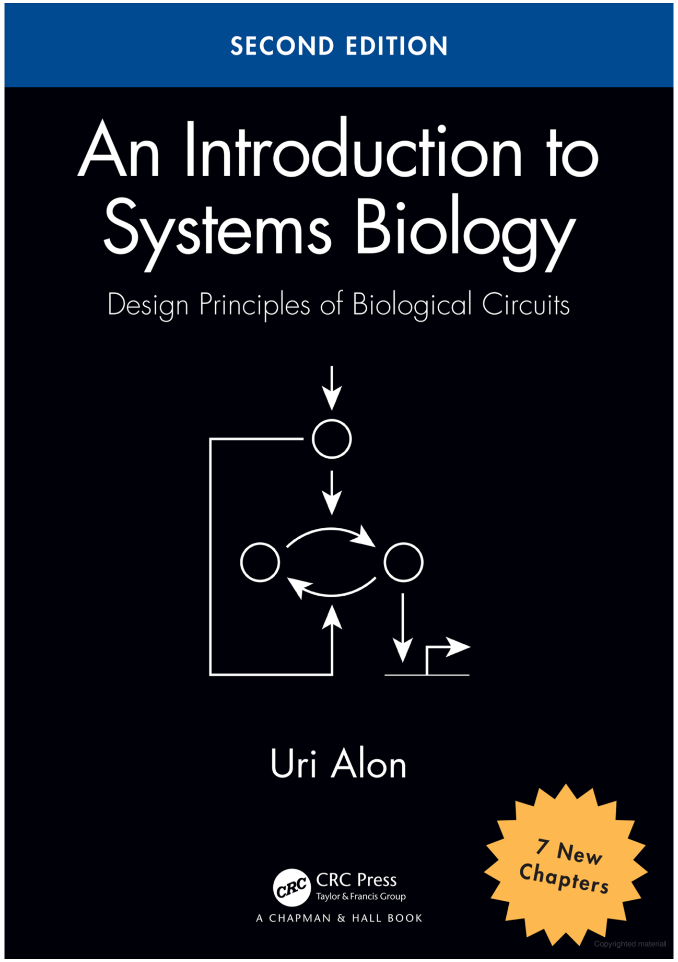
Quantitative principles in biological systems

12. Diversity and Multi-objective Optimality in Biology

Yifan Yang, Asst. Prof., Center for Interdisciplinary Studies (CIS), Westlake University

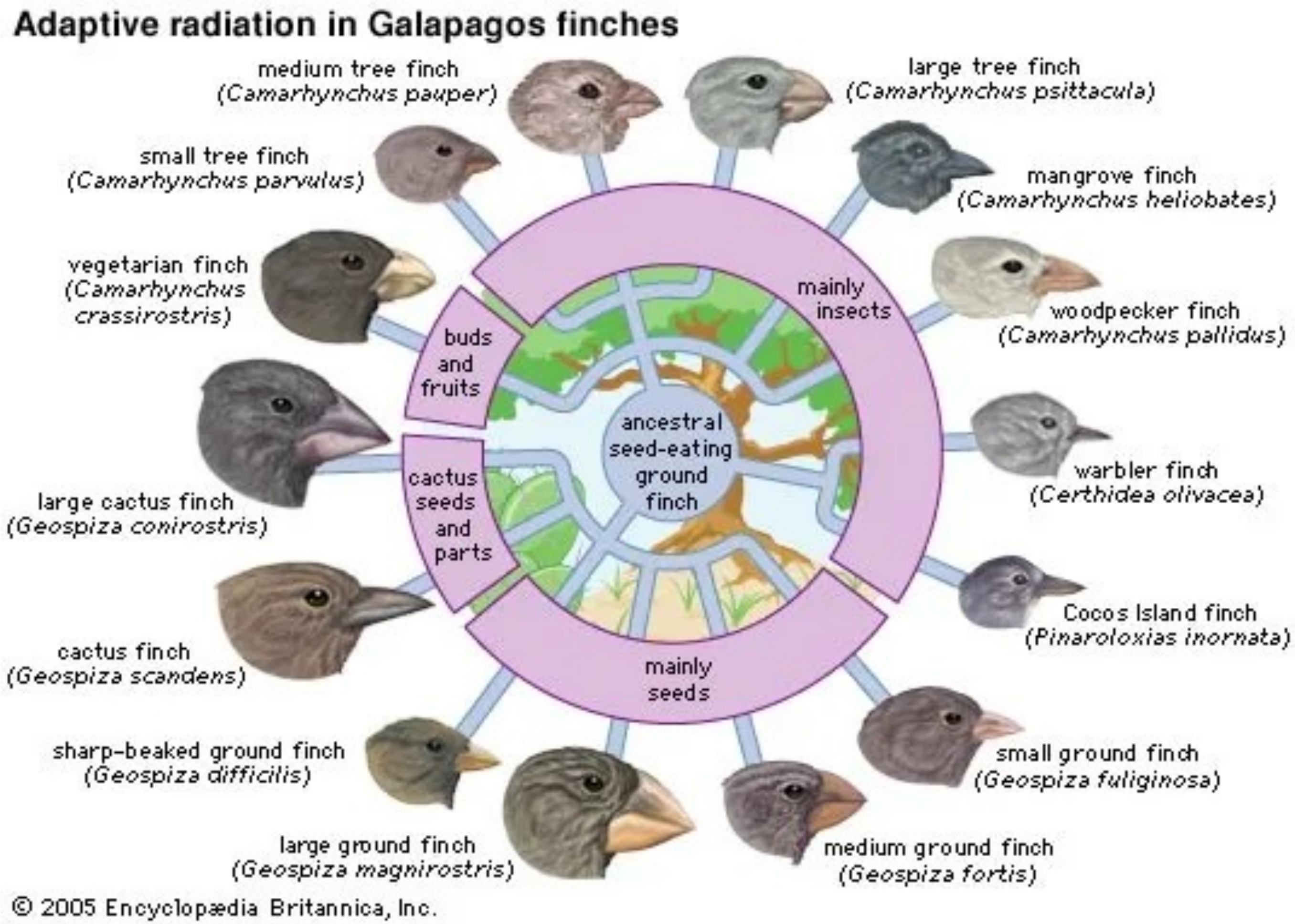
May 23rd, 2025

Chapter 14. Multi-objective Optimality in Biology

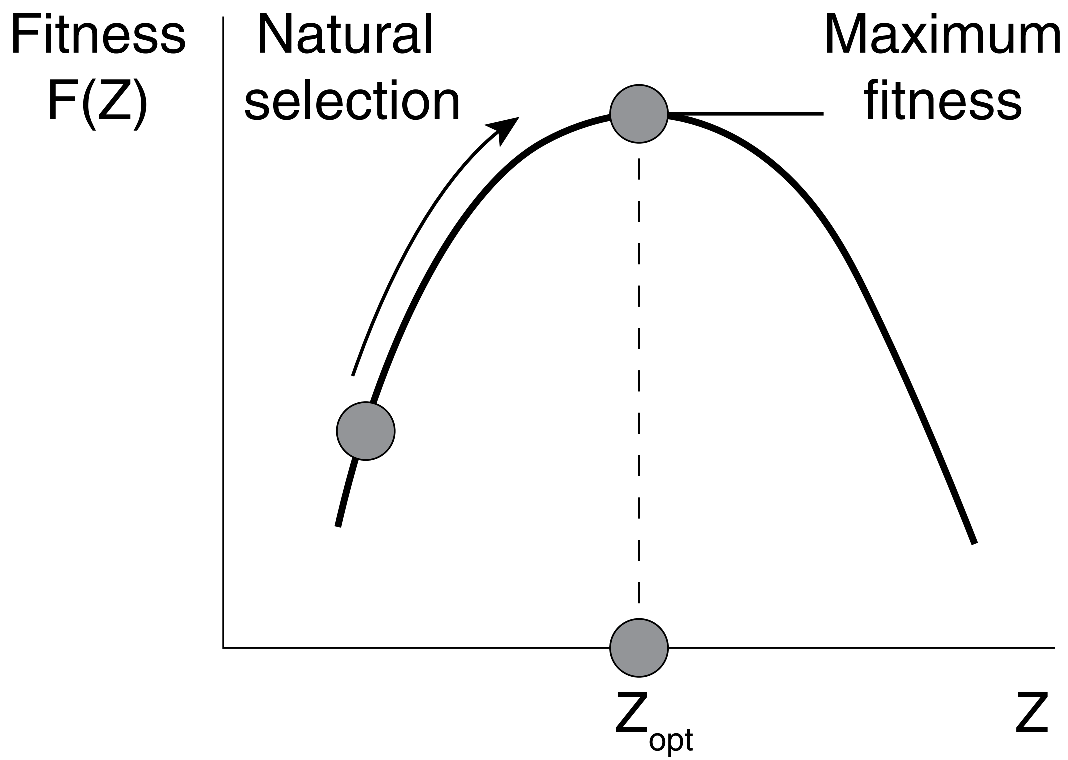


Alon (2020) 'An introduction to Systems Biology. 2nd Ed.' CRC Press

The diversity of finches on Galapagos islands inspired Darwin to develop his theory of natural selection



