dynamic processes in cells (a systems approach to biology)

jeremy gunawardena
department of systems biology
harvard medical school

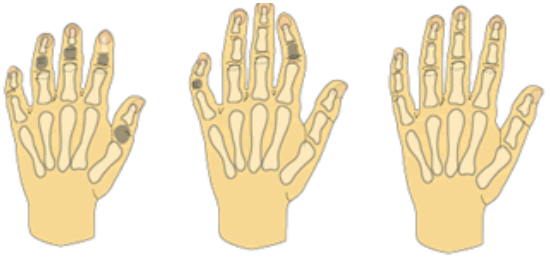
lecture 6

20 september 2016

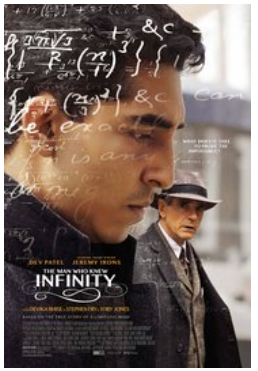
mathematical population genetics

how do genes (alleles) change over time in a population?

shortness of digits



A1 A4 normal



"To THE EDITOR OF SCIENCE: I am reluctant to intrude in a discussion concerning matters of which I have no expert knowledge, and I should have expected the very simple point which I wish to make to have been familiar to biologists. ... Mr. Yule is reported to have suggested, as a criticism of the Mendelian position, that if brachydactyly is dominant 'in the course of time one would expect, in the absence of counteracting factors, to get three brachydactylous persons to one normal'. It is not difficult to prove, however, that such an expectation would be quite groundless."

1877-1947

G H Hardy, "*Mendelian proportions in a mixed population*", Science, **28**:49-50 1908.

https://www.peds.ufl.edu/divisions/genetics/teaching/hand_malformations.htm

Hardy's calculation

two alleles (a, A) at a single locus

probabilities

genotypes $\mathrm{AA}(P) \quad \mathrm{Aa}(Q) \quad \mathrm{aa}(R) \qquad P+Q+R=1$

alleles $\mathrm{A}\left(P+\frac{Q}{2}\right) \qquad \mathrm{a}\left(R+\frac{Q}{2}\right)$

under random mating in an infinite population with non-overlapping generations, in the absence of selection, mutation, migration, etc, the next generation looks like

genotypes

$$\mathrm{AA} \qquad \mathrm{Aa} \qquad \mathrm{aa}$$

$$\left(P+\frac{Q}{2}\right)^2 \quad 2\left(P+\frac{Q}{2}\right)\left(R+\frac{Q}{2}\right) \quad \left(R+\frac{Q}{2}\right)^2$$

alleles $\mathrm{A}\left(P+\frac{Q}{2}\right) \qquad \mathrm{a}\left(R+\frac{Q}{2}\right)$ no change

the genotype frequencies become stable as soon as they satisfy $Q^2 = 4PR$ which happens after only a single generation

Hardy-Weinberg equilibrium

single locus, n alleles

alleles $A_1(p_1), A_2(p_2), \cdots, A_n(p_n)$

under random mating with no selection, mutation, migration, etc, the genotype frequencies become stable after one generation

their values are given by the respective terms in the expansion of

$$(p_1 + \cdots + p_n)^2$$

genotypes $A_iA_i(p_i^2)$ $A_iA_j(2p_ip_j)$

C Stern, "*The Hardy-Weinberg law*", Science, **97**:137-38 1943.

human polymorphisms at HW equilibrium

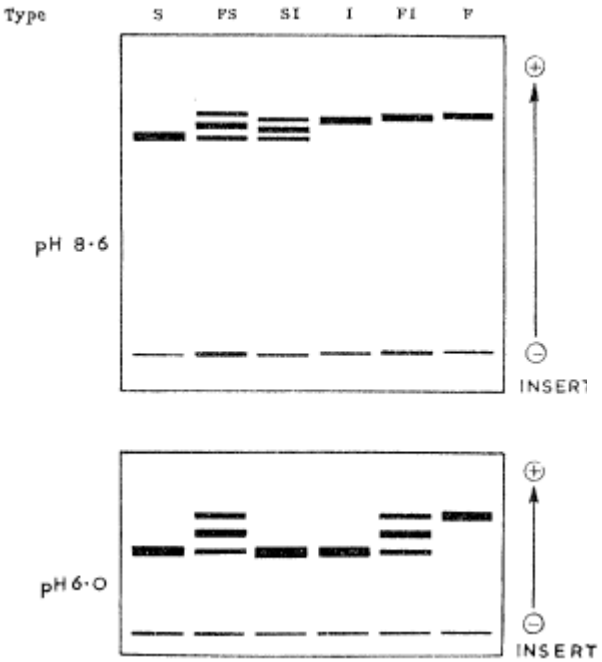


TABLE 9. OBSERVED AND EXPECTED NUMBERS OF PLACENTAL ALKALINE PHOSPHATASE TYPES IN A POPULATION SAMPLE ASSUMING A HARDY–WEINBERG EQUILIBRIUM **F I S**

($p = 0{\cdot}27$, $q = 0{\cdot}09$ and $r = 0{\cdot}64$)

placental alkaline phosphatase type	expected incidence		expected numbers in population sample	observed numbers in population sample
S	r^2	0·410	135·9	141
SF	$2pr$	0·346	114·7	111
F	p^2	0·073	24·2	28
SI	$2qr$	0·115	38·2	32
FI	$2pq$	0·049	16·1	15
I	q^2	0·008	2·7	5
totals	$(p+q+r)^2$	1·001	331·8	332

"Although one can hardly draw firm numerical conclusions from such a small series, it seems likely, unless we have been excessively lucky in our choice of enzymes, that polymorphism to a similar degree may be a fairly common phenomenon among the very large number of enzymes that occur in the human organism"

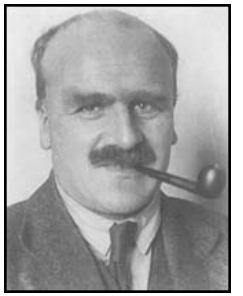
Henry Harris, "*Enzyme polymorphisms in man*", Proc Roy Soc Lond B, **164**:289-310 1966

the most successful theory in biology – I

how do allele frequencies change in populations under the influence of selection, mutation, migration, population structure, etc?


