

Predicting Evolution

v0.8 - 24.09.2026

Yannick Wurm, Programme Director

Simon Evans, Science + Technology Lead - Accelerated Adaptation

Richard Nichols, Science Contractor

CONTEXT

This working document presents an early draft of the core ideas underpinning a programme that is currently being explored at ARIA. We are sharing an early formulation and invite you to provide feedback to help us refine our thinking. This is not in itself a funding opportunity, but may end up leading to one.

An ARIA programme seeks to unlock a scientific or technical capability that

- + changes the perception of what's possible or valuable
- + has the potential to catalyse massive social and economic returns
- + is unlikely to be achieved without ARIA's intervention.

[Share your feedback and register your interest in a workshop](#)

EXECUTIVE SUMMARY

Nothing in biology makes sense except in the light of evolution.

Theodosius Dobzhansky, 1973

Dobzhansky was right, but with one important caveat: evolution makes sense of biology only in hindsight. Being able to anticipate evolutionary outcomes could transform industries: human pathogens evolving resistance, tumours escaping treatment, crop breeding failing to keep pace with climate shifts, gene therapies or engineered organisms drifting from their designed functions.

Biological evolution is the change, over successive generations, in the sequence of DNA bases—the string of four letters, A, T, G and C, that makes up the genome—within a population. Mutation and gene flow introduce new variants; selection favours those that enhance survival or reproduction; sexual reproduction swaps extended sections; and drift shifts their frequencies by chance alone. Each is, at root, a pattern-generating or -destroying process acting on a four-character alphabet across millions of positions in every individual. Yet we're not merely seeking tools that detect the genetic signal in the noise of vast datasets: we must identify the signals that will propagate into the future, rather than those that only explain the past.

We believe three technological developments can combine to transform evolutionary biology from a retrospective to a predictive science:

- + Robotics can automate experiments at a scale never previously attempted, maintaining thousands of replicate, evolving populations in parallel under precisely controlled conditions—turning experimental evolution into something closer to a production line.
- + DNA and RNA sequencing is cheap enough to connect genotypic to phenotype across tens of thousands of populations—generating the base-by-base record of change that any predictive model must be trained and tested against.
- + Modern machine learning has proven adept at detecting structure in huge, highly complex, high-dimensionality datasets—exactly what is needed to predict genome-scale change in DNA sequence.

Together, these developments can take evolutionary biology past the threshold that weather forecasting crossed 50-100 years ago: from a discipline that explains why things happened to one that specifies, with quantified uncertainty, what happens next.

No dataset yet exists at the scale needed to train and test predictive models so we will create one: massively replicated experimental evolution in a nematode worm, an animal with genetic architecture broadly representative of animals' and plants'. We will then test predictions through open, blind competitions: teams will submit their forecasts of which populations persist in the face of thermal and chemical stress, and what genetic changes underlie those outcomes.

- + These stressors exert demographic pressure by increasing mortality rate, reducing the likelihood of population persistence.
- + This elevated mortality acts as a selective agent, driving adaptive evolution if survival is attributable to genetic variants.

Why this programme?

Evolution is not a historical curiosity. It is happening all around us: in every pathogen evolving resistance to a drug, every pest escaping biocidal control, every wild population struggling to deal with climate change. In each case, the capability gap between anticipating that change and only documenting it afterwards is what separates a successful intervention from one that comes too late.

Unfortunately, evolutionary biology remains overwhelmingly an explanatory science, rather than a predictive one. It can account for why a population evolved as it did after the fact, often in exquisite detail—but breeding decisions, drug approvals, and biocontrol releases all have to be made before the evolutionary outcome is known. An explanatory science, however good its explanations, cannot inform a decision that has to be made in advance of the evidence.

Converting that explanatory power into a forecast—one that specifies what a given population will do next, at what rate, and with what degree of confidence—is something the field has rarely attempted, and never at the scale these decisions demand. This programme exists to find out whether that is possible, and if so, place a working version into the hands of the people who make these calls.

What we hope to accomplish

All useful forecasts specify a range of outcomes, their likelihood, and their timing. Weather forecasting met this standard not by building a single model of indefinite precision, but a hierarchy of models, each suited to a different time horizon, each honest about where its skill runs out. A forecast is useful not because it predicts everything, but because it states accurately what can be predicted, and for how long.

As with weather forecasting, success is not represented by a single, successful prediction. Rather, it is the ability to deliver decision-relevant forecasts¹—including accurate representation of their temporal sensitivity—that can inform both near-term interventions (e.g., selective breeding or which drug to administer) and longer-term strategic decisions (e.g., minimising risks of invasive species naturalising, agricultural pests escaping control, engineered living systems to go haywire, or cancers evolving resistance to treatment).

Achieving that means confronting a basic problem: we do not yet know whether evolutionary trajectories are predictable in principle, only that they have rarely been predicted in practice.

Overview:

- + Evolutionary trajectories currently look disordered: unpredictable, historically contingent yet also seemingly one-off.
- + Our bet: this disorder is not pure noise. Predictable structure exists beneath it, and forecasting evolution means finding that structure.
- + We will reveal it through:
 - massive-scale replicated evolutionary experiments...
 - ...tracked with comprehensive genomic profiling...
 - ...analysed by cutting-edge modelling approaches new to biology.
- + Our funding concentrates on building an unprecedented dataset that only an ARIA-scale programme could create. Its benefits are diffuse and shared across fields and industries, while its costs (wet labs, robotics, sequencing at scale) are concentrated upfront. A classic market failure.
- + Forecasting needs no large-scale subsidy. With open data competitors cover their own costs, while the models, talent, and visibility are theirs to keep. Furthermore, a prize competition keeps diverse teams searching until one does win (e.g., the DARPA driverless vehicle and the CASP protein folding challenges were won by teams from outside the founding fields). Small development awards will widen participation to strong UK-based teams short on compute or engineering.

