)** Same as c but with Catch22 identification and SVM clustering after B-SoID grammar dissection. This is a hybrid treatment combining HCTSA feature extraction to the high resolution pipeline. **e)** Same as c but using K-nearest neighbours (KNN) as cluster algorithm. KNN required a much higher training:testing ratio of 9:1, dramatically reducing the size of the testing dataset. The accuracy of a random classifier for all matrices on this figure would be 8.3% (not shown on figures for lack of space).

Figure 3 – *Coccinella* finds differences in type of sleep rebound after sleep deprivation. a) Sleep profile of flies over the period of three days. A 12 hours sleep deprivation regime starts at the beginning of the dark phase of day 0 (purple bar). The three-hours windows labelled with green boxes were analysed by *coccinella* in search of meaningful differences. The letters above refer to the panels using data in those time windows **b)** Extent of rebound as observed following sleep deprivation as performed in a. Panel a and b reproduce data from[24]. **c)** Confusion matrix showing the classification using *coccinella* of the baseline time-series. No accuracy gain compared to the random classifier. **d)** Confusion matrix of the rebound data. The classification finds two clusters, separated by the 300 seconds threshold (thick black lines).

Supplementary Figure 1. a) Explanatory drawing of how confusion matrices present the data. The blue diagonal boxes indicate the number of flies correctly classified (true positives). The vertical cells

indicate the number of flies wrongly classified (false positives). The horizontal cells indicate the number of flies that should have been classified (false negatives). **b)** Survival curves upon treatment with a panel of 12 drugs administered at different concentrations, indicated above each figure (1ppm, 100ppm, 1000ppm). **c)** Confusion matrices for treatments with 11 compounds and a solvent control, with drugs used at concentrations of 1000ppm (left), 100ppm (middle), 1ppm (right). The target icon indicates calculated accuracy. Accuracy of a random classifier would be 8.3%.

Supplementary Figure 2. a) Confusion matrix classifying the action of two selected drugs (DDT 100ppm and Dieldrin 100ppm) or solvent on wild type flies (WT) or mutants with known resistance to those drugs. The LD100 of Dieldrin is 0.1 - 1ppm and the LD100 of DDT is 10-100ppm. The three purple dotted boxes highlight the three experimental clusters (wild-type flies, *para* mutants, *Rdl* mutants)

Supplementary Table 1. List of all the compounds used in this study with bibliographic reference.