The classical picture of natural selection involves the fitness landscape

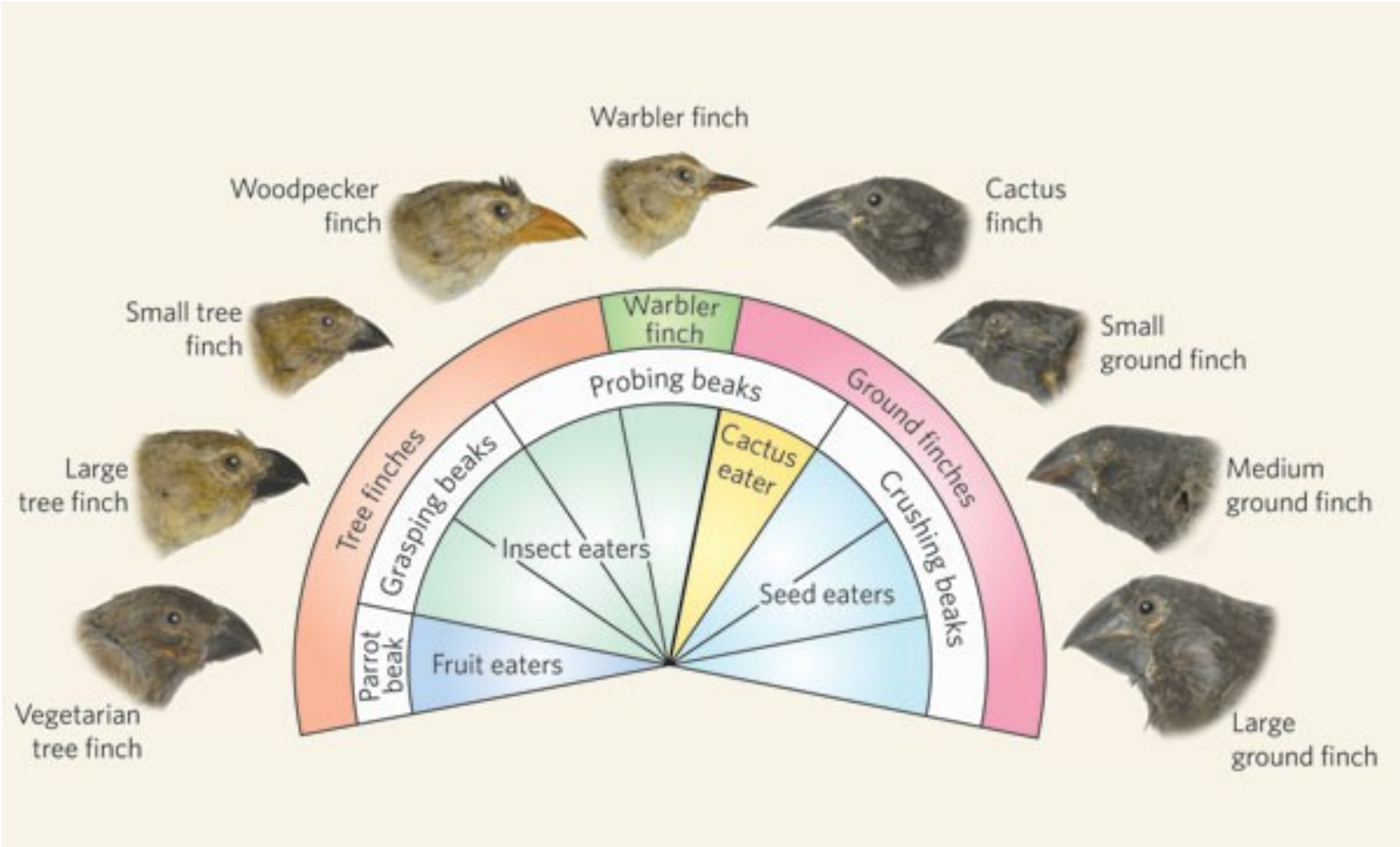


genotype → phenotype → fitness

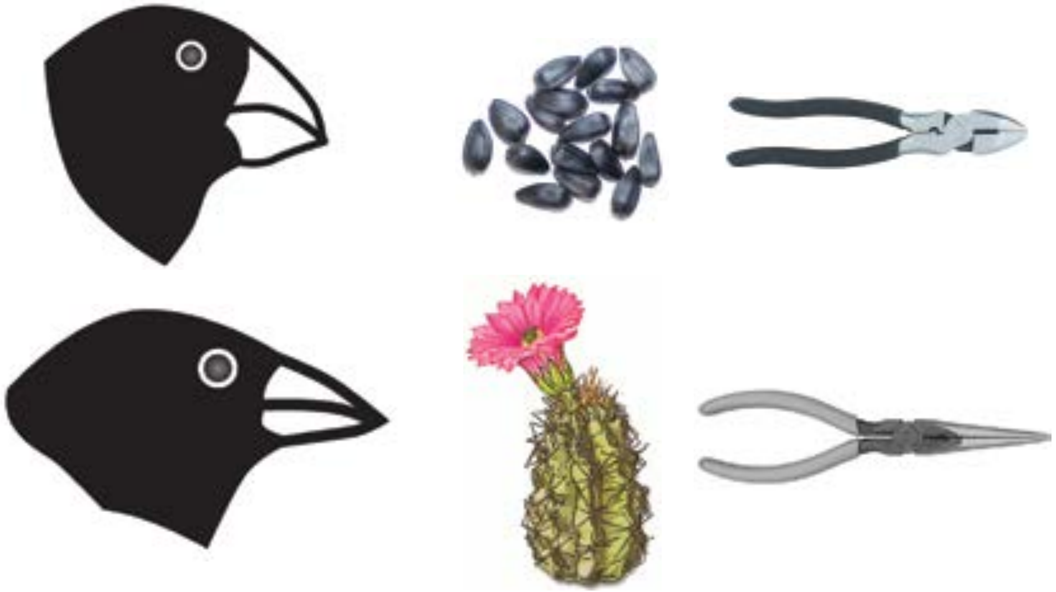
Peter and Rosemary Grant observe the evolution of Darwin’s finches in real time over 30 years

Daphne Major @ Galapagos

Finches with different beaks evolved to specialize in distinct tasks



The fitness function often involves trade-offs among multiple objectives

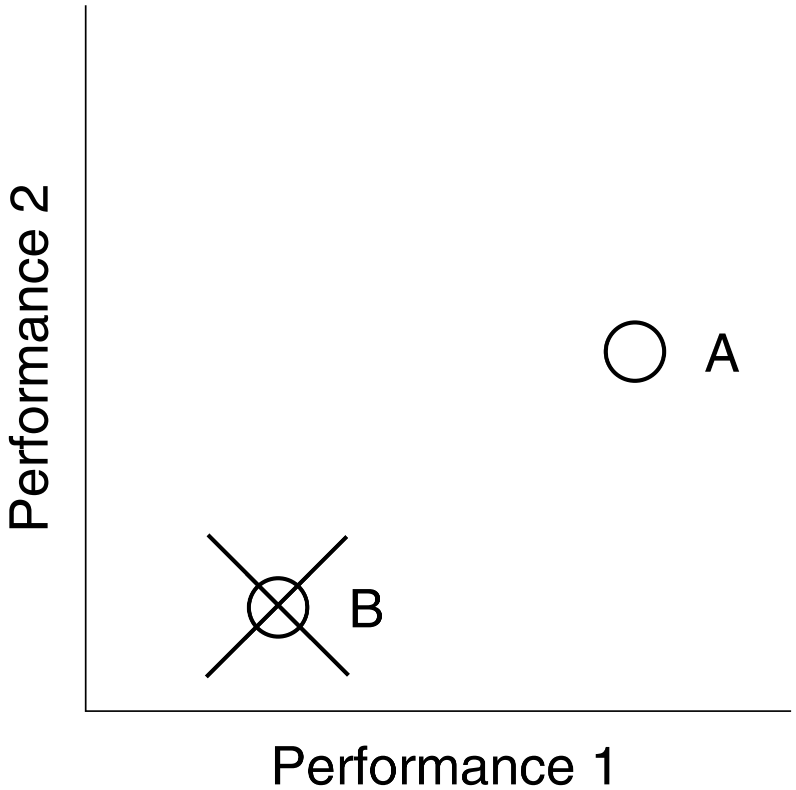


Multiple tasks are characterized by performance functions

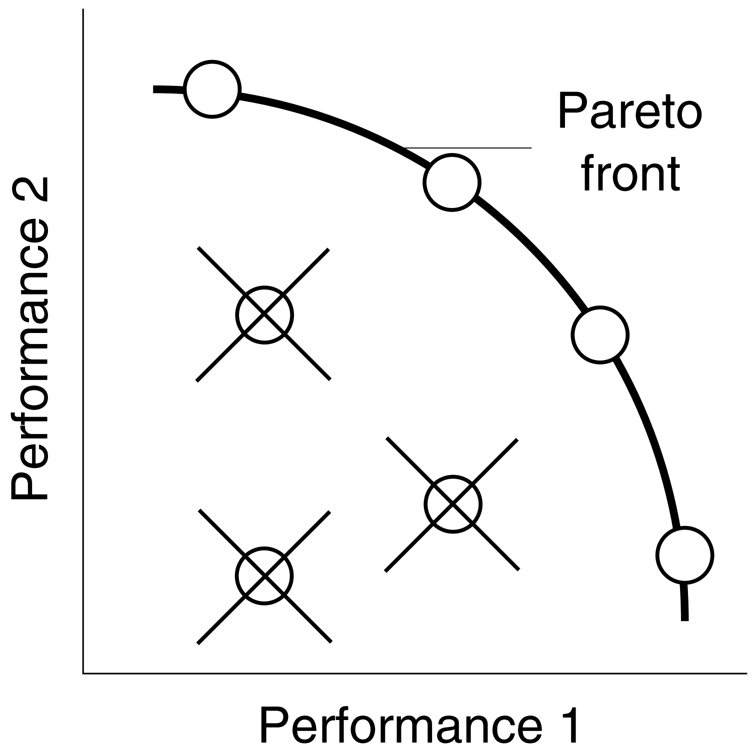
$$
\begin{array}{ccccc}
 & & \text{performance at task 1} & & \\
 & \nearrow & & \searrow & \\
\text{genotype} \rightarrow \text{phenotype} & & & & \text{fitness} \\
 & \searrow & & \nearrow & \\
 & & \text{performance at task 2} & &
\end{array}
$$

$$F = F(P_1(\vec{T}), P_2(\vec{T})) \qquad \partial F / \partial P_i > 0$$

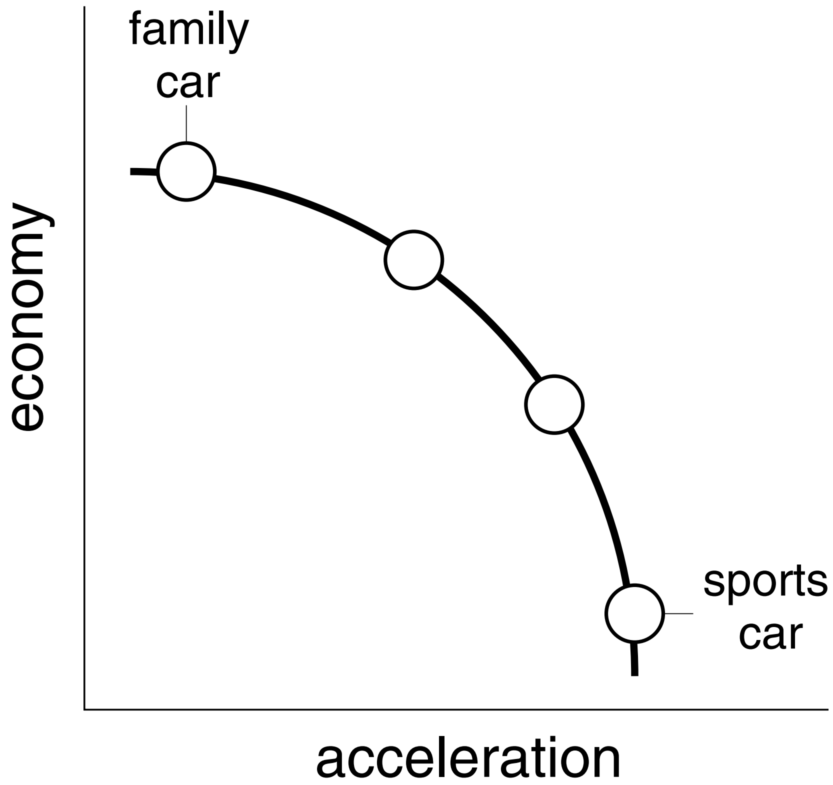
Engineers and economists use Pareto efficiency to study multi-objective optimization