Programme North star: Establish evolutionary forecasting as a quantitative discipline: predicting whether, when and how any population adapts (or collapses) with the reliability needed to steer interventions in living systems.

Programme Goals

#	Goal	Section of North Star	Outcomes	Success Metrics
1	Discipline <i>Establish a discipline, not just a result</i>	"Establish evolutionary forecasting as a quantitative discipline"	The experimental testbed, scoring platform and open dataset: analysis-ready releases of ~2,100 evolutionary trajectories, molecular-state and blinded test outcomes.	Participation spans biological, mathematical and machine-learning communities; external teams generate models, software, publications or teaching resources; the data are reused beyond the Programme; and at least one successor challenge is in development.
2	Forecastability <i>Prove evolution is forecastable</i>	"Predicting whether, when and how a population under emerging stress adapts or collapses"	Challenge 1: six intercrossed lines under heat and H ₂ O ₂ stress. Forecasts are scored against observed outcomes and baseline models.	At least one team demonstrates a reproducible success in forecasting the unseen outcomes in Challenge 1, that outperforms the relevant reference model based on random genetic drift and simple additive genetics.
3	Generalisability <i>Show the rules are generalisable</i>	"Any population"	Challenge 2 tests forecasts for 12 new wild-derived populations; Challenge 3 extends to spatial structure and interaction between stressors.	Stronger success: forecasting advantage over the published reference models is retained. North Star success: useful forecasting extends to range expansion and the spatial distribution of genetic variation.
4	Horizon <i>Establish the forecast horizon for genetic evolution</i>	"Precisely enough to inform the decisions that depend on it"	Performance is measured as a function of lead time through staged Challenge 3 competitions at t20, t40 and t60. Scores, calibration and uncertainty are tracked.	A reproducible map showing where forecasting skill holds and where its horizon lies.

Why it's worth shooting for

Evolutionary forecasts could inform decisions across timescales—from near-term interventions such as seasonal vaccine design and personalised medicine, to the longer-term strategic risks set out below.

Approaching evolution as a forecasting problem is also a distinct challenge for machine learning, not only for biology. Existing biological foundation models (e.g., AlphaFold², ESM³, Evo⁴) learn structural and functional regularities from a static snapshot of sequence or structure. Forecasting evolution requires something categorically different: models that learn how those regularities change over time under selection, and that generalise not by memorising one population's history but by capturing rules that hold for any population's future, given its genetic background and the pressures acting on it. Testing whether such models are possible—using data at a scale no existing dataset provides—is as much a contribution to machine learning as it is to evolutionary biology.

Forecasting evolution will allow us to make proactive interventions in medicine, agriculture and biosecurity by predicting when, how and how fast:

- + Tumours escape the treatments we prescribe.
- + Gene-edited cells delivered as therapy can fail to establish, or evolve beyond our control.
- + Human pathogens evolve escape routes from vaccines and antibiotics.
- + Biocontrol agents, released to control an invasive species, evolve a capacity to attack non-target organisms.
- + Evolutionary regression occurs in genetically engineered crops.
- + Agricultural pests evolve pesticide resistance.
- + Species adapt to environmental changes in situ, or go locally extinct