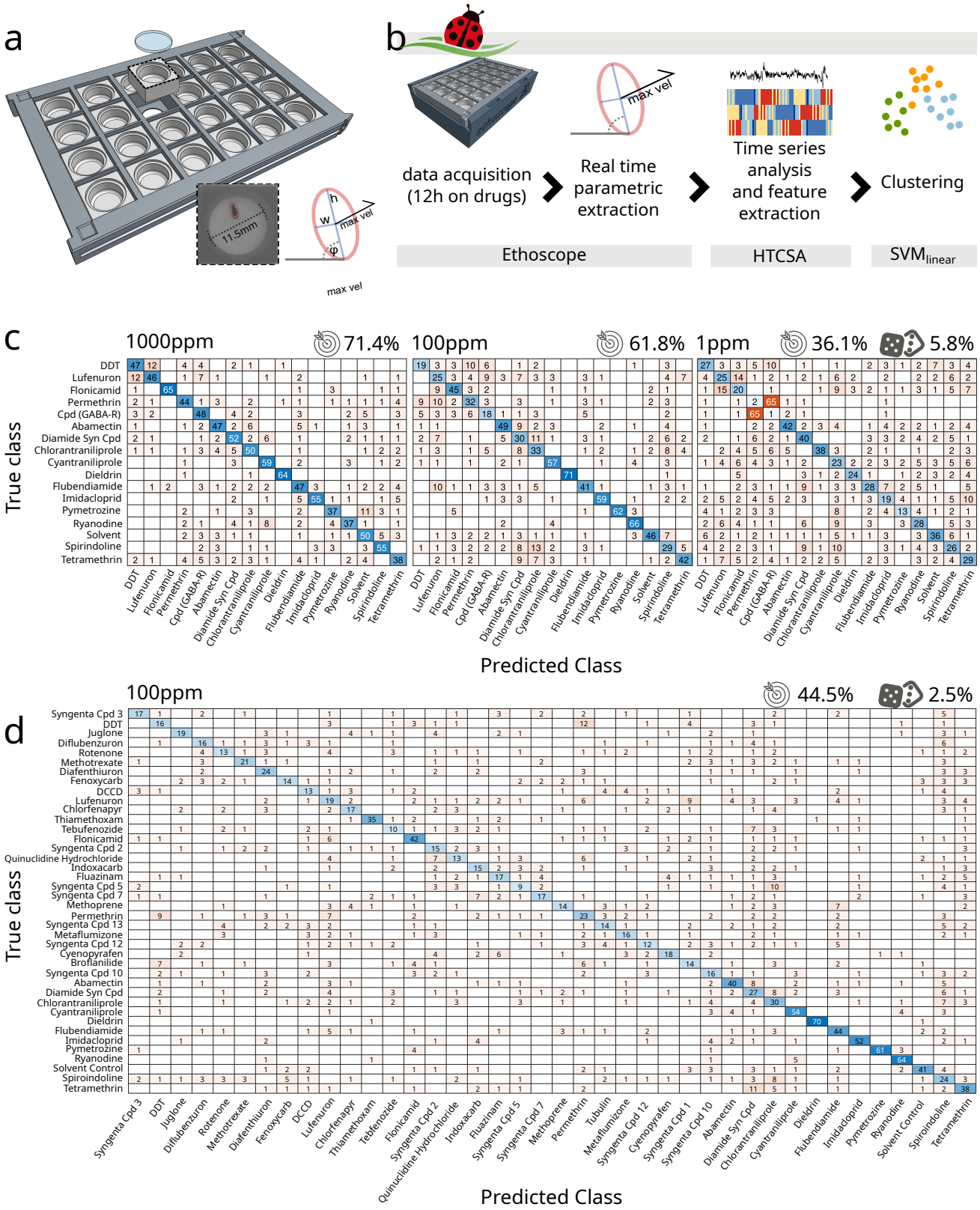


Figure 1

Jones et al 2022

