Published on Dec 14, 2024 by Arcadia Science **DOI**: 10.57844/arcadia-48b0-607a

A data-driven approach to match organisms and research problems

What if we could select research organisms that are far more relevant to human biology or more likely to unearth biological solutions not found in humans? With more sequence data, structural prediction, and phylogenetic comparative methods, a richer framework is possible.

Contributors (A-Z)

Prachee Avasthi, Audrey Bell, Megan L. Hochstrasser, Ryan York

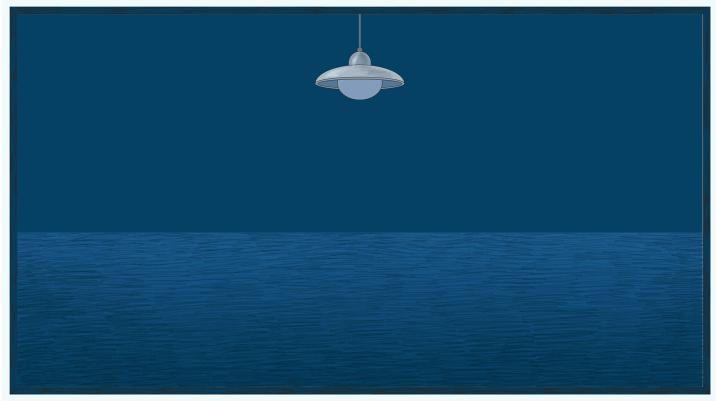
Version 1 · Mar 31, 2025

Purpose

It's critical to select the ideal organismal model to use for studying a human disease or biological process faster, cheaper, and easier than can be explored in humans. Scientists often select organisms based on historical precedent, ease of use in the lab, and similarity of genes or phenotypes. While this approach has resulted in many important advancements and certainly has its merits, relying on intuition, convention, and prior studies to select model organisms isn't always optimal for understanding the complexities of human biology, particularly in the context of therapeutic development. Discovery research and preclinical testing in animal models often fail to translate to

the clinic **[1]** and don't take the evolutionary history of mice and humans into account **[2]**.

In this pub, we describe a new framework for thinking about organismal model selection that leverages the vastness of biology, including and beyond traditional model systems. This approach has the potential to accelerate the pace of biological discovery by highlighting valuable organisms that have been historically overlooked and understudied but have outsized biological relevance to humans.



We tend to rely on a single set of model organisms, looking "under the lamppost" at the biology we know. What if we shone a light across the whole tree of life? Could we find better models?

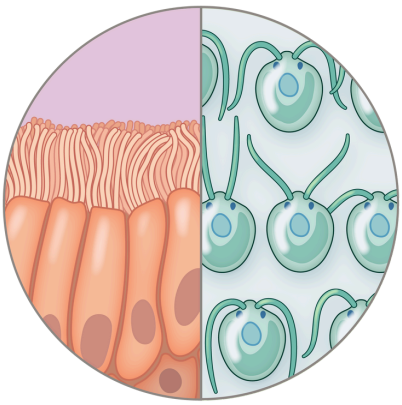
This pub is meant for a scientific audience and we'd love feedback. Would our organismal selection framework change how you'd select which organism you'd use to solve your research problem of interest? Would you use these tools to identify new research directions based on where your organismal expertise is best leveraged?

- This pub is part of the **platform effort**, "Genetics: Decoding evolutionary drivers across biology." Visit the platform narrative for more background and context.
- Read our **companion pub**, "Leveraging evolution to identify novel organismal models of human biology" **[3]**, for more details on the science underlying our organismal selection pipeline.
- For an **example of this approach in action**, check out "Rescuing *Chlamydomonas* motility in mutants modeling spermatogenic failure" **[4]**.

Traditional organism selection

There are many reasons why traditional model organism selection is suboptimal when pursuing *biological conservation,* the context most relevant to humans. Traditionally, this is done by comparing gene or protein sequences between the organism and humans and considering whether the two share relevant phenotypes. Historically, identifying the right system with conserved biology has required deep knowledge of individual organisms and the contribution of an entire field to unearth the dimensions of shared biological context.