1890-1962


1889-1988


1892-1964

two alleles (a, A) at a single locus:

selection coefficient

dominance

genotypes	AA	Aa	aa
relative fitness	*1 - s*	*1 - hs*	*1*

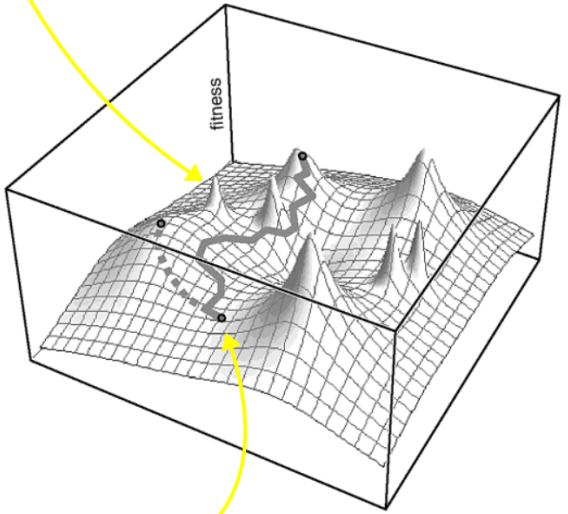
change in frequency (*p*) of allele A

$$\Delta p = \frac{pqs[ph + q(1-h)]}{1 - 2pqhs - q^2 s}$$

John Gillespie, **Population Genetics**, JHU Press, 2004; Sean Rice, **Evolutionary Theory: Mathematical and Conceptual Foundations**, Sinauer Associates, 2004

general mathematical conclusions

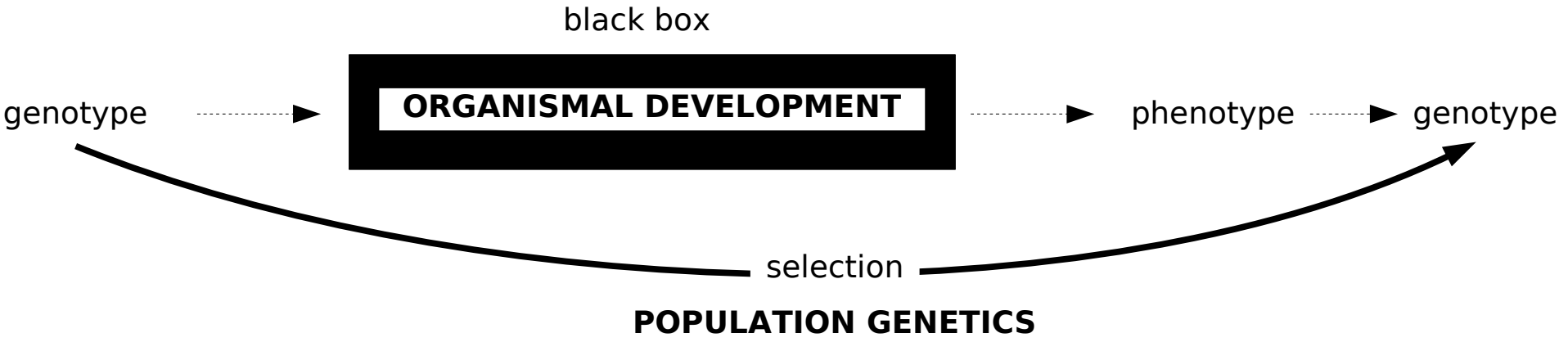
fisher – in larger populations, even weak positive selection can lead to an allele becoming fixed in the population, while deleterious alleles are quickly lost. average fitness increases monotonically to an optimum.



wright – in smaller populations, random processes (drift, mutation, recombination) can dominate over selection. deleterious alleles can become fixed, while beneficial ones can be lost. average fitness may decrease.

Eugene Koonin, **The Logic of Chance: The Nature and Origin of Biological Evolution**, FT Press Science, New Jersey, 2012

the price paid for success



- selection is assumed to act on the genotype, when it really acts on the phenotype – the organism is thereby treated as if it were a black box

out of sight, out of mind – the dark side of modelling

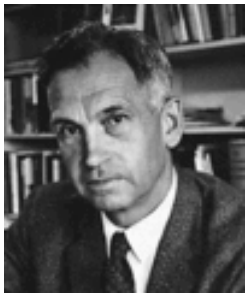
- selection appears as a parameter, not as a variable. population genetics does not tell us how selection arises. it only tells us how allele frequencies change.

- the molecular basis of the genotype was not known at the time and so the nature and extent of the genetic variation (ie: the initial conditions) could not be accurately specified

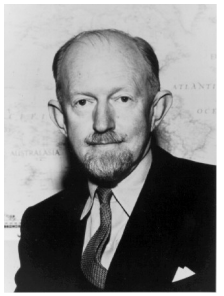
the modern ("neo-darwinian") synthesis

field biologists and paleontologists showed that variation in nature was consistent with laboratory genetics and with the predictions of population genetics

1900-1975



1904-2005



1902-1984

"Nothing in biology makes sense except in the light of evolution" (*) Dobzhansky

BUT –

- selection remained a parameter and compelling evidence for its presence was hard to find in natural populations
- the modern synthesis excluded the study of development

(*) Theodosius Dobzhansky, **Genetics and the Origin of Species**, Columbia University Press, 1937; Ernst Mayr, **Systematics and Origin of Species from the Viewpoint of a Zoologist**, Columbia Univ Press, 1942; George Gaylord Simpson, **Tempo and Mode in Evolution**, Columbia Univ Press, 1944