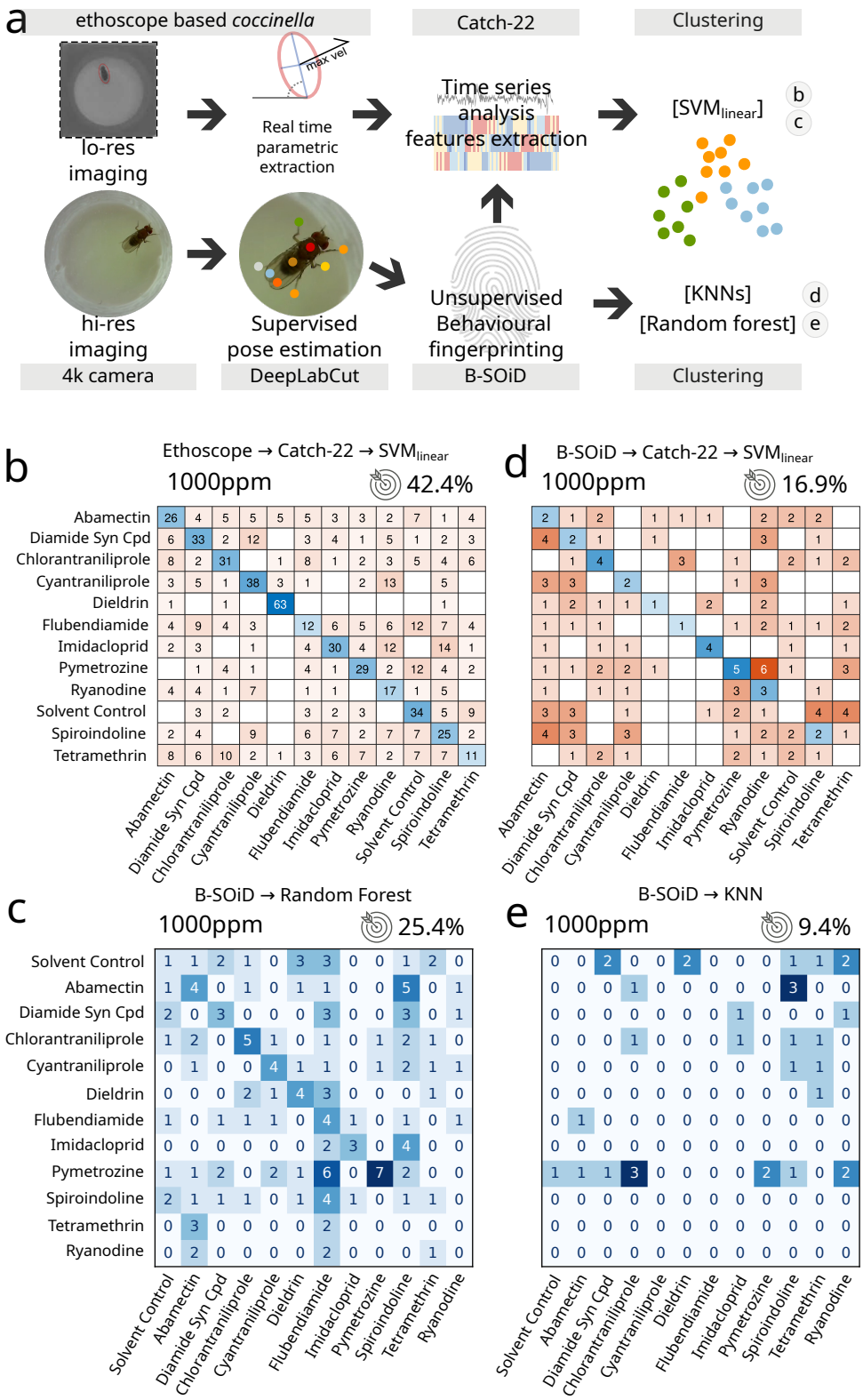


Figure 2

Jones et al 2022