Leveraging intuition about commonalities to unearth shared principles



Modeling human cilia with *Chlamydomonas* flagella

Imagine we wanted to study the movement of mucus in our airway in respiratory diseases, movement of cerebrospinal fluid in the brain in developmental disorders, or movement of eggs to the uterus in female infertility. Cilia, the finger-like protrusions in cells lining the trachea, brain ventricles, and fallopian tubes responsible for this movement are nearly identical to the flagellar structures that single-celled green algae use to swim. Both the individual proteins and the coordinated processes needed to generate force from these protrusions are conserved in algae and provide a low-cost, simple, and less invasive way to study these mechanisms to improve a range of complex diseases.

Rather than relying on intuition or luck, we wondered if it was possible to more systematically identify properties of organisms across the tree of life that might be redeployed or re-engineered to develop human therapeutics and other useful innovations. Not only might we be able to accelerate the work many organismal biologists have contributed to mechanistic understanding, but we may also be able to improve the accuracy of organismal selection for downstream application.

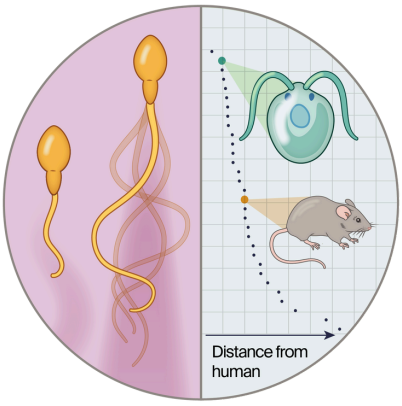
For example, around 90% of drugs that progress from preclinical testing in organismal models (95% is done in rodents) to clinical trials in humans fail. This failure rate suggests that many researchers are using convention or historical precedent and not fully leveraging available data to optimize the organism they select for their research questions. We asked whether we could use a more rigorous data-driven framework for discovery research to increase the accuracy of insights with respect to human relevance.

Rationally sourcing biological conservation

Beyond proteins and prior mechanistic studies, we've never been in a better position to leverage even more data. We can use protein structural properties inferred from amino acid sequence and take into account evolutionary history to do comparisons between species **[3]**. Sometimes we find that our intuition about model systems was spot-on, but we can be much more confident in our choices and reach conclusions quicker.

Leveraging data to speed up model selection

Spermatogenic failure is a severe form of male infertility with certain subtypes attributable to mutations in the *SPEF2* and *DNALI1* genes.



Modeling human sperm motility with *Chlamydomonas*

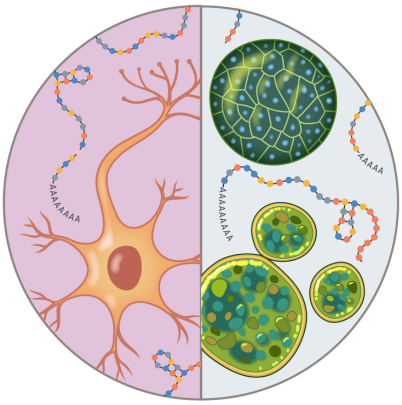
Using the data from our organism selection pipeline, we landed on the green alga *Chlamydomonas reinhardtii* as an appropriate model for spermatogenic failure. We identified motility defects in algal cells with mutations in the appropriate genes as well as rescue motility **[4]** with compounds previously found to increase sperm motility **[5]**.

While scientists have long been using *Chlamydomonas* to understand sperm motility due to structural similarities between *Chlamydomonas* and sperm flagella **[6][7][8]** due

to high-resolution electron microscopy, we were able to use our framework to identify an appropriate model and validate its relevance to human biology quickly, cheaply, and with high confidence using little additional context. In this case, the pipeline led us to an existing model, but we got there through an unbiased selection process.

The power of this data-driven approach is more readily appreciated when the results of our analyses lead to unintuitive results, identifying organisms with non-obvious similarities to human biology.