Julian Huxley, **Evolution: the Modern Synthesis**, Allen & Unwin, 1942

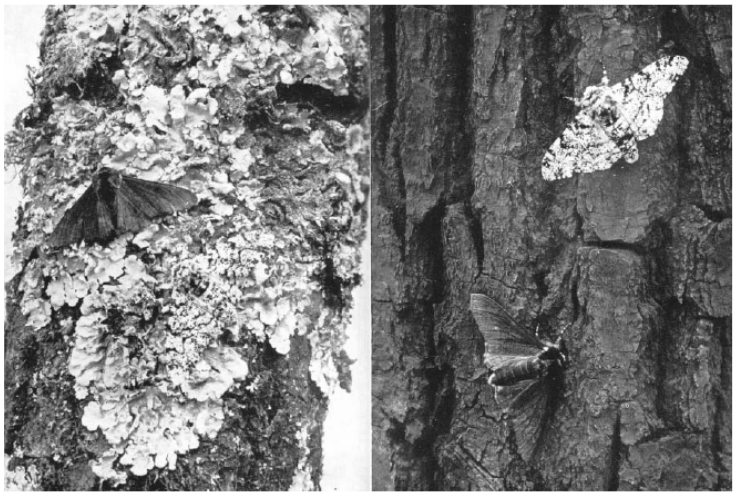
the triumph of selection

evidence for strong selection, arising from predation, came from field studies of animal behaviour in their natural ecological context

Edmund Brisco Ford, **Ecological Genetics**, Methuen & Co, 1964

1901-1988

example: bernard kettlewell's experiment on “industrial melanism” in the peppered moth *Biston betularia*



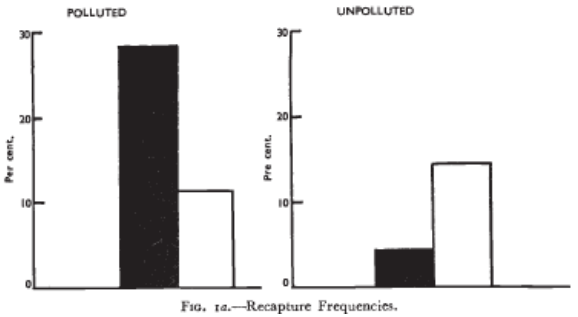
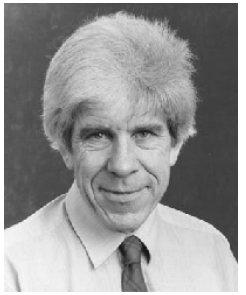


Fig. 1a.—Recapture Frequencies.

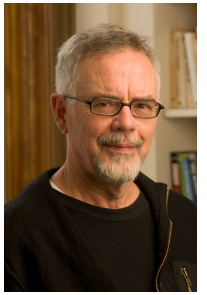
Kettlewell, “*Further selection experiments on industrial melanism in the Lepidoptera*”, Heredity, **10**:287-301 1956; Cook, Grant, Saccheri, Mallet, “*Selective bird predation on the peppered moth: the last experiment of Michael Majerus*”, Biol Lett **8**:609-12 2012; see also http://www.open.ac.uk/library/digital-archive/clip/clip%3Asci_clip14

the most successful theory in biology – II

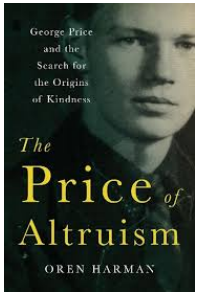
selection can exert an influence on an allele not only through an individual organism which carries that allele but also through other organisms which also carry that allele, especially those organisms which are closely related, such as parents, offspring, siblings, etc ("**kin selection**")



1936-2000



b. 1943

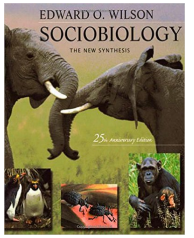


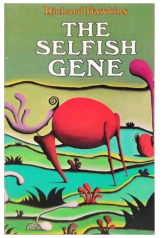
1922-1975

"I will give up my life for 2 brothers or 8 cousins" J B S Haldane

inclusive fitness – selection weighted by genetic relatedness

$$s = s_i + \sum_j r(i,j)s_j$$





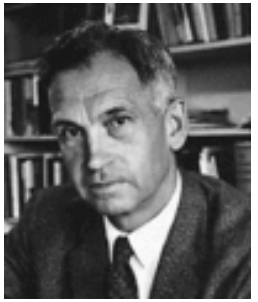
can explain social behaviour (conflict, cooperation) at many biological levels

- parasitic elements in the genome (ie: transposons)
- segregation distorters and meiotic drivers in gametes
- conflict (mother/father, mother/fetus) over parental investment in offspring
- reciprocal altruism (ie: warning cries)

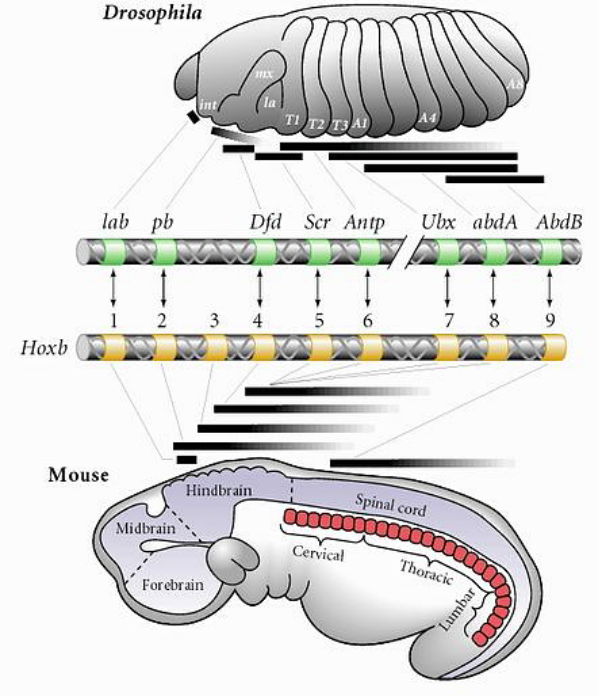
frequency-dependent selection – requires new tools like **game theory**

the revenge of the organism

"Much that has been learned about gene physiology makes it evident that the search for homologous genes is quite futile except in very close relatives."