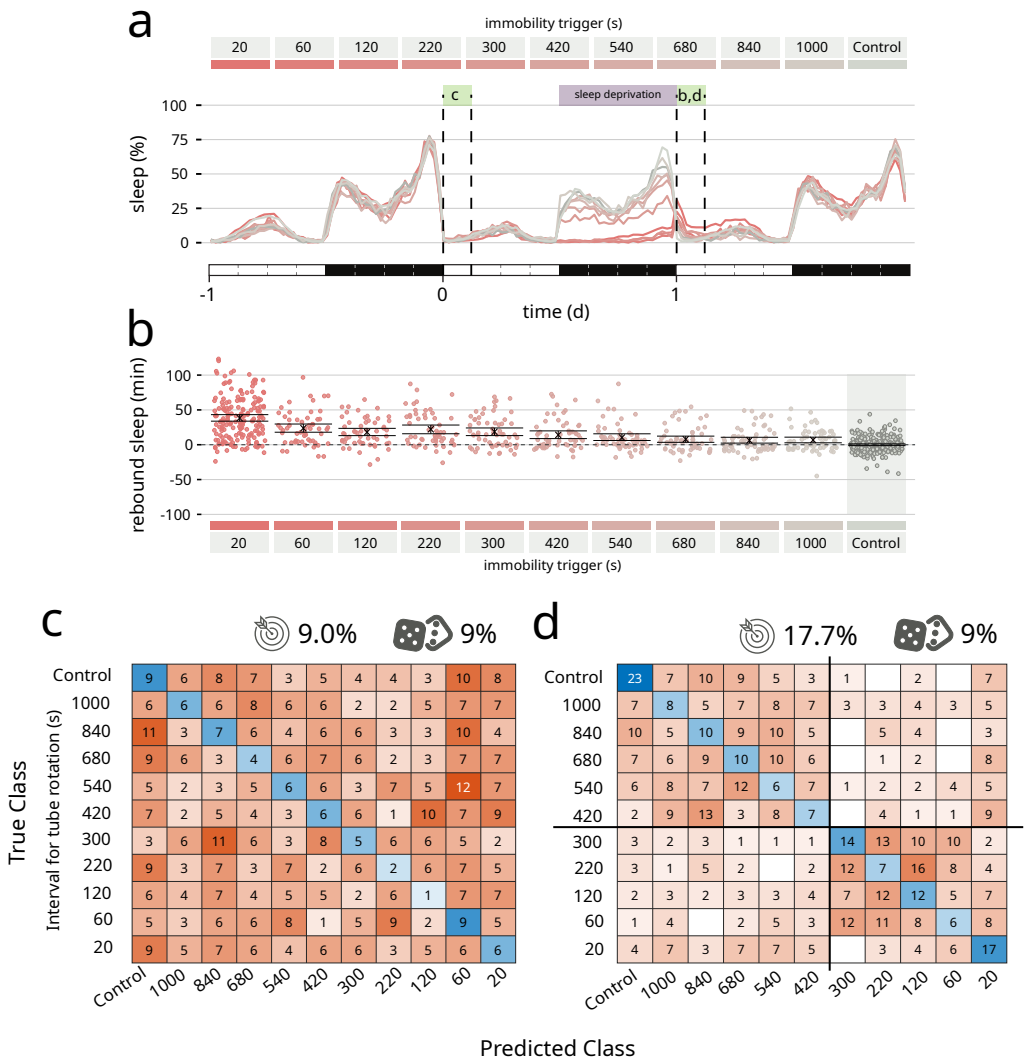
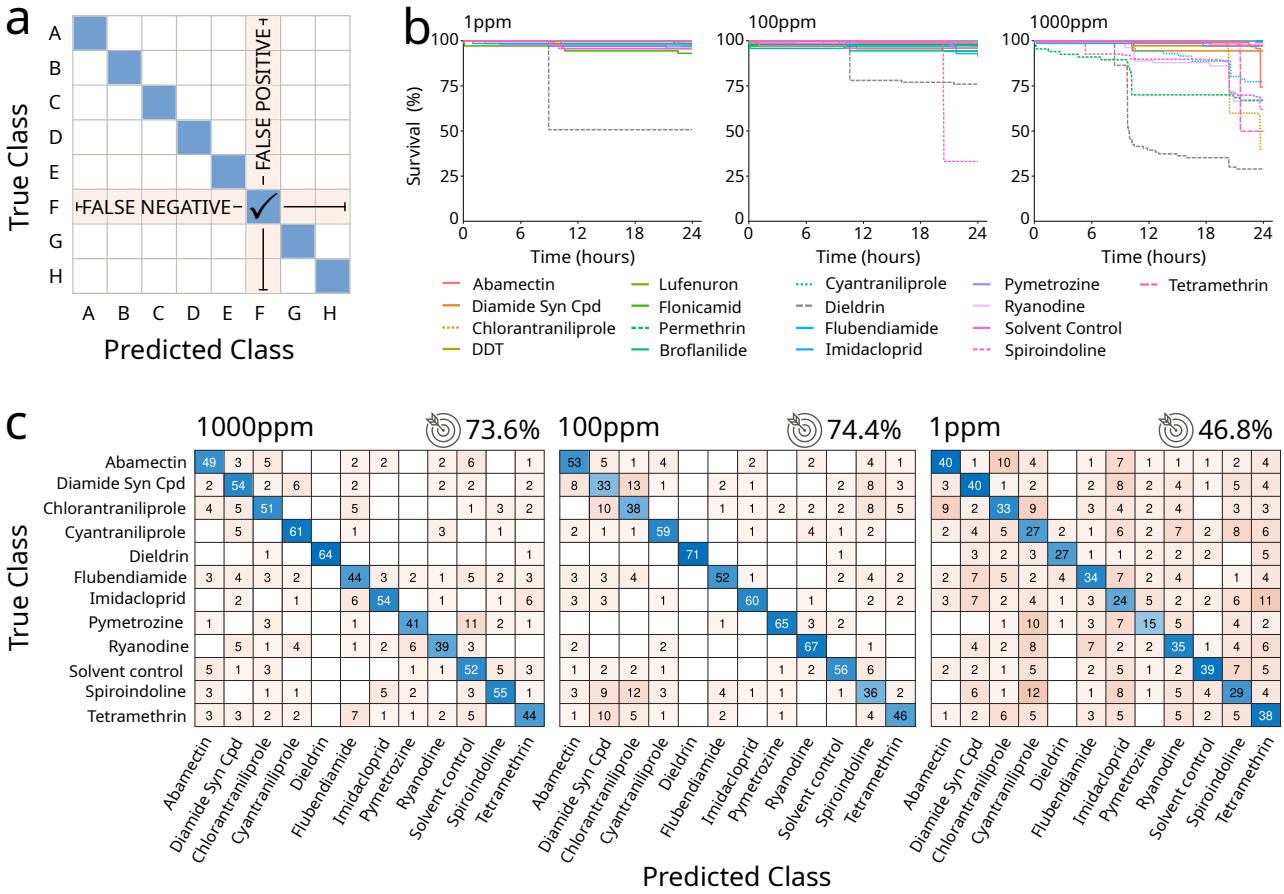