Why it's differentiated

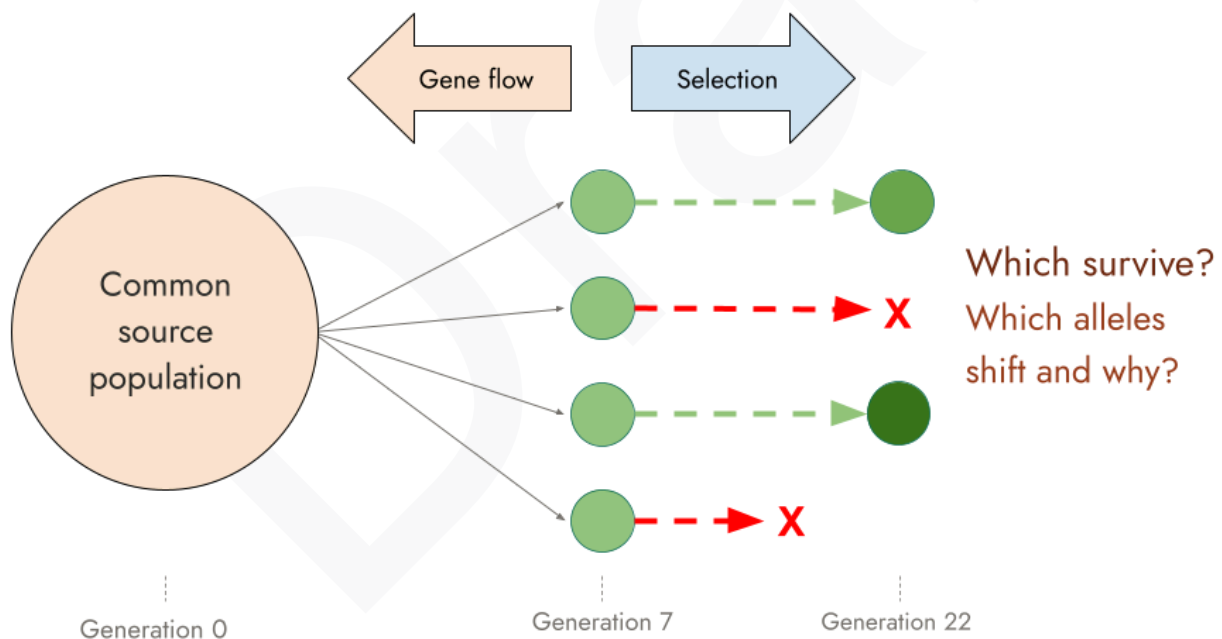
Feature	Community challenges CASP, CAFA, DREAM	Targeted evolution programmes ARPA-H ADAPT	Targeted evolution programmes DARPA Prophecy	Long-Term Evolution Experiments ⁵	Drosophila selection experiments ^{6,7}	Predicting Evolution
Blinded prediction against withheld truth	✓	✗	✗	✗	✗	✓
Creates biological ground truth	~	✓	✓	✓	✓	✓
Sets forecasting difficulty	~	~	✓	~	~	✓
Replicates sufficiently to score forecast uncertainty	✗	✗	✗	✗	✗	✓
Forecasts outcomes before they occur	✗	✓	✓	✗	✗	✓
Repeated genomic, molecular and phenotypic measurements	~	✓	~	~	~	✓
Tests transfer across genetic backgrounds, gene flow and space	✗	✗	~	✗	✗	✓
Forecasts population trajectories and fate	✗	✗	✗	✗	✗	✓
Uses empirical outcomes (i.e., not simulations)	~	✓	✓	✓	✓	✓
Open to diverse disciplines and methods	✓	~	~	✗	✗	✓
High complexity organism	✗	~	✗	✗	✗	✓
Realistic genetic diversity & genetic structure	✗	✓	~	✗	✗	✓
"Realistic" population size	✗	✓	~	✗	✗	✓

✓ fully met; ~ partially met; ✗ not met.

A key feature of our design is that we keep a check on the stochastic component of evolution while the experiment plays out. Left alone, populations under selection can drift apart unpredictably—two populations that start identical can end up nothing alike. So each of our populations stays connected, throughout the experiment, to the same starting population. Replication shows us the range of directions populations can take; the ongoing connection lets us measure how far each one pulls away, in which direction, and whether it holds that course—or collapses.

That connection is gene flow: each generation, we replace part of every evolving population with individuals from the common source population. This does two things:

- + It mirrors real-world problems, where selection acts locally but new individuals keep arriving from elsewhere, carrying genes suited to different conditions^{8,9}.
- + It makes the system easier to study, because it reduces the role of pure chance—leaving a clearer signal of what selection itself is doing.



What we expect to fund

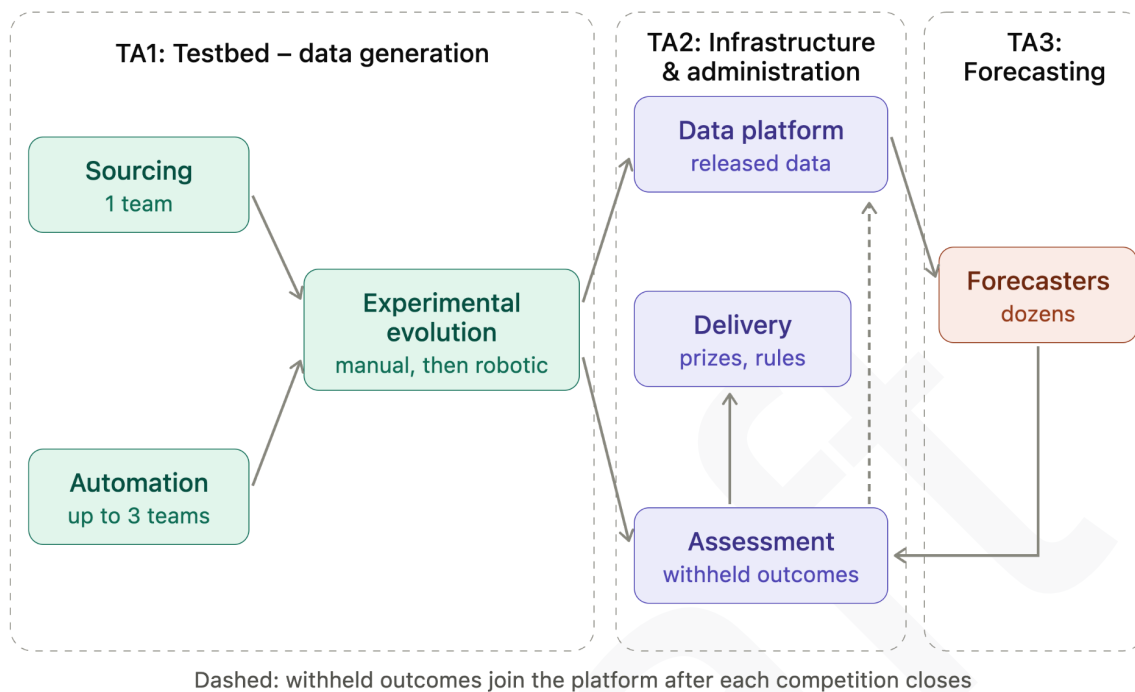
Turning evolutionary forecasting into a testable discipline requires:

- + Running evolution experiments at scale and measuring their outcomes.
- + Teams blindly forecasting those outcomes, who work from the same well-characterised evidence and testing conditions.
- + Competition machinery that controls what forecasters see and when, holding the withheld outcomes, and scores submissions against them.