Ernst Mayr, **Animal Species and Evolution**, Harvard University Press, 1963



on the contrary, there is deep conservation of certain genes and their protein functions between evolutionarily very distant organisms

Lutz, Lu, Eichele, Miller, Kaufman, "*Rescue of Drosophila labial null mutant by the chicken ortholog Hoxb-1 demonstrates that the function of Hox genes is phylogenetically conserved*", Genes & Dev **10**:176-84 1996

“evo-devo” or misunderstanding?

“Since the Modern Synthesis, most expositions of the evolutionary process have focused on microevolutionary mechanisms. Millions of biology students have been taught the view (from population genetics) that ‘evolution is change in gene frequencies.’ Isn’t that an inspiring theme? This view forces the explanation toward mathematics and abstract descriptions of genes, and away from butterflies and zebras ... The evolution of form is the main drama of life’s story, both as found in the fossil record and in the diversity of living species. So, let’s teach that story. Instead of ‘change in gene frequencies,’ let’s try ‘evolution of form is change in development’.”

Sean Carroll, **Endless Forms Most Beautiful: The New Science of Evo-Devo**, W W Norton & Co, 2005

“Even ignoring the fact that most species are unicellular and differentiated mainly by metabolic features, this statement illustrates two fundamental misunderstandings. Evolutionary biology is not a story-telling exercise, and the goal of population genetics is not to be inspiring, but to be explanatory. ... Nothing in evolution makes sense except in the light of population genetics.”

Michael Lynch, “*The frailty of adaptive hypotheses for the origins of organismal complexity*”, PNAS **104**:8597-604 2007, commenting on Sean Carroll's comment

the evolution of complexity revisited

the evidence from developmental biology suggests, in contrast to population genetics, that the “black box” of the organism must have specific properties in order for selection to act and for complexity to evolve

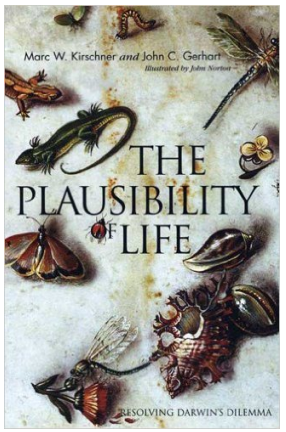
weak linkage – biological processes are able to interact without having to know too much about each other

this facilitates evolvability and the emergence of complexity

The theory of facilitated variation

John Gerhart*† and Marc Kirschner‡

PNAS **104**:8582-9 2007



INTERVIEW Open Access

Beyond Darwin: evolvability and the generation of novelty

Marc Kirschner

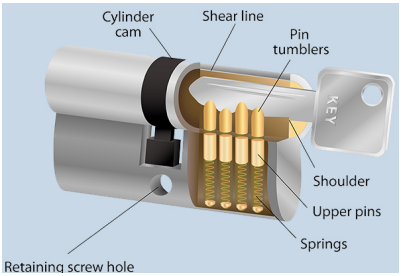
Kirschner M *BMC Biology* 2013, **11**:110

BMC Biology **11**:110 2013

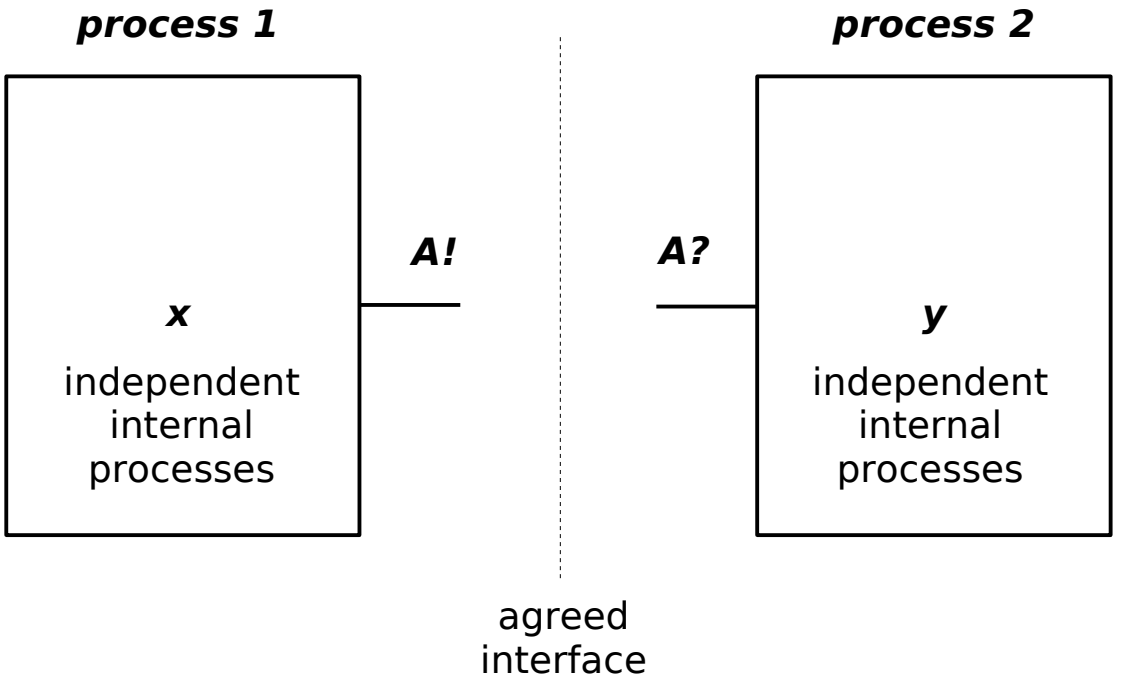
but how has weak linkage evolved ???