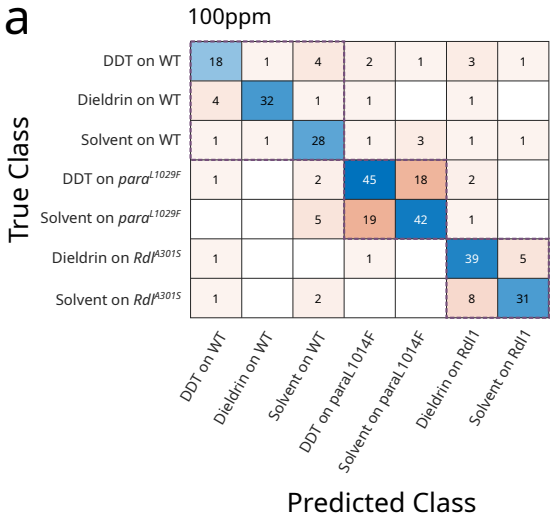
Figure 3

Jones et al 2022



Supp. Figure 1

Jones et al 2022



Supp. Figure 2

Jones et al 2022

Compound	Mode of action	Ref	Figure
Syngenta Cpd 3	Mitochondrial ATP synthesis		1D
DDT	Voltage-gated sodium channel activator	1	1C
Juglone	ROS	2,3	1D
Diflubenzuron	Chitin synthesis	4,5	1D
Rotenone	Complex I of the mitochondrial respiratory chain (MRC)	6	1D
Methotrexate	Dihydrofolate reductase inhibitor	7	1D
Diafenthiuron	Mitochondrial ATPase	8	1D
Fenoxycarb	Juvenile hormone mimic	9	1D
DCCD	Mitochondrial FIFO ATP synthase	10	1D
Lufenuron	Chitin synthesis inhibitor	11	1C
Chlorfenapyr	Mitochondrial oxidative phosphorylation	12	1D
Thiamethoxam	NACh-R13	13	1D
Tebfenozide	Ecdysone receptor	14	1D
Flonicamid	Sodium channels/TRP-V channels	15	1C
Syngenta Cpd 2	NACh-R16	16	1D
Quinuclidine Hycl.	Muscarinic Ach-R	17	1D
Indoxacarb	Voltage-gated sodium channels	18	1D
Fluazinam	Mitochondrial ATP synthesis	19,20	1D
Syngenta Cpd 5	Unknown		1D
Syngenta Cpd 7	Octopamine receptor		1D
Methoprene	Juvenile hormone mimic	21,22	1D
Permethrin	Voltage-gated sodium channels	23	1C
Tubulin	Unknown		1D
Metaflumizone	Voltage-gated sodium channels	24	1D
Syngenta Cpd 12	Unknown		1D
Cyenopyrafen	Succinate dehydrogenase	25,26	1D
Syngenta Cpd 1	GABAA receptors		1C
Syngenta Cpd 10	Octopamine receptor	27	1D
Abamectin	Glutamate-gated chloride channels	28,29	1C,1D,2
Diamide Syn Cpd	Ryanodine receptors		1C,1D,2
Chlorantraniliprole	Ryanodine receptors	30	1C,1D,2
Cyantraniliprole	Ryanodine receptors	31	1C,1D,2
Dieldrin	GABAA receptors	32	1C,1D,2