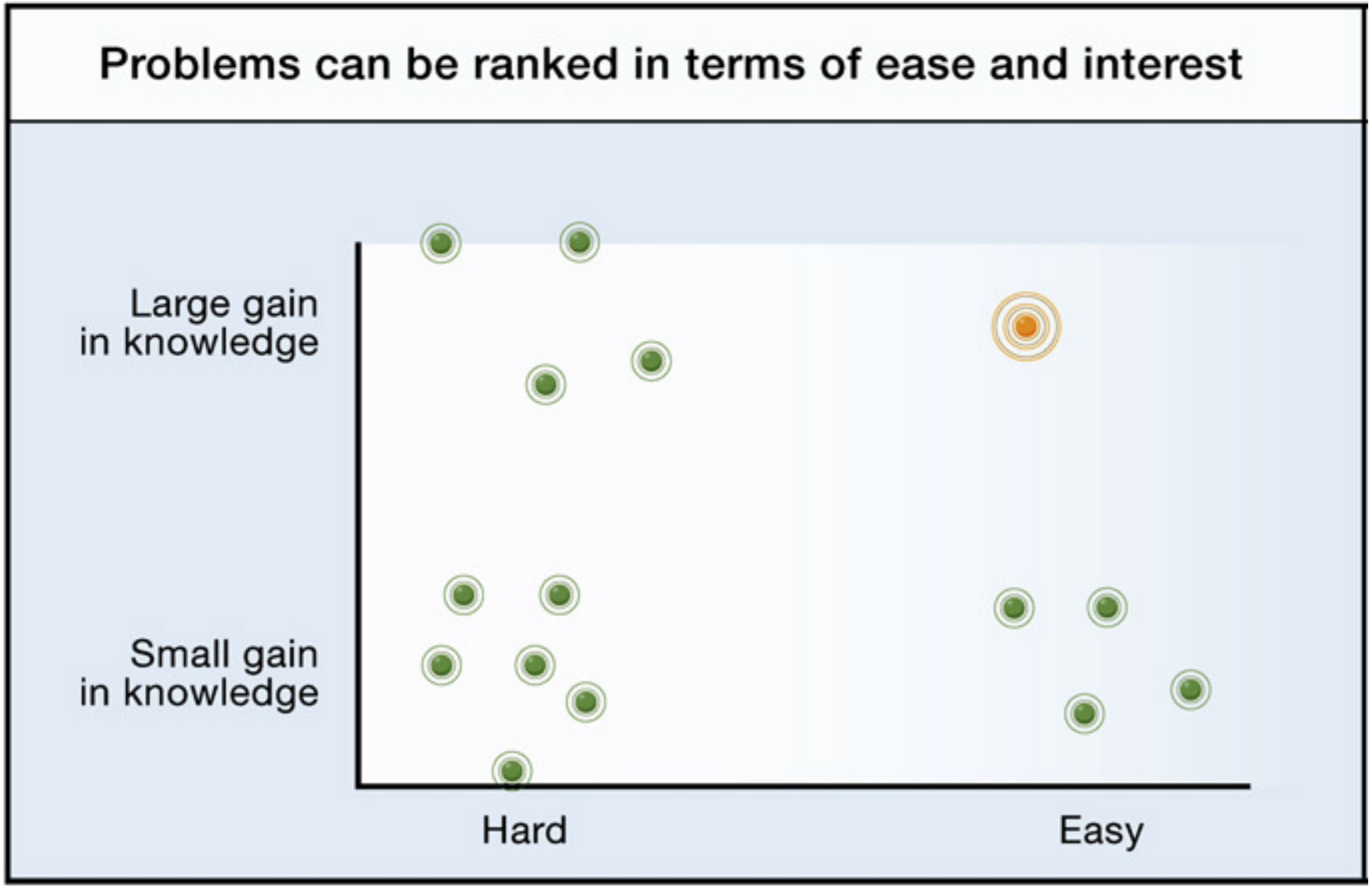
Pareto improvements lead to the Pareto front, the set of all Pareto efficient solutions



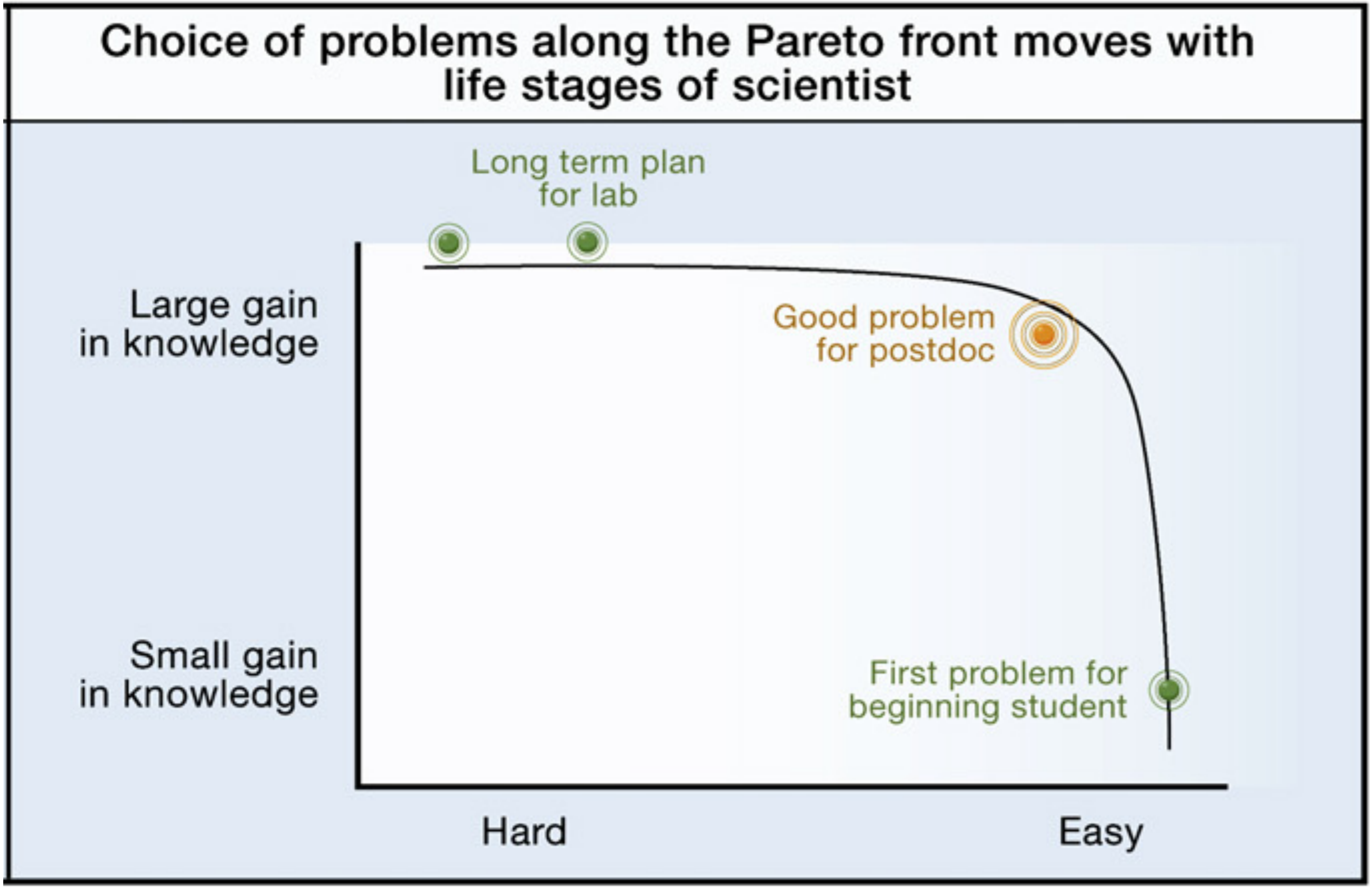
In engineering, we choose optimal designs based on existing performances under trade-offs



Another application of Pareto optimality is the choice of research projects



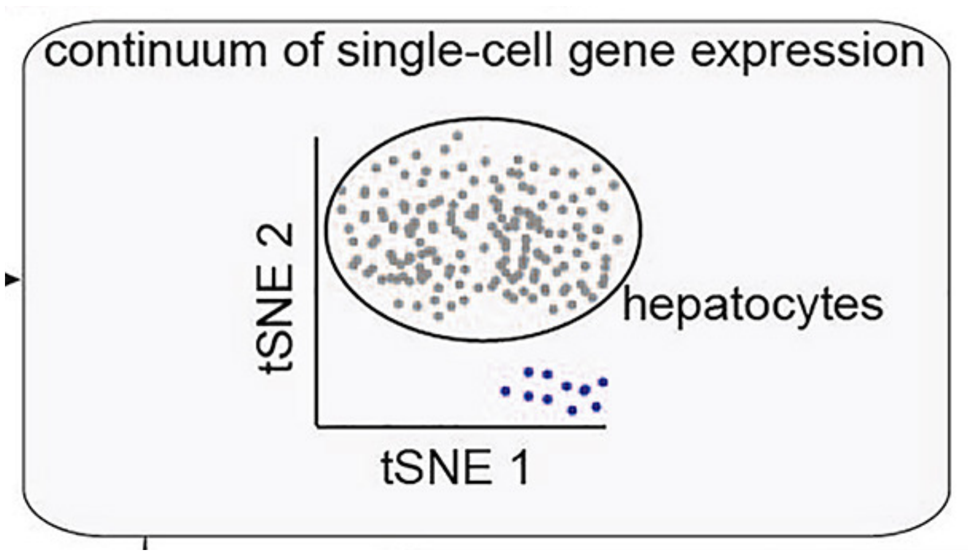
Alon (2009) Mol. Cell



Alon (2009) Mol. Cell

In biology, we often observe phenotypes without knowing the performances

Ammonite fossiles



Single-cell gene expressions

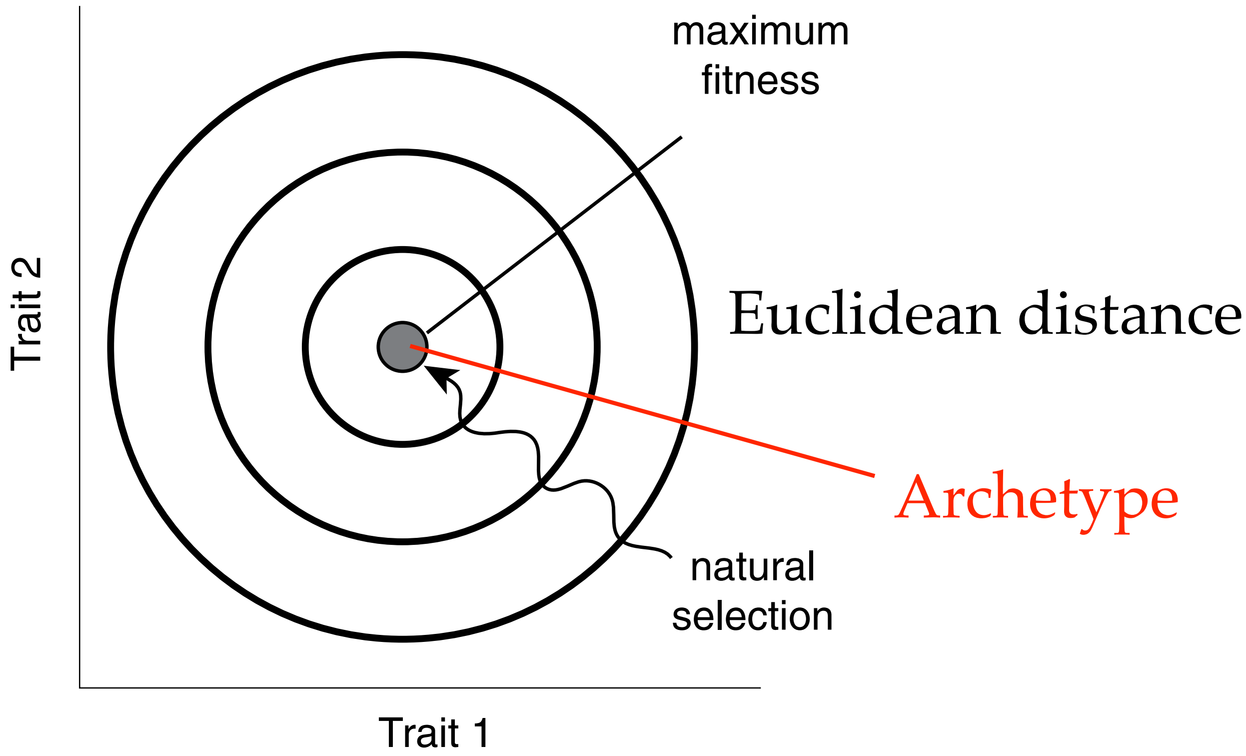
Can we infer the objectives using multi-dimensional phenotypic data?