Leveraging data to find unexpected models with human relevance



Modeling human neuronal mRNA processing in unicellular organisms

Imagine we want to develop a treatment for spinal muscular atrophy (SMA), a neuromuscular disease caused by mutations in SMN1, a protein involved in RNA processing that's critical for motor neuron function and survival. Let's say we're trying to decide which research organism to use upstream of pharmacology and toxicity assessments to unearth relevant biological assays and mechanisms of action for therapeutic assets.

If we use standard model organism selection, we'd likely start by considering a mouse model or another well-established organism. In other words, when studying a neuromuscular disease, you might assume that the right organism to study this in has neurons and muscles. However, our analysis based on multiple physical and chemical protein properties beyond primary sequence suggests that unicellular *Sphaeroforma arctica* and *Chlorella vulgaris* have a more conserved biological context relative to other species and are well-suited to tackle SMA.

Well-established models like mice aren't just expensive to maintain — they also don't necessarily recapitulate the specifics of the human disease. The conservation of relevant properties in a much simpler system may signal that the etiology of the disease is in a more ancient and conserved biological process that makes muscles

and nerves particularly vulnerable. And that more complex tissue-level phenotypes may be a consequence rather than a cause of the disease.

Our strategy lets us rationally and agnostically consider less-studied organisms that may be more biologically relevant to the disease or trait in question.

A call for change

We've developed an approach that allows scientists to rationally identify research organisms for modeling human traits by incorporating genomic data, protein structure, and other biological contexts **[3]**. Knowing that not all researchers can dynamically spin up new infrastructure for every new research organism they land on, the other major utility of our framework is that for a fly or fish or worm lab, we can help agnostically identify the focus areas where these species are most relevant and can make the most headway. We hope this data-driven approach will increase our ability to leverage the full diversity of the natural world for scientific discovery.

Weigh in!

Would you use our workflow to identify an appropriate research organism, a biological area the model you have expertise in can best tackle, or use these data to support your choices when seeking funds, in publications, or for drug development? This platform relies on access to high-quality, annotated genomes across a wide range of organisms. What species for which you already have expertise or tools would you like to be integrated into our platform?

References

1 Hartung T. (2024). The (misleading) role of animal models in drug development. https://doi.org/10.3389/fddsv.2024.1355044

2 Perlman RL. (2016). Mouse Models of Human Disease: An Evolutionary Perspective. https://doi.org/10.1093/emph/eow014

3 Avasthi P, McGeever E, Patton AH, York R. (2024). Leveraging evolution to identify novel organismal models of human biology. https://doi.org/10.57844/ARCADIA-33B4-4DC5

4 Essock-Burns T, Lane R, MacQuarrie CD, Mets DG. (2024). Rescuing Chlamydomonas motility in mutants modeling spermatogenic failure. https://doi.org/10.57844/ARCADIA-FE2A-711E

5 Gruber FS, Johnston ZC, Norcross NR, Georgiou I, Wilson C, Read KD, Gilbert IH, Swedlow JR, Martins da Silva S, Barratt CLR. (2022). Compounds enhancing human sperm motility identified using a high-throughput phenotypic screening platform. https://doi.org/10.1093/humrep/deac007

6 Inaba K. (2011). Sperm flagella: comparative and phylogenetic perspectives of protein components. https://doi.org/10.1093/molehr/gar034

7 Leung MR, Zeng J, Wang X, Roelofs MC, Huang W, Zenezini Chiozzi R, Hevler JF, Heck AJR, Dutcher SK, Brown A, Zhang R, Zeev-Ben-Mordehai T. (2023). Structural specializations of the sperm tail. https://doi.org/10.1016/j.cell.2023.05.026

8 Neilson LI, Schneider PA, Van Deerlin PG, Kiriakidou M, Driscoll DA, Pellegrini MC, Millinder S, Yamamoto KK, French CK, Strauss JF III. (1999). cDNA Cloning and Characterization of a Human Sperm Antigen (SPAG6) with Homology to the Product of the Chlamydomonas PF16 Locus. https://doi.org/10.1006/geno.1999.5914