Sandboxing

Tasks are organisationally isolated: no organisation may work on more than one task, such that forecasting and its assessment are fully isolated from data generation (and each other), preserving the competition's integrity.

We will test this through experimental evolution at unprecedented scale — increasing across the three challenges to more than 2,000 evolving populations of *Caenorhabditis remanei* populations under selection, each population characterised by genome, expression and chromatin measurement — released openly in a series of blinded prize competitions of escalating ambition. The challenge is deliberately application-agnostic, and designed to be accessible to teams coming from outside the life sciences.



Technical Area 1 Testbed - data generation

The testbed: the experimental platform hosting multi-parallel evolution. These teams hand off to each other every generation.

Nematode sourcing (1 team)

- + Collects, expands and maintains the *C. remanei* source pool used to initiate experimental populations.
- + Provides those populations to the experimental evolution management teams.
- + Designs and tests the selective challenges that will force adaptive evolution.
- + Proof-tests archiving protocols that will allow populations to be stored indefinitely and then revived, if needed.

Automation and robotics (up to three teams for the first 12 months)

- + Deliver the integrated systems Challenges 2 + 3 depend on: transfer, environmental control, stress application, imaging, counting, sampling and spatially structured gene flow.
- + Challenge 2 doubles the experimental scale of Challenge 1, and Challenge 3 is another quadrupling; automation makes that practicable.

- + We are particularly interested in adaptable technologies that can be folded into the platform as they mature, rather than automation that only works within a closed system.

Experimental evolution (likely different teams for Challenge 1 and 2+3)

- + Run Challenge 1 manually, with trained technicians executing established protocols: manage gene flow, apply the selective challenges, record growth and collapse, and sample at defined generations.
- + Productionise the reproducible selective challenges.
- + Commission sequencing and molecular measurement externally, with quality control rigorous enough that apparent evolutionary signals can be told apart from technical variation. Generated data include reference genome assembly, deep pooled-genome sequencing, gene expression, small-RNA sequencing and regulatory state.
- + Delivers data to the data infrastructure team (TA2) in the appropriate formats, and withheld outcomes to the independent assessors.
- + Run Challenges 2+3 similarly to Challenge 1, but introducing robotics and automation methods.

Technical Area 2: Programme infrastructure & administration

Data infrastructure, standards and visualisation (1 team)

This role sits at the centre of the programme. It handles released data only, received from the experimental evolution teams (TA1). After each competition closes, it adds the withheld outcomes so that later challenges can build on earlier results.

- + Turn experimental output into a forecasting resource: analysis-ready releases, common metadata and provenance standards, versioned datasets, secure storage and reliable distribution of large files.
- + Provides:
 - + reference baselines that forecasts must beat;
 - + submission interfaces;
 - + visualisations that make trajectories, genetic change, uncertainty and results legible to the community of potential stakeholders.

Competition administrator (1 team)

- + Ensure the forecasting challenges are credible, accessible and fair.

- + Manage rules, deadlines, participant support, data releases and publication of results.
- + Adjudicate the competition (including managing leaderboards) and distribute prizes. They will also manage the leaderboards and ranking systems.

Independent assessor (1 team)

- + Independent technical assessor and a constituted judging team retain the withheld outcomes, received directly from the experimental evolution team (TA1), and score submissions against them.
- + Verify scoring, manage conflicts of interest and adjudicate unexpected cases.
- + Under NDA and in a manner that respects competitors' confidential IP, verify protocols, models, and code provided.

Technical Area 3: Forecasting

(Entry numbers are uncapped but ARIA will have a call to support a portfolio of funded efforts, only some of which receive development funds)

- + Predicting genetic composition of populations, ideally with calibrated uncertainty.

We welcome academic groups, startups and SMEs, frontier research organisations, research institutes, national facilities, established companies, independent researchers, and consortia of these. We are looking for teams that pair deep expertise with a willingness to work across disciplinary boundaries, share methods and standards, and adapt as the experiments reveal where evolutionary predictability holds and where it breaks.

A small academic team, startup or independent researcher may have a strong forecasting idea without the compute, software engineering or staff to develop it at a scale that allows them to be competitive. We therefore expect to offer a limited number of development awards ahead of the competitions, covering:

- + Compute
- + Temporary research or data-engineering staff
- + Access to other specialist facilities, or the work of turning an early method into a reproducible system.
- + Technical advice and help finding collaborators may also be available.

Importantly, **the competitions are open.**

- + Subject to final eligibility, safety and data-access rules, the challenges will remain open to unfunded teams that produce valid submissions.
- + Development support exists to widen the range of serious approaches tested, not to predetermine which can win.
- + Openness extends to the forecasts themselves: at the conclusion of the competition, all participants must share their models. Prizes will not be awarded unless and until this has been verified.

Challenge structure

Details of the challenge design and the scoring system are given in the Appendix.
All forecasts fail but for three principal reasons.