Flubendiamide	Ryanodine receptors	33	1C,1D,2
Imidacloprid	NACh-R	34,35	1C,1D,2
Pymetrozine	TRP-V channels	36	1C,1D,2
Ryanodine	Ryanodine receptors	37	1C,1D,2
Spiroindoline	Cholinergic transmission	38	1C,1D,2
Tetramethrin	Sodium channel modulator	39	1C,1D,2

References

1. O'Brien, R. D. & Matsumura, F. DDT: A New Hypothesis of Its Mode of Action. Science 146, 657–658 (1964).
2. Islam, A. K. M. M. & Widhalm, J. R. Agricultural Uses of Juglone: Opportunities and Challenges. Agronomy 10, 1500 (2020).
3. Maleita, C. et al. Juglone and 1,4-Naphthoquinone—Promising Nematicides for Sustainable Control of the Root Knot Nematode Meloidogyne luci. Front. Plant Sci. 13, 867803 (2022).
4. Cohen, E. & Casida, J. E. Inhibition of Tribolium gut chitin synthetase. Pestic. Biochem. Physiol. 13, 129–136 (1980).
5. Meola, S. M. & Mayer, R. T. Inhibition of Cellular Proliferation of Imaginal Epidermal Cells by Diflubenzuron in Pupae of the Stable Fly. Science 207, 985–987 (1980).
6. Palmer, G., Horgan, D. J., Tisdale, H., Singer, T. P. & Beinert, H. Studies on the Respiratory Chain-linked Reduced Nicotinamide Adenine Dinucleotide Dehydrogenase. J. Biol. Chem. 243, 844–847 (1968).
7. Stone, S. R., Montgomery, J. A. & Morrison, J. F. Inhibition of dihydrofolate reductase from bacterial and vertebrate sources by folate, aminopterin, methotrexate and their 5-deaza analogues. Biochem. Pharmacol. 33, 175–179 (1984).
8. Ruder, F. J. & Kayser, H. The Carbodiimide Product of Diafenthiuron Inhibits Mitochondria in Vivo. Pestic. Biochem. Physiol. 46, 96–106 (1993).
9. Grenier, S. & Grenier, A.-M. Fenoxycarb, a fairly new Insect Growth Regulato: review of its effects on insects. Ann. Appl. Biol. 122, 369–403 (1993).
10. Ruder, F. & Kayser, H. The thiourea insecticide diafenthiuron inhibits mitochondrial ATPase in vitro and in vivo by its carbodiimide product. Biochem. Soc. Trans. 22, 241–244 (1994).
11. Wilson, T. G. & Cryan, J. R. Lufenuron, a chitin-synthesis inhibitor, interrupts development ofDrosophila melanogaster. J. Exp. Zool. 278, 37–44 (1997).
12. Black, B. C., Hollingworth, R. M., Ahammadsahib, K. I., Kukel, C. D. & Donovan, S. Insecticidal Action and Mitochondrial Uncoupling Activity of AC-303,630 and Related Halogenated Pyrroles. Pestic. Biochem. Physiol. 50, 115–128 (1994).
13. Wiesner, P. & Kayser, H. Characterization of nicotinic acetylcholine receptors from the insectsAphis craccivora,Myzus persicae, andLocusta migratoria by radioligand binding assays: Relation to thiamethoxam action. J. Biochem. Mol. Toxicol. 14, 221–230 (2000).
14. Retnakaran, A. et al. Mode of action of the ecdysone agonist tebufenozide (RH-5992), and an exclusion mechanism to explain resistance to it. Pest Manag. Sci. 57, 951–957 (2001).
15. Taylor-Wells, J. et al. Toxicity, mode of action, and synergist potential of flonicamid against mosquitoes. Pestic. Biochem. Physiol. 151, 3–9 (2018).
16. Chambers, C. et al. Insecticidal spider toxins are high affinity positive allosteric modulators of the nicotinic acetylcholine receptor. FEBS Lett. 593, 1336–1350 (2019).