engineering uses strong linkage

process interaction takes place through agreed interfaces

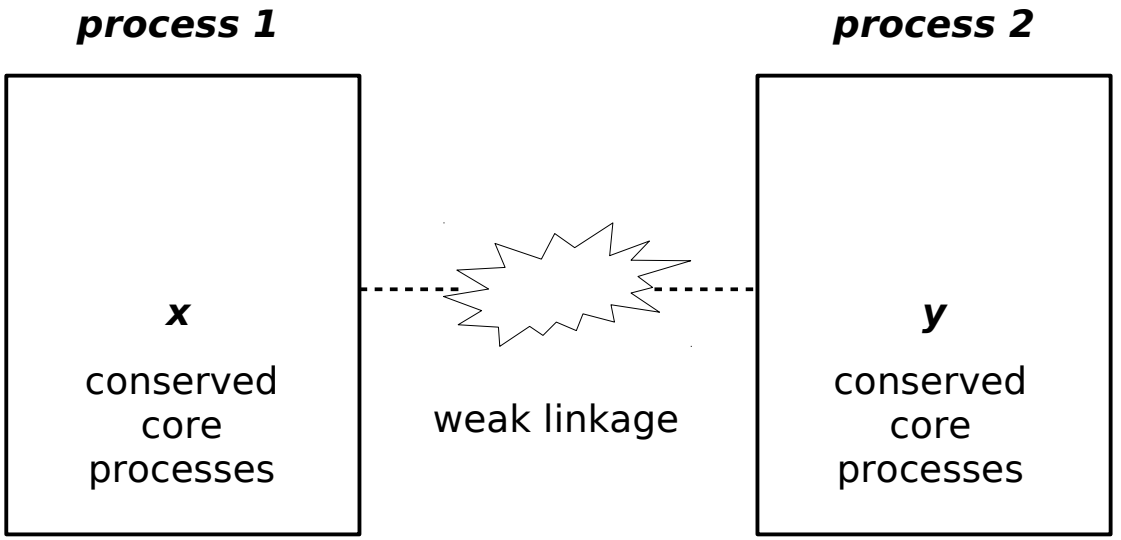


lock & key



constrained interfaces allow de-constrained innovation within modules

biology uses weak linkage



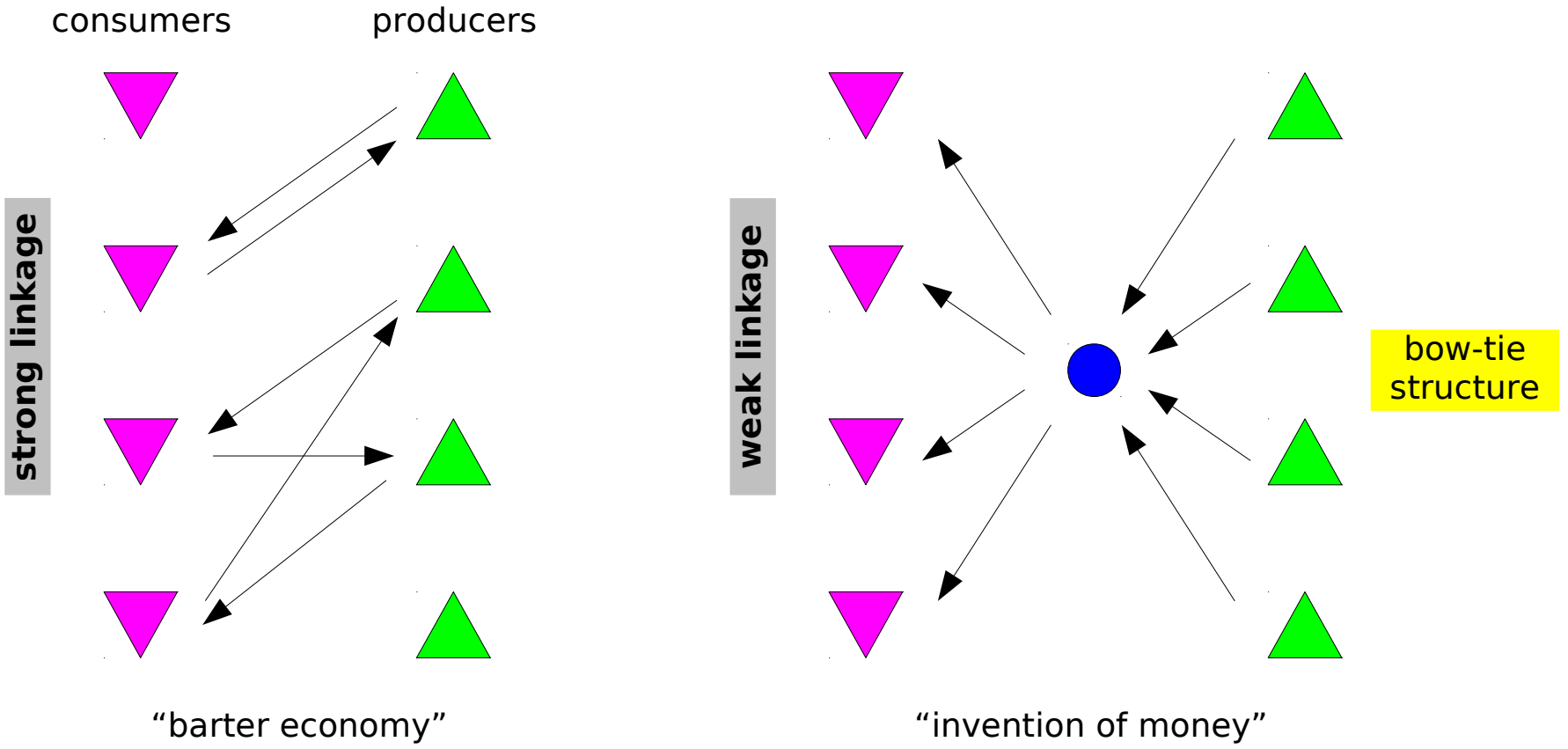
de-constrained interfaces allow conserved core components to evolve new functionality within processes while maintaining overall system behaviour