Challenges			
Challenge	Design	Forecasting innovation	Rationale
1	Variation from a single wild population, assorted to create diverse genetic starting points exposed to the same selective challenges.	Using early generation patterns to forecast outcomes for populations with genetic variants in different genetic backgrounds.	Attracting teams and engaging them with this rich source of data to train models that can exploit the combination of genomic and latent state data, and broader data from across the tree of life on eukaryote biological systems.
2	Variation from multiple wild populations used as novel starting points for the same selective challenges	Using insights from (1) to forecast outcomes for populations with multiple new variants which have co-evolved in different combinations.	Raising the demands of the challenge so that forecasting success progressively depends on capturing the deeper dynamics governing evolutionary trajectories, rather than extrapolating short-term trends.
3	Expose the same populations, each in a 2D matrix of interconnected sites in which the intensity of selective challenge increases between adjacent sites.	Using insights from (1) and (2) to predict how each population will colonise this landscape, and how the genetic variants will distribute across the landscape.	Moving the challenge towards the laboratory exit door: successful forecasting must now confront the processes that shape species' range limits and more generally spatially varying selection acting in the face of gene flow

Even if forecasting ultimately fails beyond a particular time horizon or under particular conditions, a reproducible map of where prediction holds, how rapidly it degrades and what information extends its reach would be a field-defining outcome.

What we are still trying to figure out

- + What amounts of cash should we allocate to small, innovative forecasting teams we want to support?
- + Who else might complement our efforts to make this prediction competition successful?
 - + Might cloud computing providers/hyperscalers host datasets and/or contribute compute credits? This could help attract users to their platform.
 - + Might an organisation want to contribute to prize pots or give stipends to creative teams with high potential? Supporting this initiative could complement internal R&D to make the field advance faster.
- + Which data formats should datasets be shared in? Current thinking is that we provide some or all of:
 - + Reference genome: FASTA.gz + .fai
 - + Raw Illumina reads: FASTQ.gz
 - + Read alignments: CRAM + .crai
 - + Variants / pooled allele frequencies: VCF.gz + .tbi or .csi
 - + Metagenome contigs: FASTA.gz
 - + Metagenome assembly graph: GFA.gz
 - + Metagenome taxonomic / functional profiles: BIOM or TSV
 - + mRNA- and small-RNA abundance: TSV or Parquet
 - + ATAC-seq peaks: narrowPeak.gz + .tbi
 - + ATAC-seq signal tracks: bigWig
 - + Single-worm phased genomes: FASTQ.gz, phased VCF.gz + index, and haplotype-resolved FASTA.gz
 - + Phenotyping videos + some level of CSV summary
- + Teams will be evaluated based on differences between predictions and realised outcomes. The relative rankings of teams will be shared among competition participants. However, to help prevent cheating, teams should also deposit the

materials from which the predictions were generated (protocol, model code, links to training data) to an independent verification team under NDA. The verification team should audit or verify competitors' submissions in a manner with no unnecessary disclosure, in order to preserve competitors' IP and commercial confidentiality.

- + Should teams be allowed to use datasets or models that are not fully publicly available to train their model? To focus innovation on modeling approaches, we suspect that a rule similar to "using a widely accessible closed-source dataset or tool" is ok (e.g., broadly available closed-source closed-weight LLMs). But that "using a dataset my company has exclusive access to" should not be.
- + Are there other things we should consider to maximise the likelihood of the programme's success or impact?

SOURCES

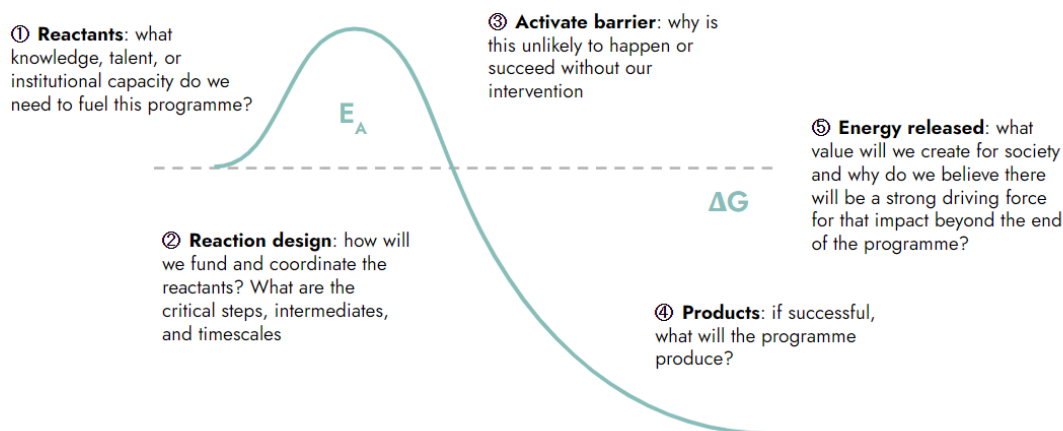
References cited in this document.