Yes, we can. By ParTI-ing (Pareto task inference)

Shoval et al (2012) Science

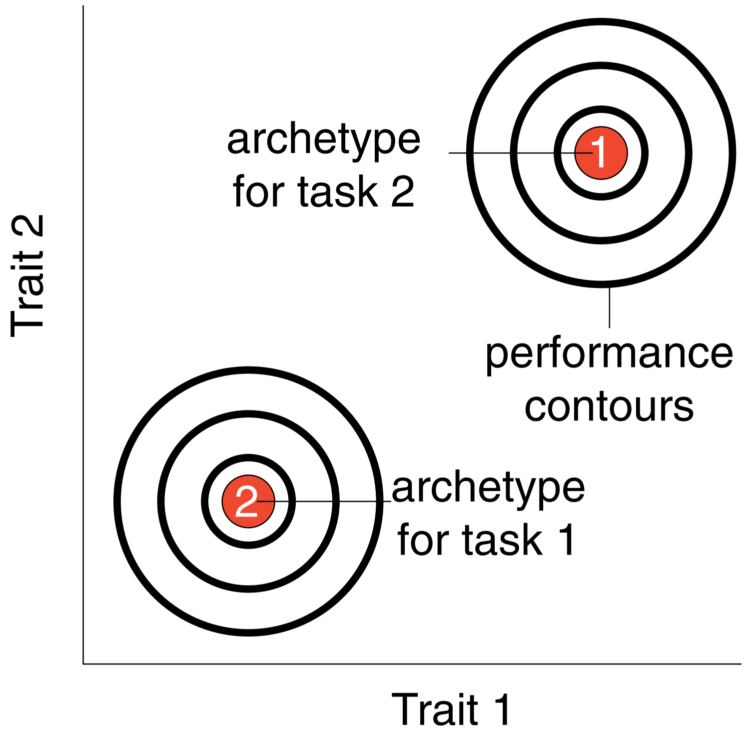
Hart et al (2015) Nat. Methods

Assumption 1: Performances drop with increasing metric distances in the trait space



Archetype: Best traits for a given task

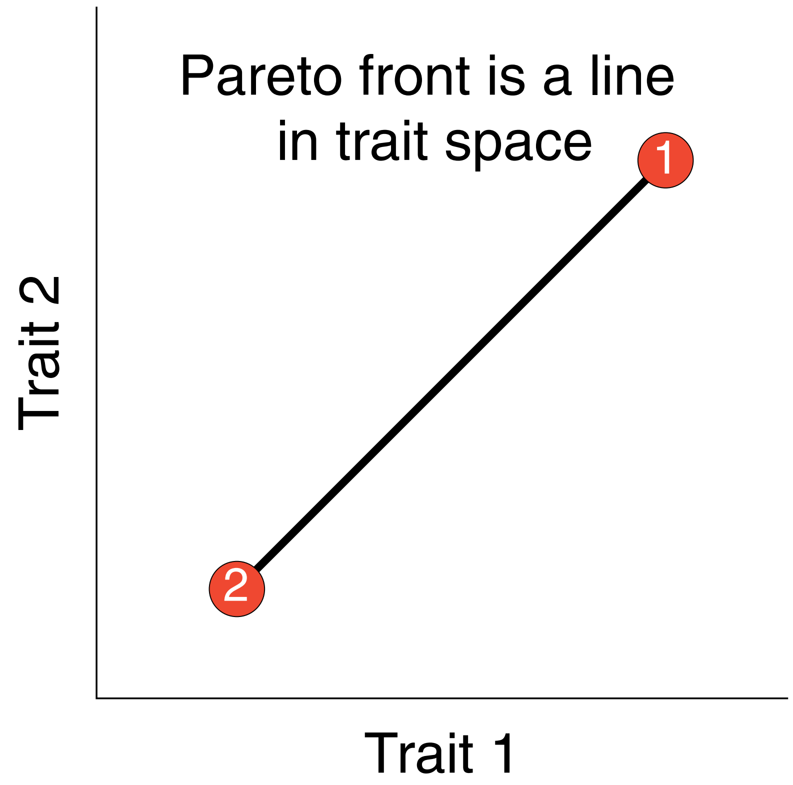
Assumption 2: Fitness is unknown function of 2 different tasks



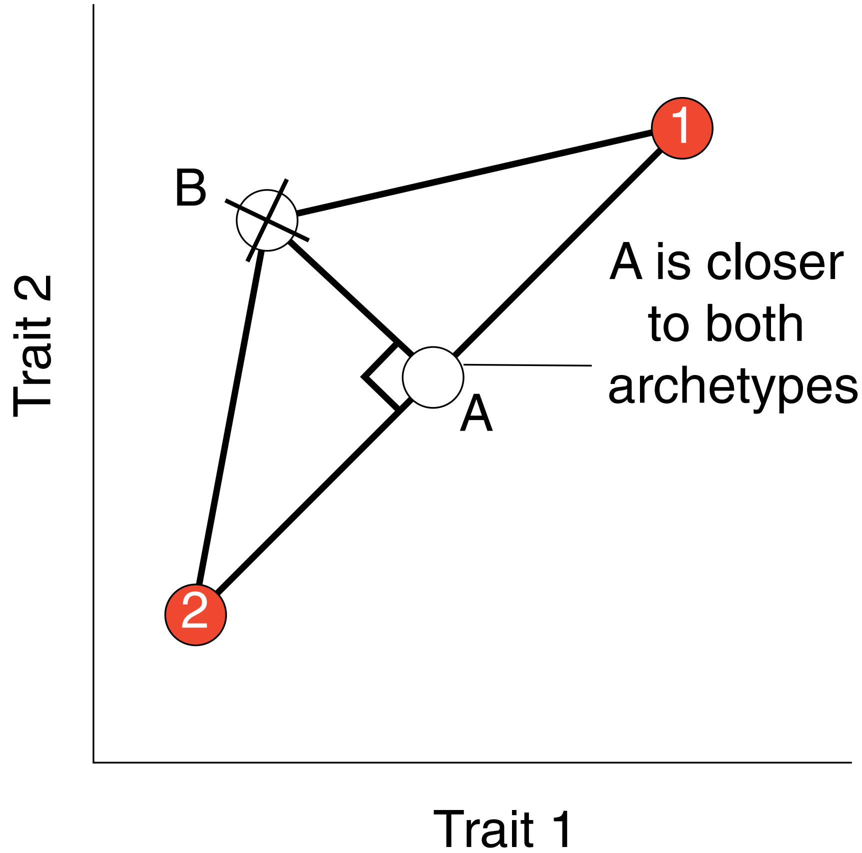
$$F = F(P_1(\vec{T}), P_2(\vec{T}))$$

$$\partial F / \partial P_i > 0$$

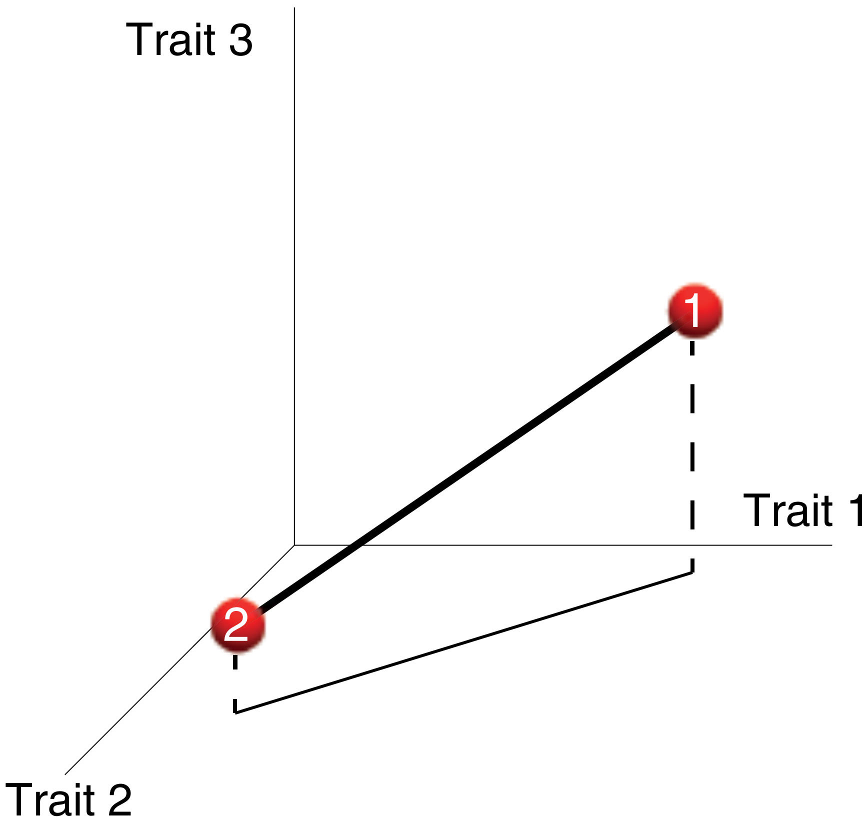
Theorem: The optimal solution must fall on the line segment that connects the 2 archetypes



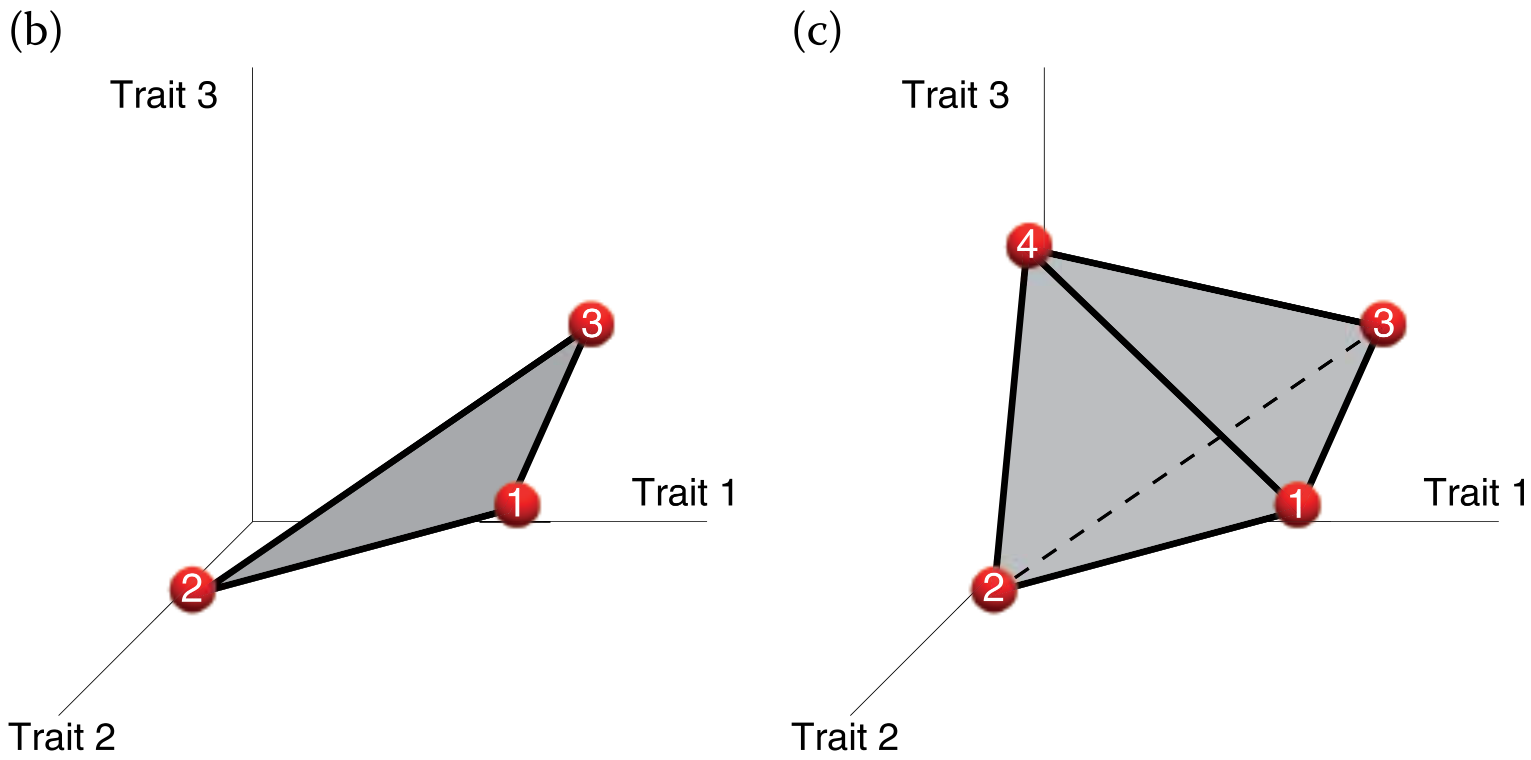
We can prove the theorem by comparing arbitrary phenotype B not on the line, with its counter part A on the line