weak linkage through intermediation

metabolism – biological polymer construction requires dehydrating condensations

$$X\text{-}OH + Y\text{-}H \longrightarrow X\text{-}Y + H_20$$

which consume energy and must be coupled to energy-producing reactions



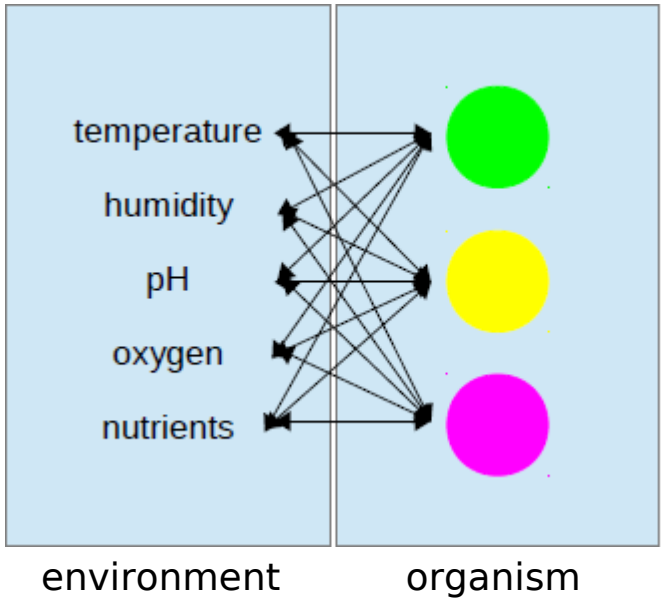
Csete, Doyle, “*Bow ties, metabolism and disease*”, Trends Biotechnol **22**:446-50 2004

weak linkage through intermediation

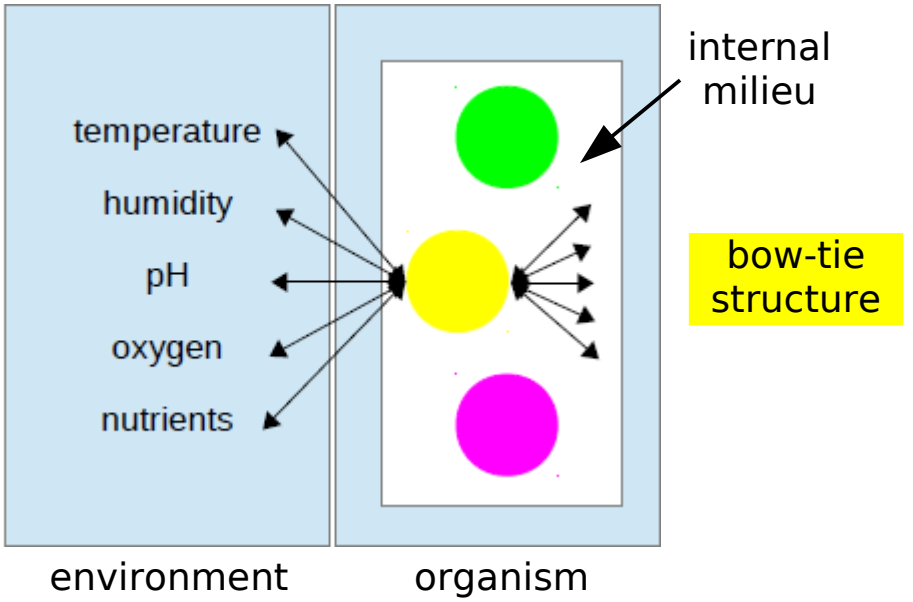
claude bernard's internal milieu

"The stability of the internal environment is the condition for the free and independent life."

strong linkage



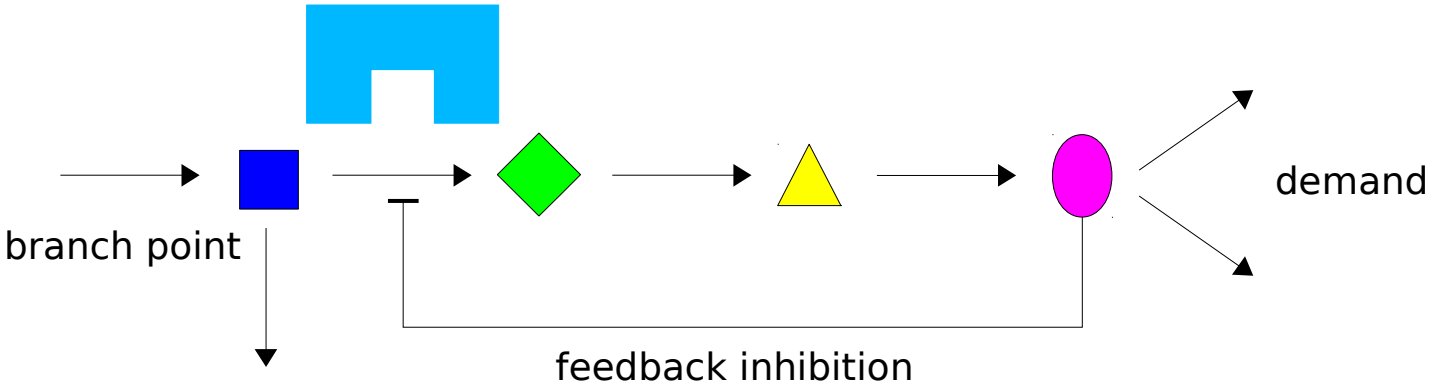
weak linkage



Claude Bernard, from **Lectures on the Phenomena Common to Animals and Plants**, 1878. Quoted in C Gross, *"Claude Bernard and the constancy of the internal environment"*, The Neuroscientist, **4**:380-5 1998