1. Gneiting, T. & Raftery, A. E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **102**, 359–378 (2007).
2. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
3. Hayes, T. *et al.* Simulating 500 million years of evolution with a language model. *Science* **387**, 850–858 (2025).
4. Brix, G. *et al.* Genome modelling and design across all domains of life with Evo 2. *Nature* **652**, 1349–1361 (2026).
5. Whiting, J. R. *et al.* The genetic architecture of repeated local adaptation to climate in distantly related plants. *Nat. Ecol. Evol.* **8**, 1933–1947 (2024).
6. Mascolo, E. *et al.* Long-term evolution of regulatory DNA sequences. Part 2: theory and future challenges. *Curr. Opin. Genet. Dev.* **98**, 102472 (2026).
7. Lenski, R. E. Revisiting the Design of the Long-Term Evolution Experiment with *Escherichia coli*. *J. Mol. Evol.* **91**, 241–253 (2023).
8. Garlapow, M. E. *et al.* Genetic and genomic response to selection for food consumption in *Drosophila melanogaster*. *Behav. Genet.* **47**, 227–243 (2017).
9. Barghi, N. *et al.* Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLoS Biol.* **17**, e3000128 (2019).
10. Tautz, D. *et al.* Beyond Mendel: a call to revisit the genotype-phenotype map through new experimental paradigms. *Genetics* **232**, (2026).
11. Mackay, T. F. C. & Anholt, R. R. H. Pleiotropy, epistasis and the genetic architecture of quantitative traits. *Nat. Rev. Genet.* **25**, 639–657 (2024).
12. Hendry, A. P., Day, T. & Taylor, E. B. Population mixing and the adaptive divergence of quantitative traits in discrete populations: a theoretical framework for empirical tests. *Evolution* **55**, 459–466 (2001).
13. Durkee, L. F., Olazcuaga, L., Melbourne, B. A. & Hufbauer, R. A. Immigration delays but does not prevent adaptation following environmental change: experimental evidence. *J. Evol. Biol.* **37**, 665–676 (2024).
14. Polechová, J. & Barton, N. H. Limits to adaptation along environmental gradients. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6401–6406 (2015).
15. Walsh, B. & Lynch, M. *Evolution and Selection of Quantitative Traits*. (Oxford university press, 2018).

REACTION DIAGRAM SUMMARY

We can metaphorically think of an ARIA programme as a chemical reaction. We present a simple reaction diagram to summarise the key elements of the imagined programme.

Think of each programme as a reaction



Predicting Evolution

① **Reactants:** The UK boasts successful frontier AI laboratories, commercial breeders, pharma companies, specialist AI-for-biology companies and emerging evolutionary-data platforms. Their approaches span sequence and multimodal foundation models, representation learning, generative models, latent-dynamics and world models, causal and uncertainty-aware learning, differentiable simulation, and scientific agents. These communities rarely encounter evolution as a clearly specified forecasting problem.

Complementary to these is the UK's strong and long-established communities—commercial and academic—for quantitative and environmental genetics, animal and plant breeding, evolutionary ecology and experimental evolution.

② **Reaction Design:** Reduce the process to its core process: evolution as the change in a population's allelic frequencies across generations. This refines the challenge of forecasting evolution to one of mapping the plausibility of gene frequency trajectories. Forecasting scenarios

will become successively more complex through the programme, demanding continuing improvement of forecaster skill. Compute and targeted support will allow smaller teams with promising ideas to participate in the challenges.

③ **Activation barrier — market failure:** Evolutionary forecasting has the characteristics of a public good: its benefits would be diffuse, long-term and shared across many scientific fields and industries, while the costs and risks of creating it are concentrated upfront. No individual laboratory, company or conventional funding programme has the remit or incentive to build the experimental scale, automation, repeated molecular measurements, compute, storage and sustained data presentation required. Existing mechanisms also struggle to fund the organisational work: translating the biology for other disciplines, recruiting and motivating unfamiliar communities, and supporting participants. ARIA can absorb these costs and risks while acting as the catalyst, opening the lower barrier pathway.

④ **Products:** Forecasting models that outperform reference models and transfer to new settings. An unprecedented, machine-learning-ready open dataset of thousands of evolutionary trajectories, spanning genomes, expression, chromatin state, phenotype and population fate across up to 80 generations. A standing benchmark, scoring framework and automated experimental platform; and a community capable of extending the applications to many other life science contexts.

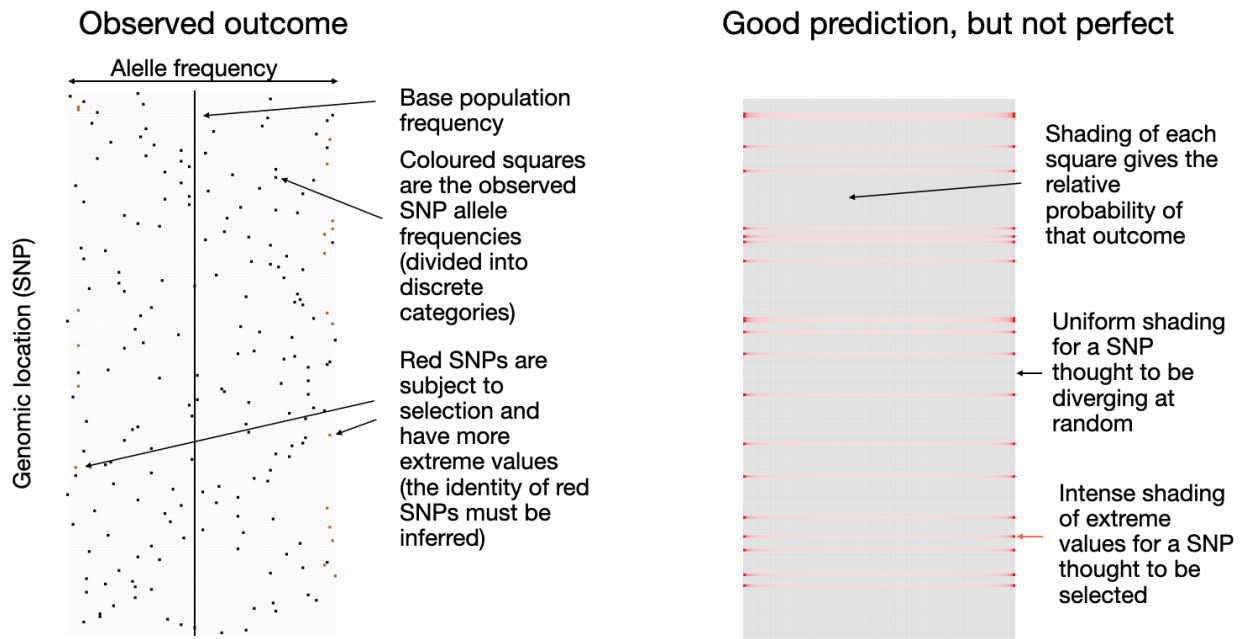
⑤ **Energy Released:** reliable evolutionary forecasting could reduce costly failures in oncology, gene and cell therapies, vaccine and drug escape, invasive-species control, selective breeding of domesticated crops and animals, and climate adaptation. An ability to forecast evolution would be a game-changer, as would identification of the forecast horizon.