17. Dick, M. R., Dripps, J. E. & Orr, N. Muscarinic Agonists as Insecticides and Acaricides. Pestic. Sci. 49, 268–276 (1997).
18. Lapied, B., Grolleau, F. & Sattelle, D. B. Indoxacarb, an oxadiazine insecticide, blocks insect neuronal sodium channels: DCJW blocks neuronal Na + channels. Br. J. Pharmacol. 132, 587–595 (2001).
19. Lee, J. E. et al. Fluazinam targets mitochondrial complex I to induce reactive oxygen species-dependent cytotoxicity in SH-SY5Y cells. Neurochem. Int. 60, 773–781 (2012).
20. Guo, Z., Miyoshi, H., Komyoji, T., Haga, T. & Fujita, T. Uncoupling activity of a newly developed fungicide, fluazinam [3-chloro-N-(3-chloro-2,6-dinitro-4-trifluoromethylphenyl)-5-trifluoromethyl-2-pyridinamine]. Biochim. Biophys. Acta BBA - Bioenerg. 1056, 89–92 (1991).
21. Miura, K., Oda, M., Makita, S. & Chinzei, Y. Characterization of the Drosophila Methoprene -tolerant gene product: Juvenile hormone binding and ligand-dependent gene regulation. FEBS J. 272, 1169–1178 (2005).
22. Jindra, M., Palli, S. R. & Riddiford, L. M. The Juvenile Hormone Signaling Pathway in Insect Development. Annu. Rev. Entomol. 58, 181–204 (2013).
23. Vais, H., Williamson, M. S., Devonshire, A. L. & Usherwood, P. N. The molecular interactions of pyrethroid insecticides with insect and mammalian sodium channels. Pest Manag. Sci. 57, 877–888 (2001).
24. Salgado, V. L. & Hayashi, J. H. Metaflumizone is a novel sodium channel blocker insecticide. Vet. Parasitol. 150, 182–189 (2007).
25. Nakahira, K. Strategy for discovery of a novel miticide Cyenopyrafen which is one of electron transport chain inhibitors. J. Pestic. Sci. 36, 511–515 (2011).
26. Wei, P. et al. Downregulation of carboxylesterase contributes to cyflumetofen resistance in Tetranychus cinnabarinus (Boisduval): Downregulated esterase genes in cyflumetofen resistance in mites. Pest Manag. Sci. 75, 2166–2173 (2019).
27. O'Sullivan, A. C. et al. Synthesis and Insecticidal Activity of New Benzyl- and Indanyl-Oxazolines, Thiazolines and Alkoxy-Alkyl-Imidazolines. in ACS Symposium Series (eds. Maienfisch, P. & Stevenson, T. M.) vol. 1204 411–430 (American Chemical Society, 2015).
28. Cully, D. F., Paress, P. S., Liu, K. K., Schaeffer, J. M. & Arena, J. P. Identification of a Drosophila melanogaster Glutamate-gated Chloride Channel Sensitive to the Antiparasitic Agent Avermectin. J. Biol. Chem. 271, 20187–20191 (1996).
29. Arena, J. P., Liu, K. K., Paress, P. S., Schaeffer, J. M. & Cully, D. F. Expression of a glutamate-activated chloride current in Xenopus oocytes injected with Caenorhabditis elegans RNA: evidence for modulation by avermectin. Mol. Brain Res. 15, 339–348 (1992).
30. Cordova, D. et al. Anthranilic diamides: A new class of insecticides with a novel mode of action, ryanodine receptor activation. Pestic. Biochem. Physiol. 84, 196–214 (2006).
31. Selby, T. P. et al. Discovery of cyantraniliprole, a potent and selective anthranilic diamide ryanodine receptor activator with cross-spectrum insecticidal activity. Bioorg. Med. Chem. Lett. 23, 6341–6345 (2013).
32. ffrench-Constant, R. H., Rocheleau, T. A., Steichen, J. C. & Chalmers, A. E. A point mutation in a Drosophila GABA receptor confers insecticide resistance. Nature 363, 449–451 (1993).
33. Masaki, T. et al. Flubendiamide, a Novel Ca 2+ Channel Modulator, Reveals Evidence for Functional Cooperation between Ca 2+ Pumps and Ca 2+ Release. Mol. Pharmacol. 69, 1733–1739 (2006).
34. Bodereau-Dubois, B. et al. Transmembrane Potential Polarization, Calcium Influx, and Receptor Conformational State Modulate the Sensitivity of the Imidacloprid-Insensitive Neuronal Insect Nicotinic Acetylcholine Receptor to Neonicotinoid Insecticides. J. Pharmacol. Exp. Ther. 341, 326–339 (2012).