A trade-off between 2 tasks predicts a line in trait space, no matter what exactly the traits are



3 tasks lead to a triangle and 4 tasks lead to a tetrahedron



k tasks will lead to a polytope with k vertices, given that trait space is at least (k-1) dimensional

We can prove that the optimal traits should be convex combination of the archetypes

$$T_{opt} = \Sigma_i w_i A_i, w_i > 0, \Sigma_i w_i = 1$$

$$\frac{dF}{dT}(T_{opt}) = 0, \frac{d^2F}{dT^2}(T_{opt}) < 0$$

$$\partial F / \partial P_i > 0$$

$$\frac{dP_i}{dr_i} < 0 \qquad r_i = \|T - A_i\|$$

Consider the derivative of fitness when we move in trait space relative to the archetypes

$$\frac{dr_i}{dT} = 2(T - A_i)$$

$$T_{opt} = \Sigma_i \frac{\theta_i A_i}{\Sigma_j \theta_j}$$

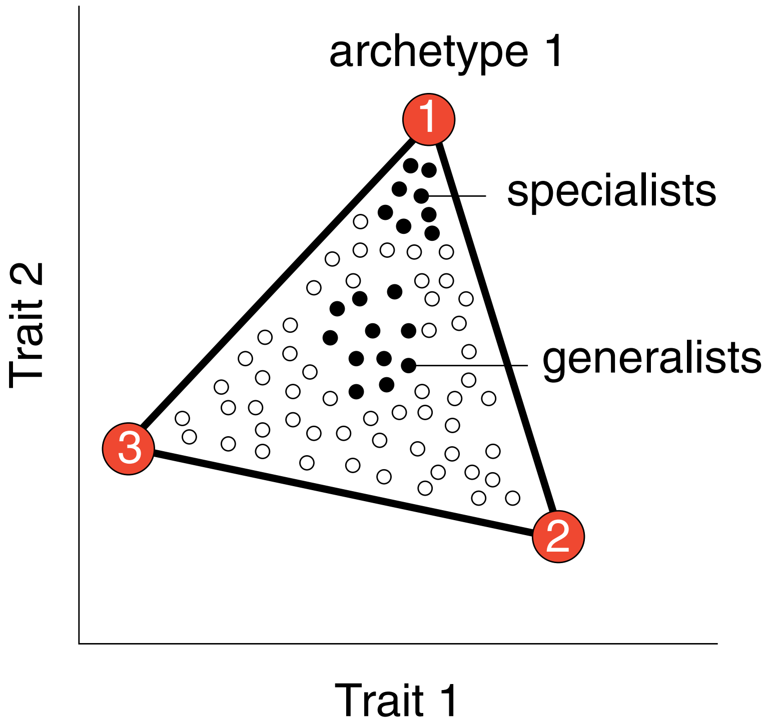
$$\frac{dF}{dT} = \Sigma_i \frac{\partial F}{\partial P_i} \frac{dP_i}{dr_i} 2(T - A_i) = 0$$

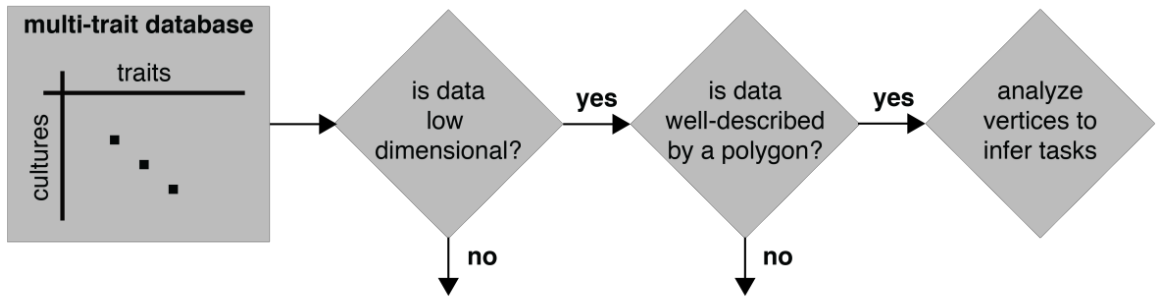
$$w_i = \frac{\theta_i}{\Sigma_j \theta_j} > 0, \Sigma_i w_i = 1$$

$$\theta_i = \frac{\partial F}{\partial P_i} \frac{dP_i}{dr_i} < 0$$

If the distance is not Euclidean or the trains are transformed nonlinearly, the polytope will be deformed

We can use the polytopes to discover what the tasks are directly from phenotypic data





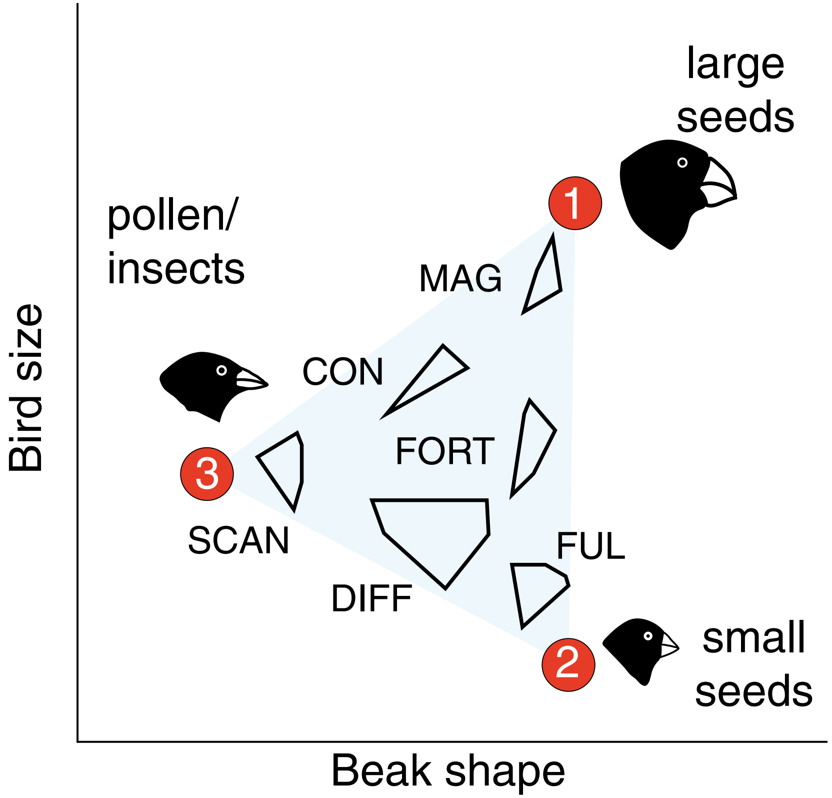
Hart et al (2015) Nat. Methods

Before we see some examples in biology,

Take a deep breath,

Talk to the person next to you, and ask questions.