ANNEX: Competition challenges and measures of prediction

- + **Predicting persistence:** for each challenge there will be a prize for forecasting demographic fate: whether existing populations persist or collapse (Scenarios 1 + 2), and which habitats become populated as a species evolves to expand its range (Scenario 3).
- + **Predicting genetic change:** the parallel competition for teams forecasting the genetic changes will run after the results of the demographic competition have been released, so the identity of the surviving populations is known. In this case the teams will forecast the allele frequencies at variable sites (loci) throughout the genome.

ANNEX: Competition challenges and measures of prediction

Genetic forecasts

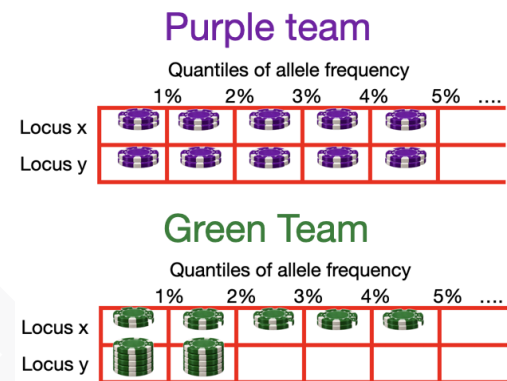


Demographic forecasts

For challenges 1 + 2: Once the experiment is run, we will release the proportion, f , of the 20 replicates that persisted (produce surviving larvae before gene flow) in each of the lines for each selection regime. The teams' goal is to forecast which those are. The forecast will be returned as the persistence probability, p_i , for each replicate. Their score will be $\log(p_i / f)$ for populations that persist and $\log(1 - p_i / 1 - f)$ for those that fail. The sum will be zero if they simply predict f for every population, and will be larger the more accurate their p_i values are. For challenge 3 the value of f will only be released for the first generation of the 20 generation series.

Evolutionary forecasts.

For challenges 1 + 2: Where selection is applied to single populations, the forecasting procedure can be thought of as placing tokens on a board in a game of chance. The squares represent a range of allele frequencies which are equiprobable under random genetic drift. The team's score is the number of tokens staked on the correct outcome (they have 100 in total per locus, and can stake a proportion of a token). In these illustrative examples the Purple Team forecasts that allele divergence is shaped by random drift rather than by selection - hence they spread their bets evenly. The Green Team agrees that locus x is shaped by genetic drift, but predicts selection will act strongly on locus y (an intense selective sweep). If the reality for the population is that the alternative allele is being driven towards fixation, the Green team will therefore beat the Purple team (but Purple wins otherwise).



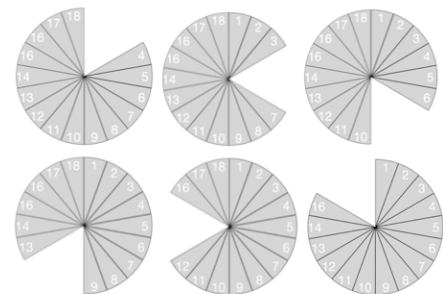
For challenge 3: where selection is applied to a system of interconnected populations the teams will be asked to specify **where** each allele is found. They will do this by forecasting the relative occurrence of the allele in each population, f_i ($\sum_i f_i = 1$), so the expected frequency of an allele with mean frequency $\bar{p}$ over n populations is $p_i = n\bar{p}f_i$. The score rewards the team that matches the observed frequency, o , with the greatest accuracy: $-\sum_i (p_i - o_i)^2 / o_i(1-o_i)$. The values of $\bar{p}$ will be released along with the results of the demographic competition.

Forecasting Challenges

Scenario 1

We will establish base populations from 18 *Caenorhabditis remanei* inbred isofemale lines from a single wild source population. Different combinations will be created by intercrossing the 18 lines, excluding 3 of the lines from each cross ($\{1,2,3\}, \{4,5,6\} \dots \{16,17,18\}$). In this manner approximately $\frac{1}{3}$ of the segregating sites will be entirely missing from at least one combination.