35. Perry, T. et al. Effects of mutations in Drosophila nicotinic acetylcholine receptor subunits on sensitivity to insecticides targeting nicotinic acetylcholine receptors. Pestic. Biochem. Physiol. 102, 56–60 (2012).
36. Nesterov, A. et al. TRP Channels in Insect Stretch Receptors as Insecticide Targets. Neuron 86, 665–671 (2015).
37. Fleischer, S., Ogunbunmi, E. M., Dixon, M. C. & Fleer, E. A. Localization of Ca2+ release channels with ryanodine in junctional terminal cisternae of sarcoplasmic reticulum of fast skeletal muscle. Proc. Natl. Acad. Sci. 82, 7256–7259 (1985).
38. Sluder, A. et al. Spiroindolines Identify the Vesicular Acetylcholine Transporter as a Novel Target for Insecticide Action. PLoS ONE 7, e34712 (2012).
39. Yamamoto, D., Quandt, F. N. & Narahashi, T. Modification of single sodium channels by the insecticide tetramethrin. Brain Res. 274, 344–349 (1983).

Jones et al. 2022 Supplementary methods

Generation of ParaL1029F via CRISPR-CAS9 based genome editing

CRISPR/Cas9-mediated genome editing was used to introduce a point mutation L1029F in para-PBG isoform, CTT to TTT, L to F by homology-dependent repair (HDR) using one guide RNA and a dsDNA plasmid donor. The strategy design, molecular biology and screening was completed by Well Genetics Inc., Taiwan (R.O.C.). The cassette PBacDsRed contains Piggy Bac 3' terminal repeats, the selection marker 3xP3-DsRed, and Piggy Bac 5' terminal repeats. The selection marker 3xP3-DsRed contains Piggy Bac 3' terminal repeats, 3x Pax3 and hsp70 promoter, DsRed2, SV40 3'UTR, and Piggy Bac 5' terminal repeats. The dsRed marker facilitates the genetic screening and was excised by Piggy Bac transposase. Only one TTAA motif was left after transposition embedded in mutated intron sequence, and create a mutation G to A on X:16,486,649; X:16,486,649~ X:16,486,646, CTAA to T TAA in intron. The CRISPR Target Site [PAM]: CACAAGATTGCCGATGACAA[CGG]. Guide RNA Primers: Sense oligo5'-CTTCGCACAAGATTGCCGATGACAA and Antisense oligo5' -AAACTTGTCATCGGCAATCTTGTGC. Upstream Homology Arm: 1,027bp,the +34,097nt to +35,123 nt from ATG of para. Forward Oligo5'-GTTCACCAAACTCGGAATCG; Reverse Oligo5'-GTGGCCAAGAAGAAGGGAAT. Downstream Homology Arm: 1,022 bp, the +35,128nt to +3,6149 nt from ATG of para Forward Oligo: 5'-CCATGGCTTTAAGCATCGCA; Reverse Oligo: 5'-TTATGACGGATACGGTTACGG. Synthesis fragment: 5'–GGTTGTCATCGGCAATTTTGTGgtgagtactcttatcgaactgctgacttgtaaacgatgtttactggctataatgctgacttatcgcct

The *Drosophila* injection strain was *white*[1118]. 206 embryos were injected. 36 G0 crosses were established. Of the 78 positive lines crosses, in the F1 screen 25 positive lines were identified. Seven lines were positively validated by PCR and 1 line was sequenced confirming no unexpected changes in *para*. Lines were isogenized and balanced. DsRed was excised using PiggyBac (PBac) Transposase Bloomington Stock RRID:BDSC_8285. Excision was validated by genomic PCR and sequencing. Resulting lines were hemizygous viable. The line used in study had internal identifier: 20256ex1.