We will start from animal evolution before moving to proteins and genes



Darwin's finches

Grant (1986) Princeton University Press

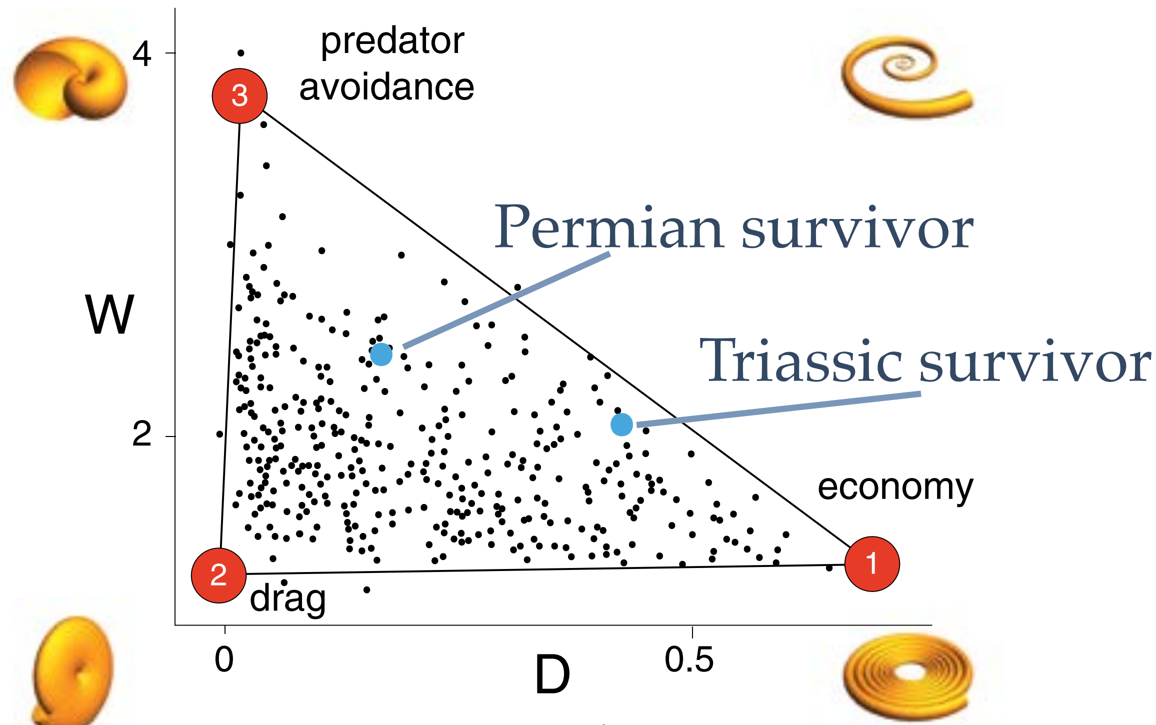
Shoval et al (2012) Science

Archetypes can last over geological timescales

$$R \sim e^{W\phi}$$

Ammonites

菊石

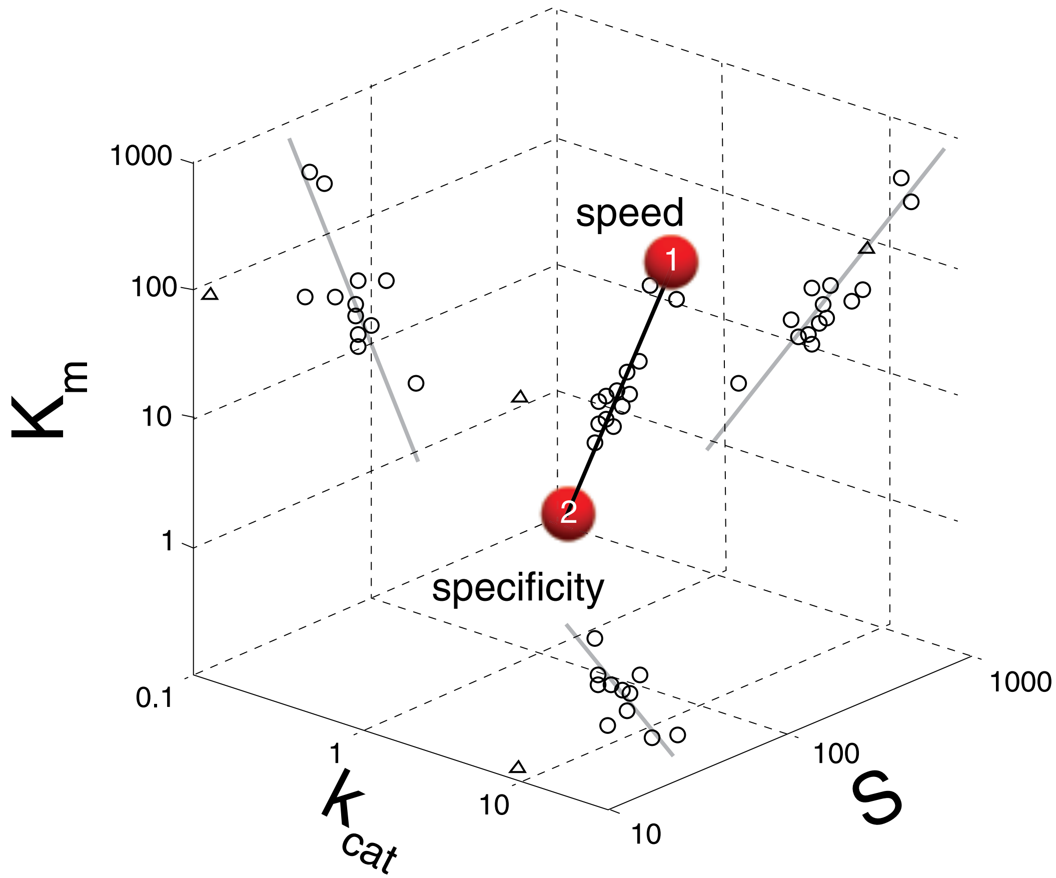


$$D = R1/R2$$

Kennedy (1989) Proc. Geo. Asso.

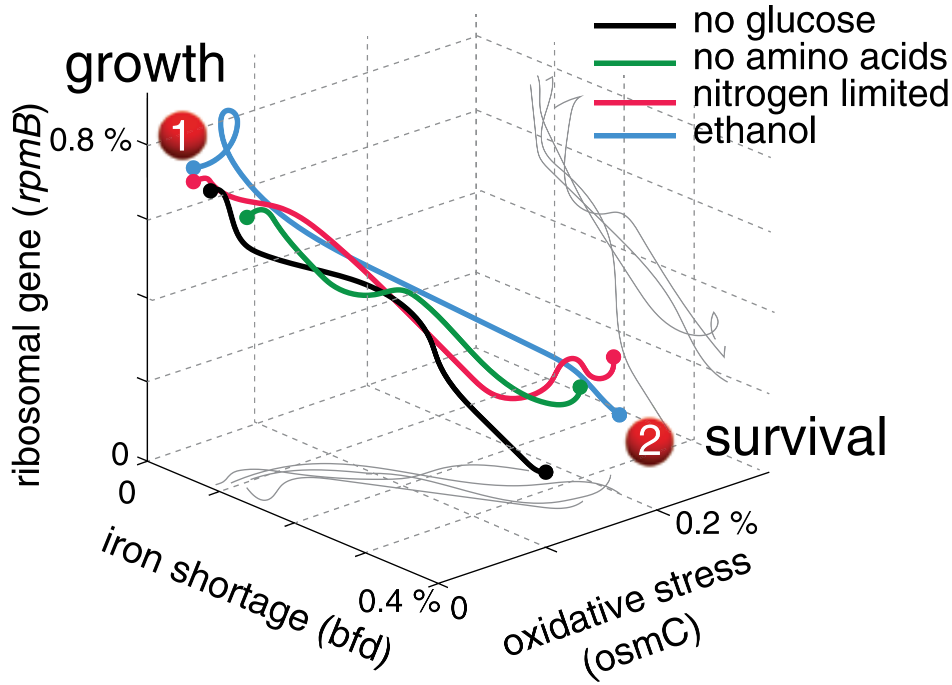
Tender, Mayo and Alon (2015) BMC Systems Biol.

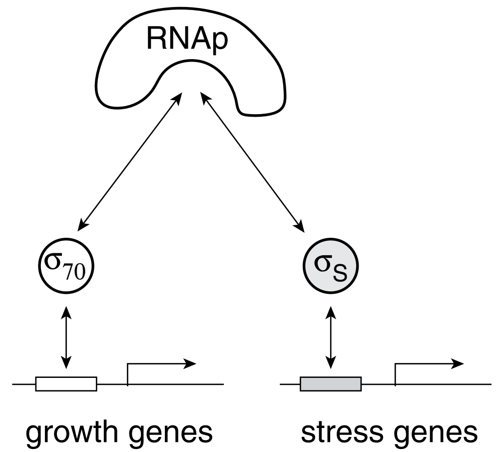
Specificity-Speed trade-off define the diversity of Rubisco in a 4 kinetic traits



Savir et al (2010) PNAS

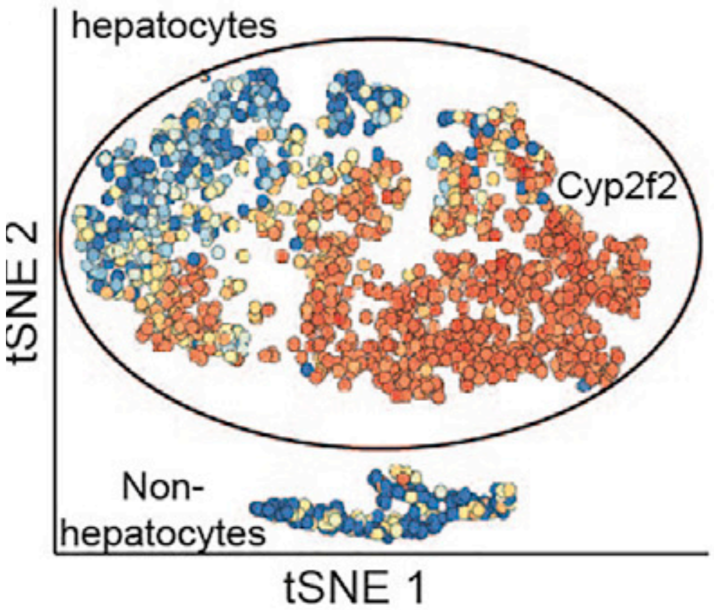
Growth-survival trade-off explain the gene expression variation of the top 200 promoters in *E. coli*

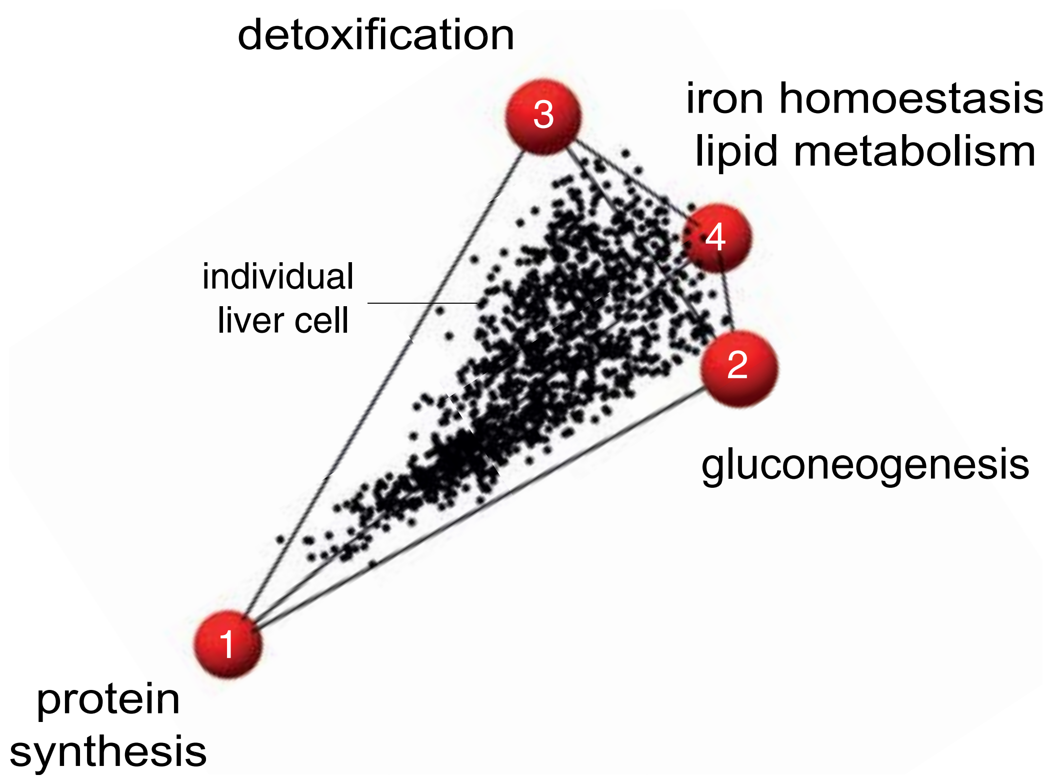




Shoval et al (2012) Science

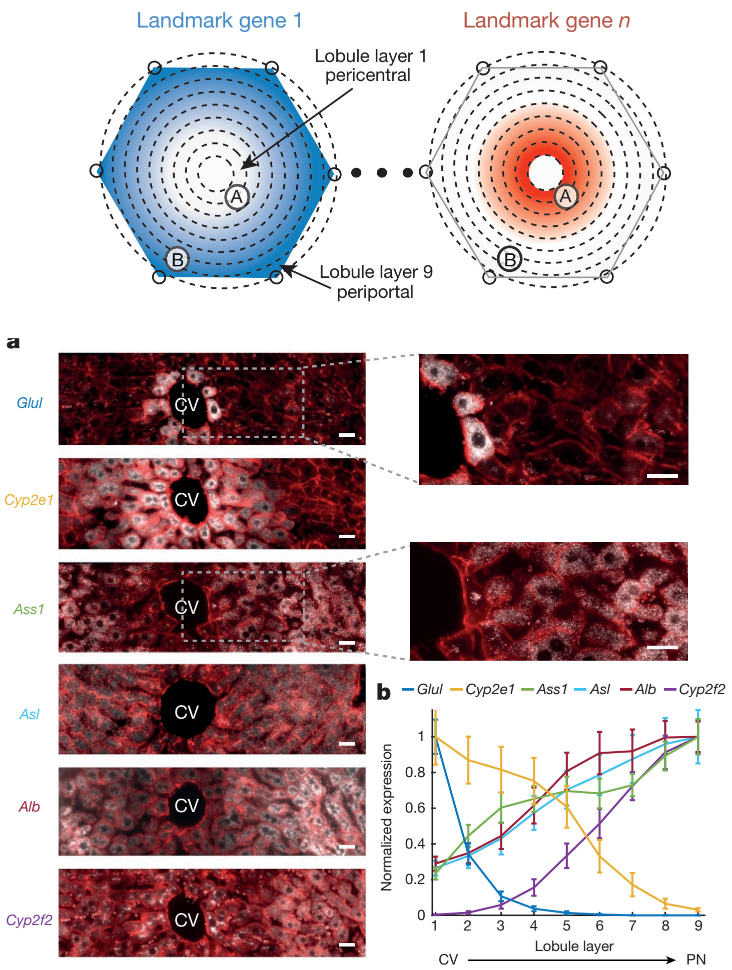
ParTI can infer division of labor within a differentiated cell type