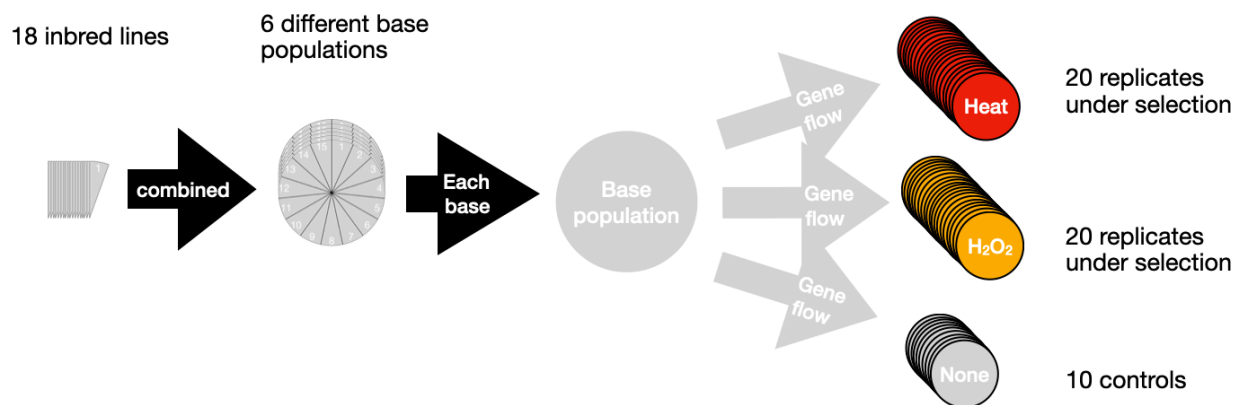
Each base population excludes 3 lines



Will expose the same genetic variants to selective challenges in different combinations.. After 10 generations of adaptation to the culture conditions, each line will be exposed to

two forms of selection of the larval stage in liquid suspension and a control:

- + 20 replicate populations will be exposed to acute H_2O_2 stress
- + 20 replicate populations will be exposed to acute heat stress
- + 10 replicate control populations will not be exposed to any stress.



From generation 0 (population founding) to generation 7, selection will be at an intensity which kills 40% of individuals, after which the intensity will increase by 3.5 percentage points per generation or until half (10) of the replicate populations have gone extinct. We will generate the following data from pools for each replicate and the base population:

- genomic sequences (9,000x for selected populations, 90,000x for the base);
- expression data for synchronised day-4 adults (polyA+ RNA and sncRNA) and ATAC-seq.

Sampling will occur at generations 0, 4 and 7 and then released. The demographic challenge will be to identify which populations survive. Each inbred line will be characterised by assembling the genome of 5 individuals using Picogram input multimodal sequencing (PiMmS) sequencing offered by the Tree of Life initiative at the Sanger in support of this project. After the demographic competition results have been released, teams participating in the genetic challenge will be asked to use that demographic information (along with the BCF variant fields specifying the SNPs and alleles, including any mutations since t_7).

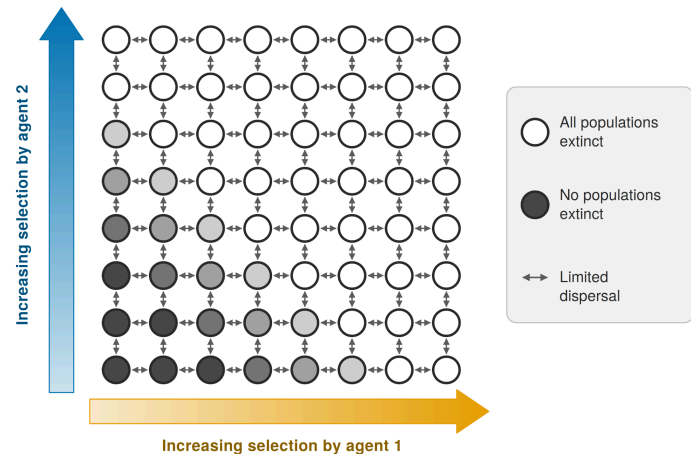
Scenario 2

This challenge is a step up in difficulty, as it tests the teams' ability to forecast how genetic effects are reshaped in independently evolved genetic backgrounds⁷. The lines will be created by collections from 12 new populations. They will be expanded up from large

founding populations taken from new collections in the wild, without the usual inbreeding which generates blocks of correlated alleles around the genome (local linkage disequilibrium). We will assemble 100 individual genomes from each base population to characterise the haplotype structure. Line construction can be underway from the start of the project and the challenge set before the end of year 2. The timing of the challenges and the release of data will proceed as in Challenge 1.

Scenario 3

This most ambitious test will require the teams to extend their predictions to the evolutionary processes that allow a species to adapt to spread into new environments. They will forecast both the successes – adaptation to colonise habitats that were initially outside their range (empty populations in the schematic showing the start of the experiment) and the limits to adaptation at the range margin.



The same source populations as used in Challenge 2 will be introduced across an array of 'habitats', each of which represents a breeding population whose persistence and genetic evolution will be the subject of the forecasting challenge, as before. Importantly, these habitats are not isolated: each generation there will be limited demographic exchange between adjacent habitats. Trials will be underway before the end of year 2 of the Programme, and completed 18 months later.

Artificial selection will be fixed for each habitat, with selection intensities of both agents spanning a range from no artificial selection to lethal conditions. Over time, viable (i.e., sustained) breeding populations are expected to establish in the more extreme environments. At the range margins we expect to see loss of genetic diversity¹⁰, a change in the distribution of the effects of surviving alleles, selection for new genetic combinations (changes to linkage disequilibrium and genomic architecture) and regulatory interactions. If innovations in automation allow. For each of the 12 base populations, the matrix will be replicated 3-fold (making a total of $64 \times 12 \times 3 = 2,304$ evolutionary trajectories). The experiment will be run for 80 generations. The teams will be invited to predict which populations produce offspring that survive the selective challenge (a list of populations for each generation in a 20-generation time course), and the distribution of each allele among the populations. Competitions will be run at 20 generation intervals after the release of data at generations 20, 40 and 60.