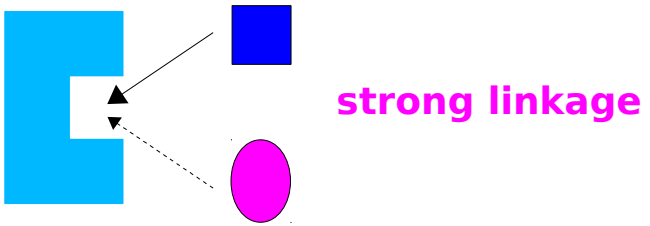
weak linkage in molecular regulation – allostery

the enzyme catalysing the first committed step in a biosynthetic pathway is usually inhibited by the terminal metabolite in the pathway



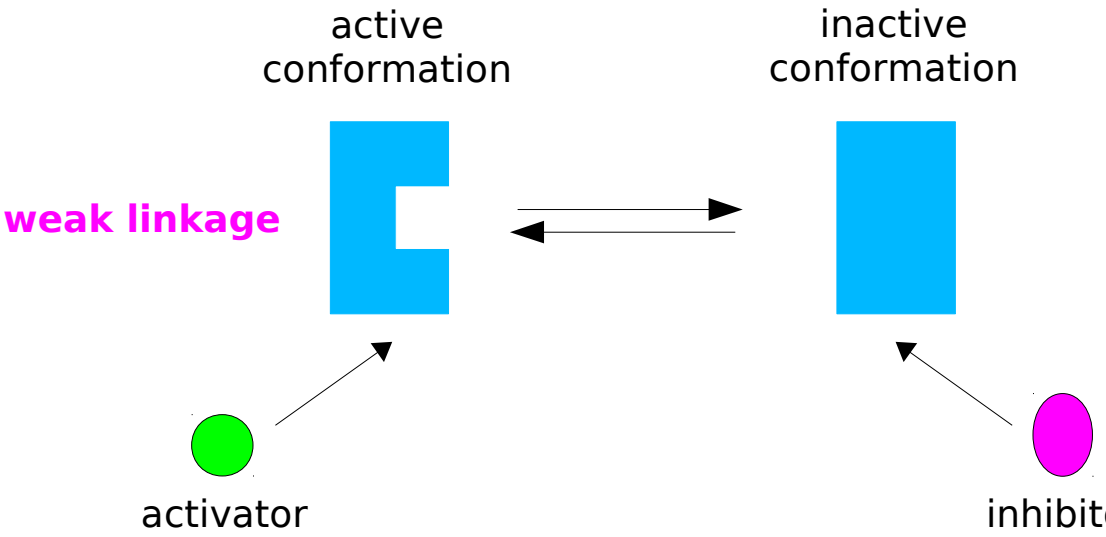
Pardee, Yates, *"Control of pyrimidine biosynthesis in Escherichia coli by a feed-back mechanism"*, J Biol Chem **221**:757-70 1956; Umbarger, *"Evidence for a negative-feedback mechanism in the biosynthesis of isoleucine"*, Science **123**:848 1956

but the terminal metabolite can be structurally very different from the enzyme's substrate, so that competitive inhibition at the catalytic site is no longer feasible

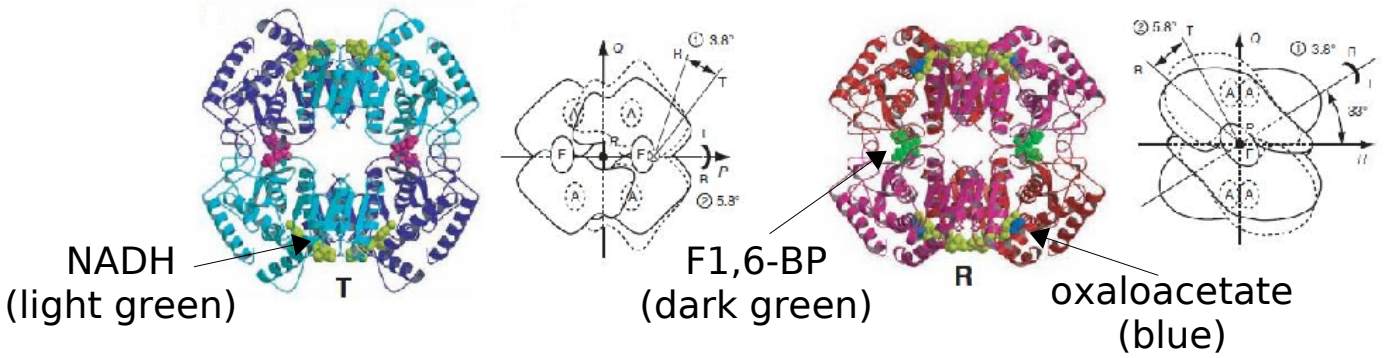


weak linkage in molecular regulation – allostery

protein allostery – exploit existing molecular states (conformations) having distinct functional properties