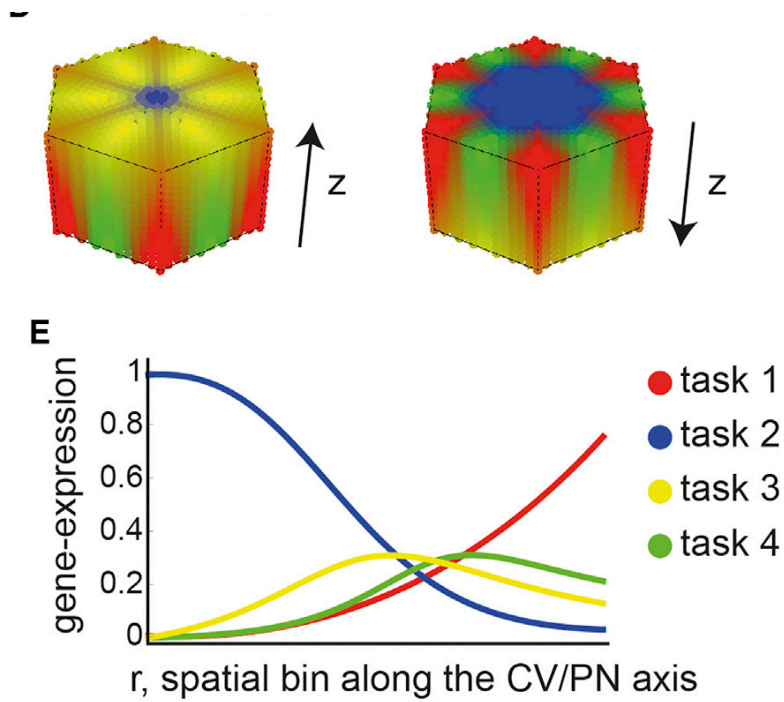


Adler et al (2019) Cell Systems

Spatial zonation of task-specialist cells arises from task-performance gradients

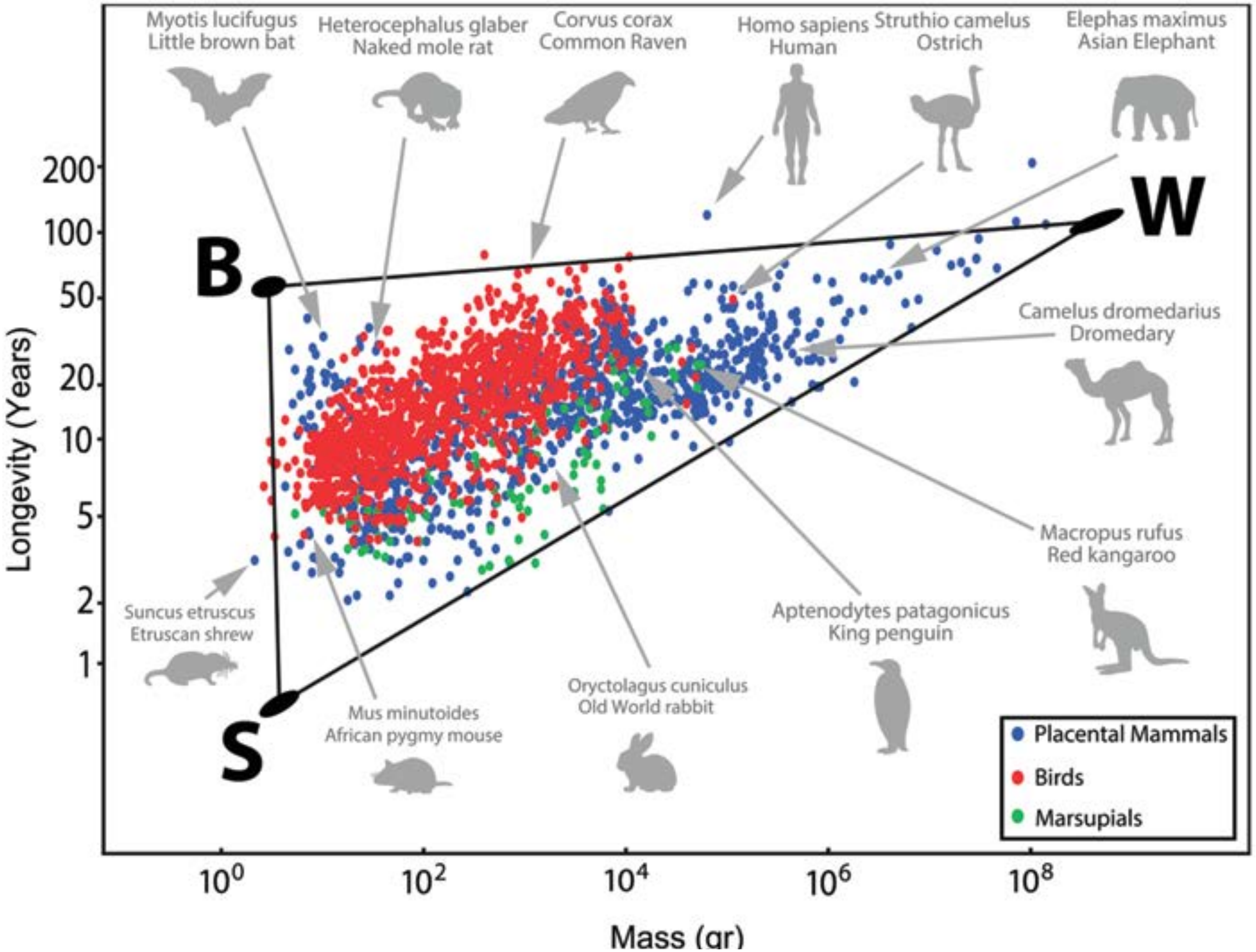


Halpern et al (2017) Nature



Adler et al (2019) Cell Systems

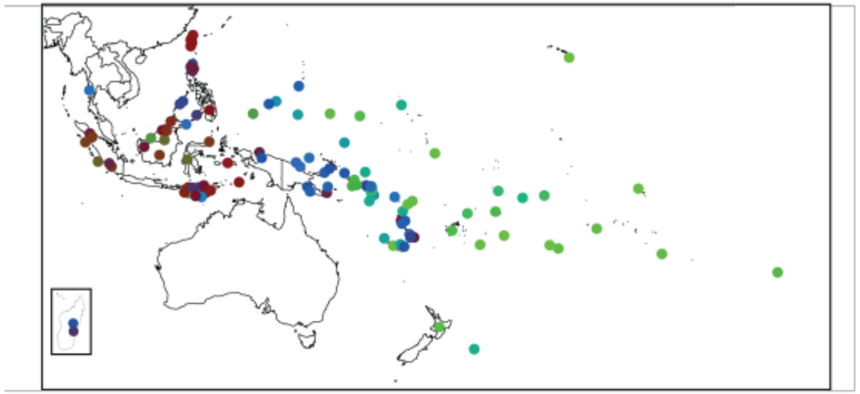
3 archetypes in the mass-longevity space define the life-history trade-offs of mammals and birds

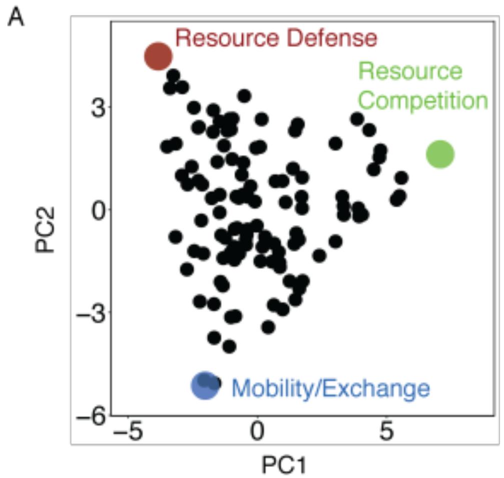


Szekely et al (2015) Science

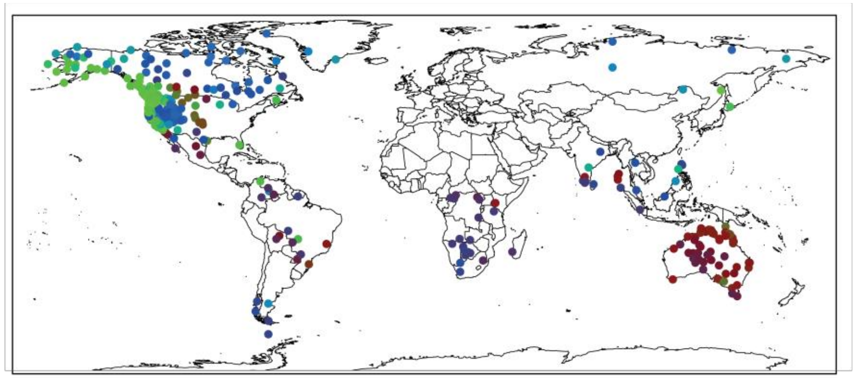
The same 3 archetypes of human culture can be inferred from linguistic/cultural diversity

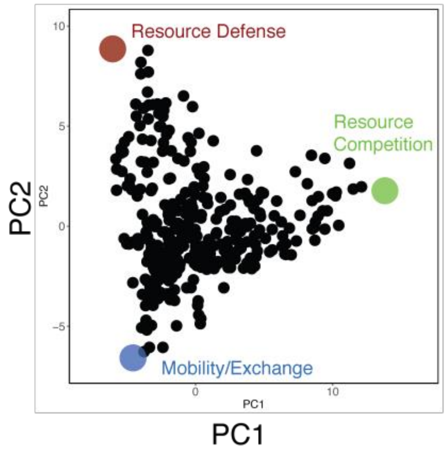
Austronesian linguistic traits





Hunter-gather cultural traits

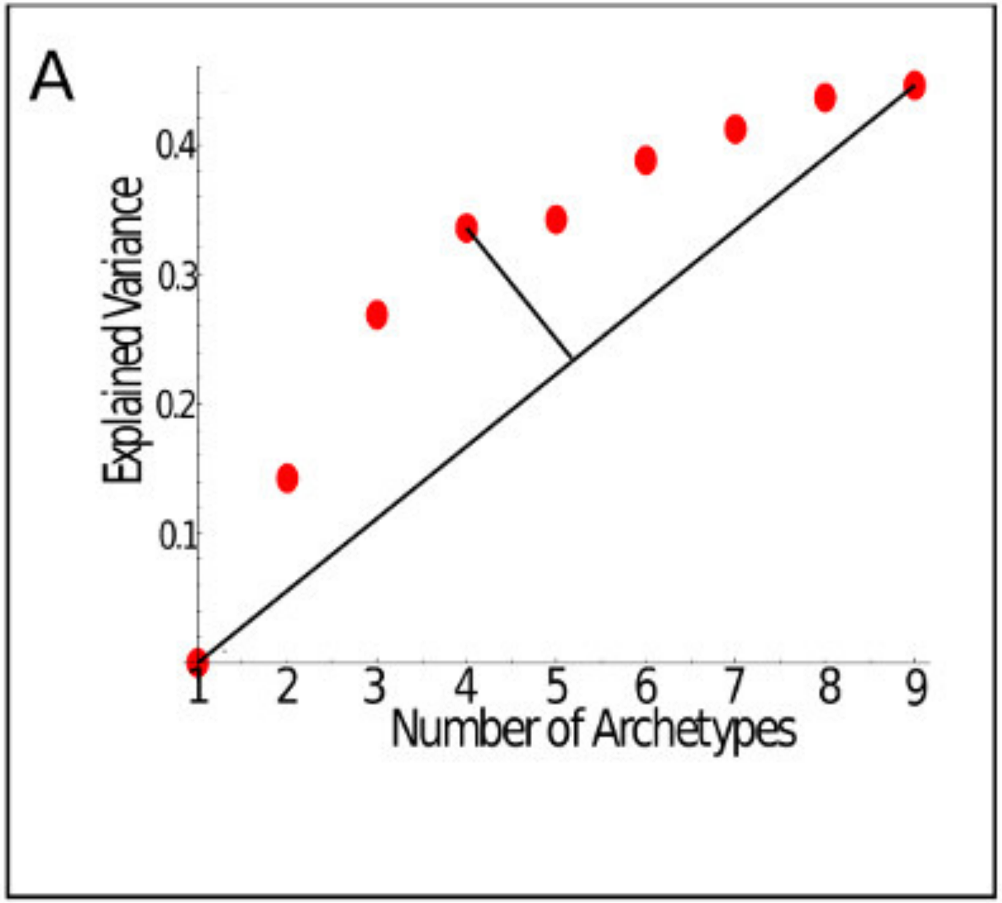




Karin and Alon (2018) BioRxiv

The first step is to determine the number of archetypes using PCHA and the elbow method