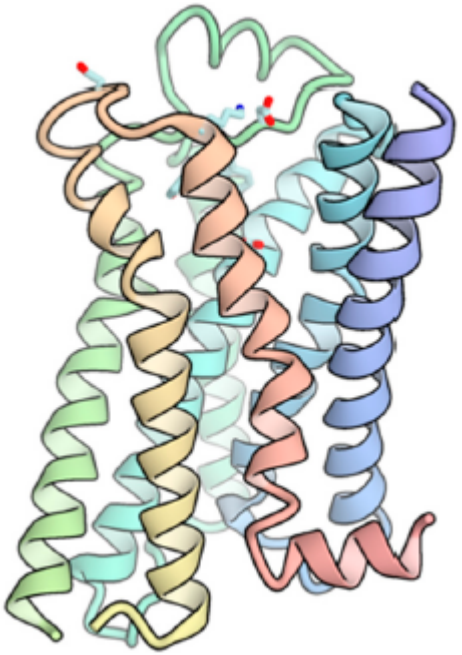
lactate dehydrogenase

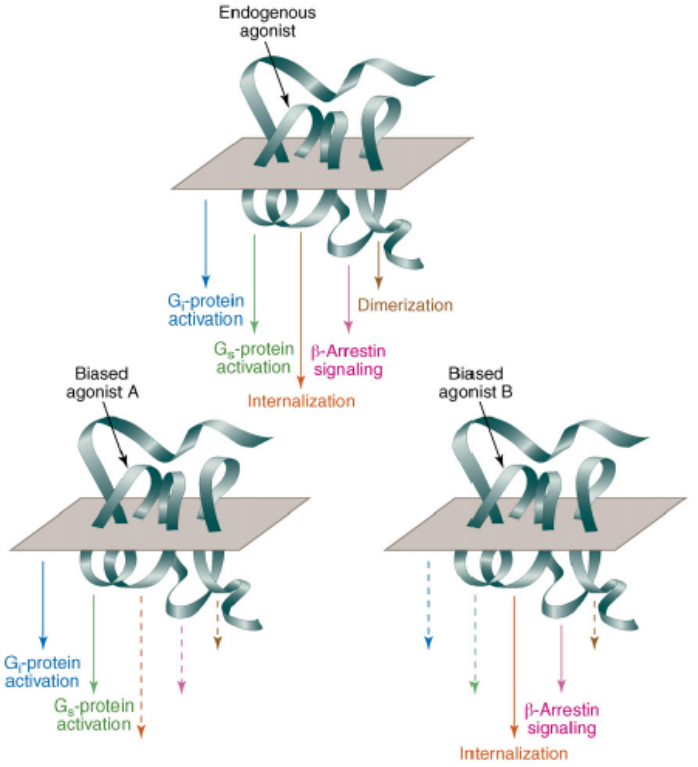


Monod, Changeux, Jacob, "*Allosteric proteins and cellular control systems*", J Mol Biol **6**:306-29 1963

weak linkage in molecular regulation – allostery

G-protein coupled (7 transmembrane) receptors





combinatorial output

"biased agonism" or "collateral efficacy" or "functional selectivity"

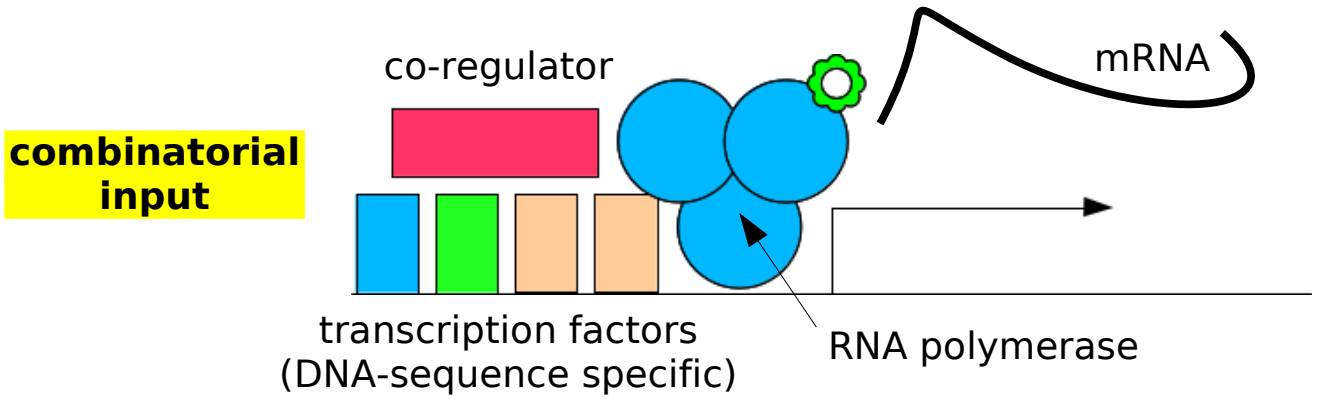
Kenakin, "*Collateral efficacy in drug discovery: taking advantage of the good (allosteric) nature of 7TM receptors*", Trends Pharmacol Sci **28**:407-15 2007

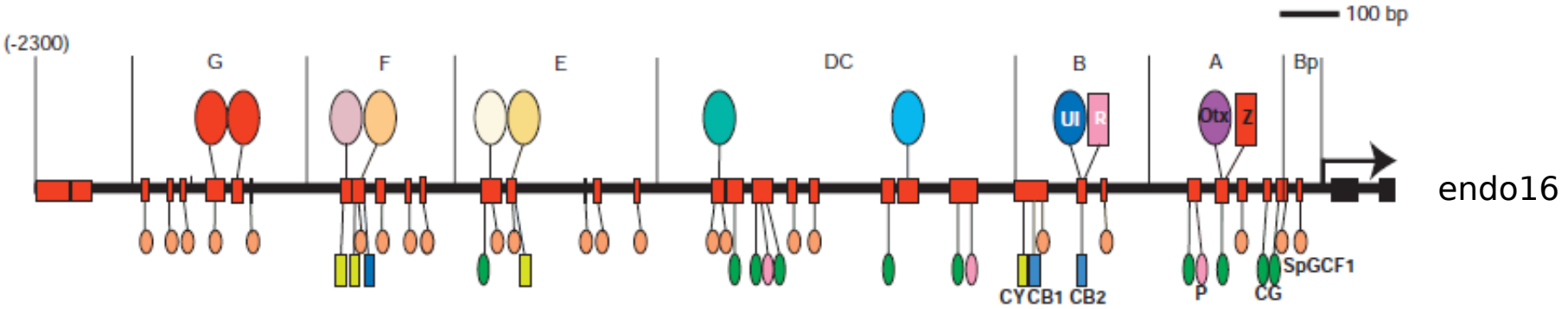
weak linkage in molecular regulation – gene regulation

construct new molecular states having distinct functional properties

gene regulation – how to regulate the catalytic production of mRNA from DNA?

indirect binding to DNA and recruitment of factors that activate/inhibit transcription





Yuh, Bolouri, Davidson, *"cis-regulatory logic in the endo-16 gene"*, Development **128**:617-29 2001