$$\mathrm{EV}(n) = \frac{1}{N}\sum_{i=1}^{N}(1 - \| \vec{p}_i - \vec{s}_i \| / \| \vec{p}_i \|)$$

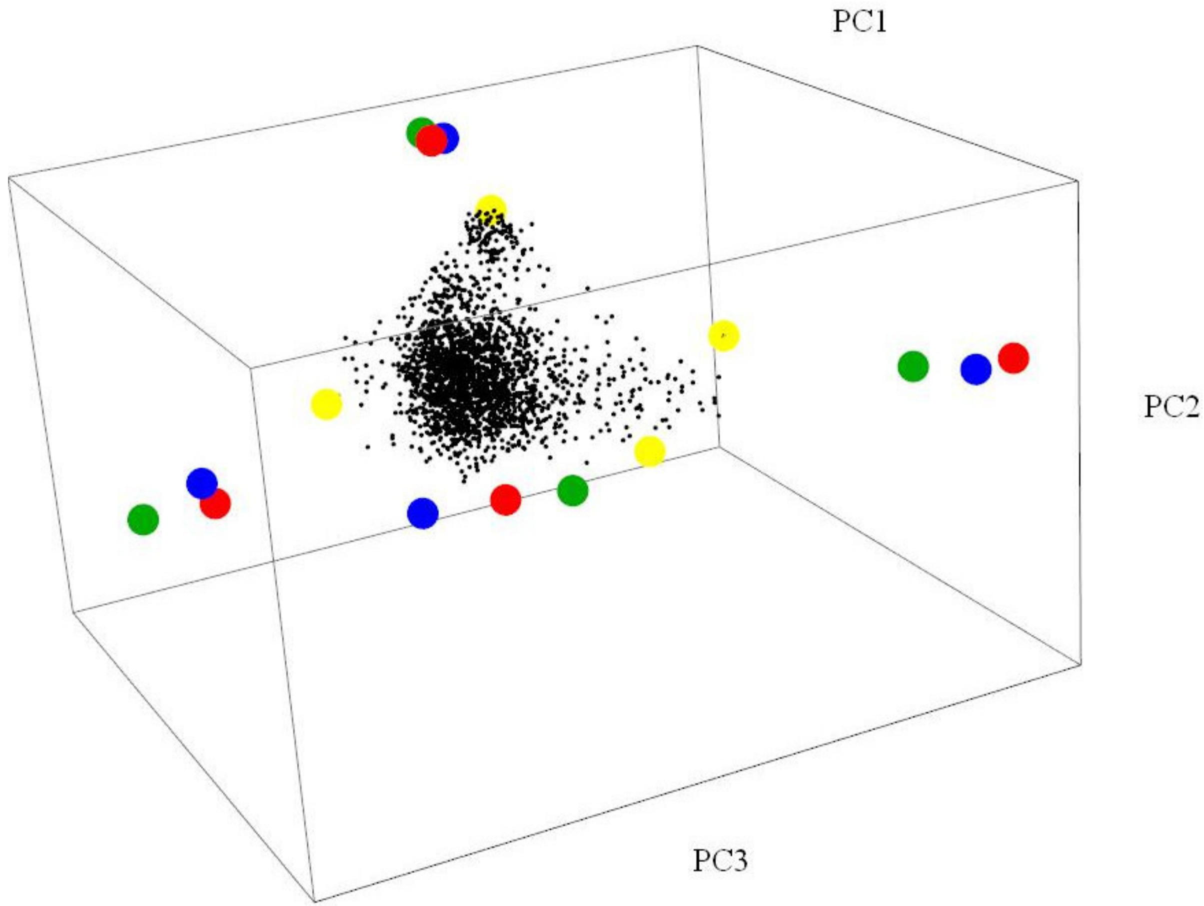


Principle Context Hull Analysis

Mørup & Hansen (2012) Neurocomputing

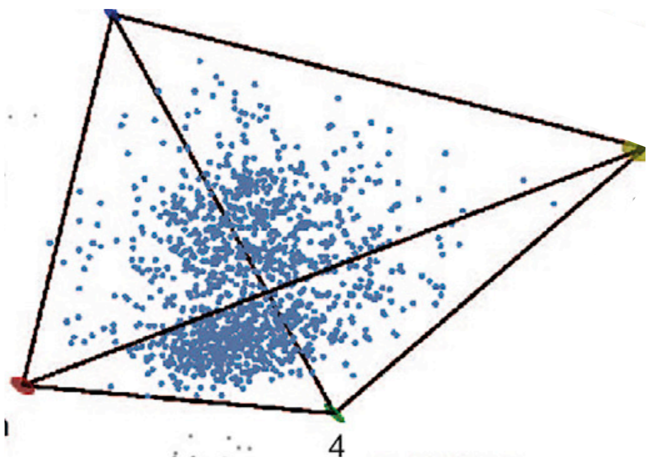
Hart et al (2015) Nat. Methods

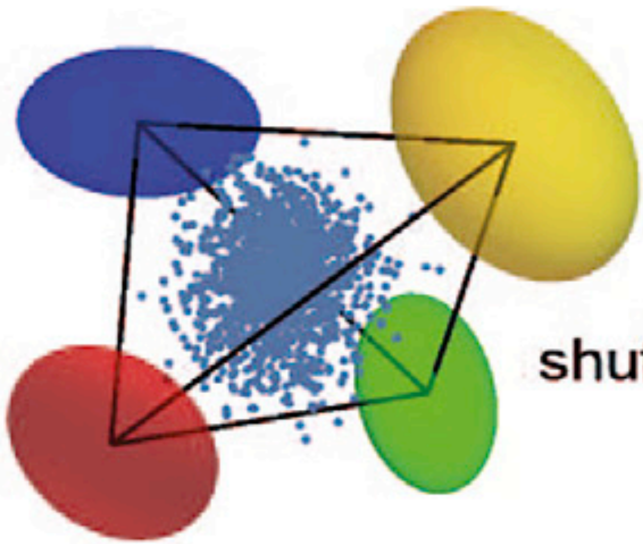
The second step is to estimate archetype positions using unmoving algorithms



Hart et al (2015) Nat. Methods

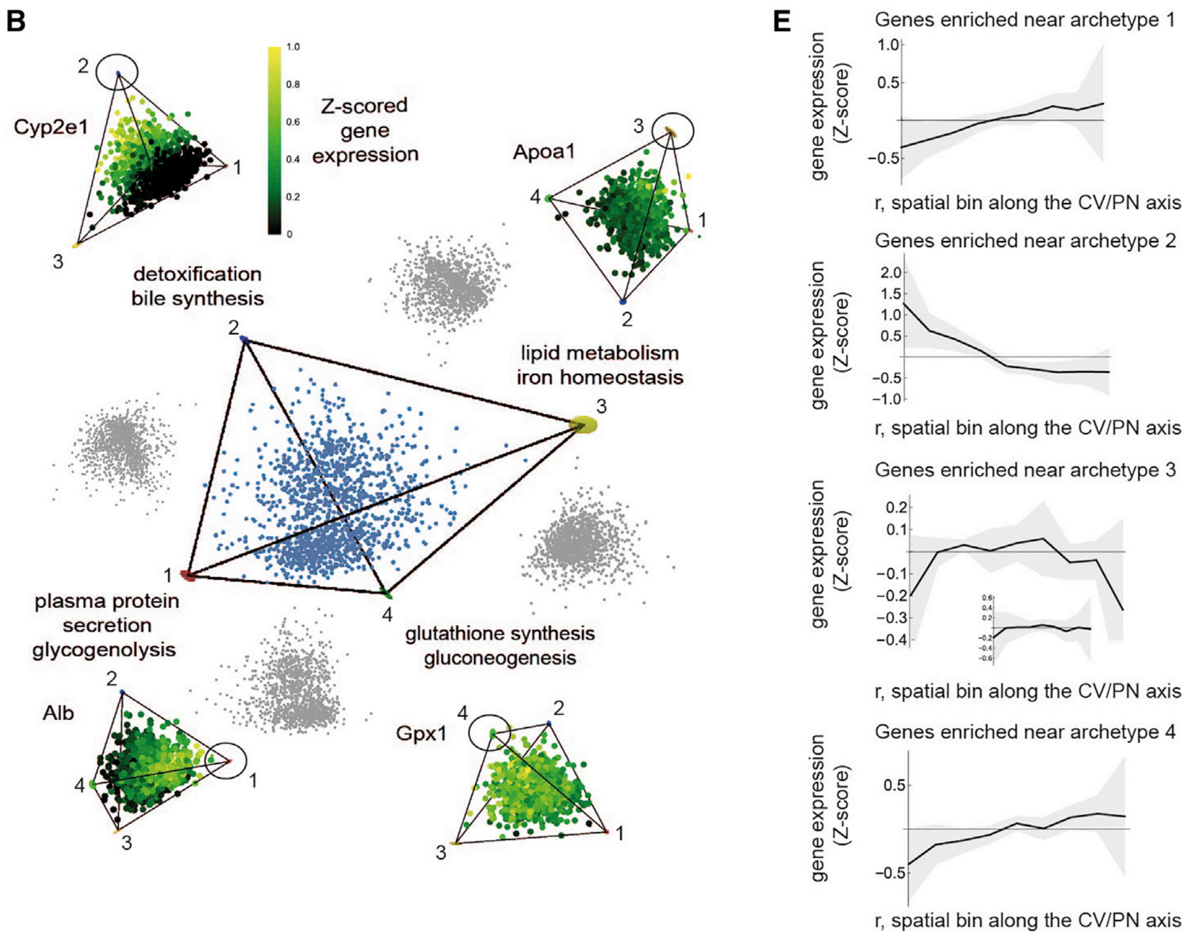
The third step is to evaluate significance of the best-fitting polytope by calculating convex hull of shuffled data





Hart et al (2015) Nat. Methods

Last step is to determine the enriched features of each archetype



Hart et al (2015) Nat. Methods

Adler et al (2019) Cell Systems