Institut Jean Roget

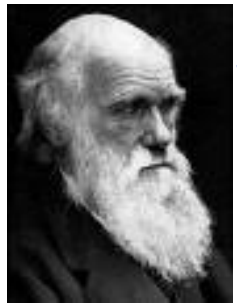
UNIVERSITE JOSEPH FOURIER
SCIENCES.TECHNOLOGIE.SANTÉ
GRENOBLE 1

cnrs
dépasser les frontières

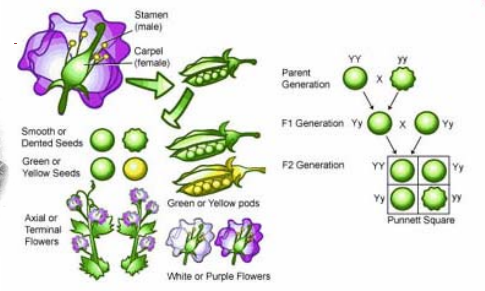
Laboratoire Adaptation et Pathogénie des Microorganismes, UMR5163, Grenoble

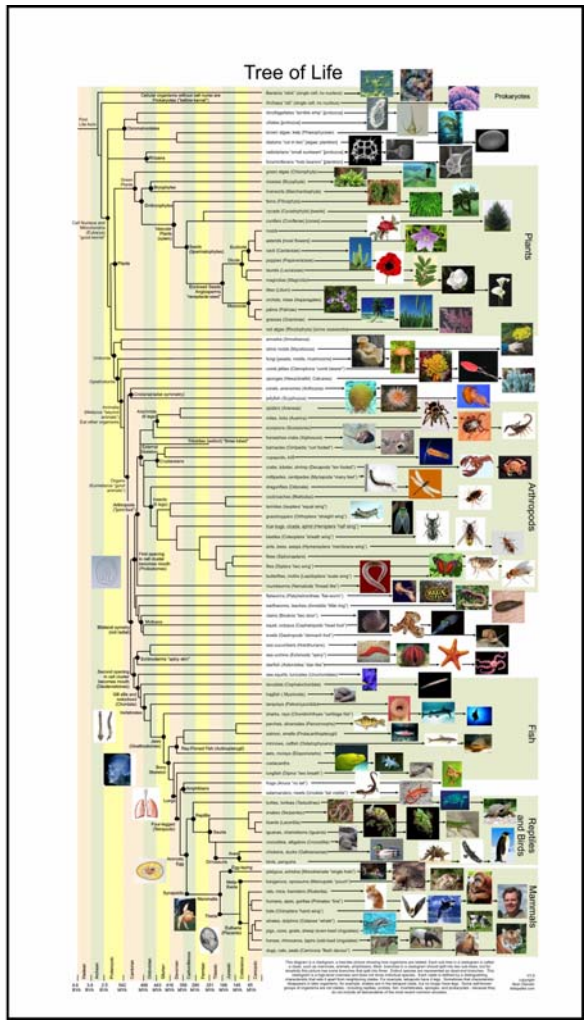
Long-term phenotypic and genomic evolution in *Escherichia coli*

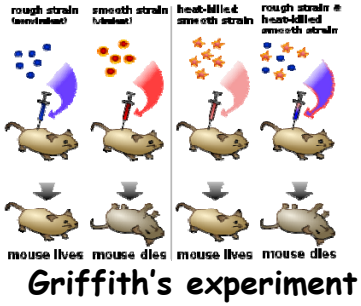
Evolution theory



"… from so simple a beginning endless forms most beautiful and most wonderful have been, and are being evolved."
On the Origin of Species, C. Darwin, 1859.







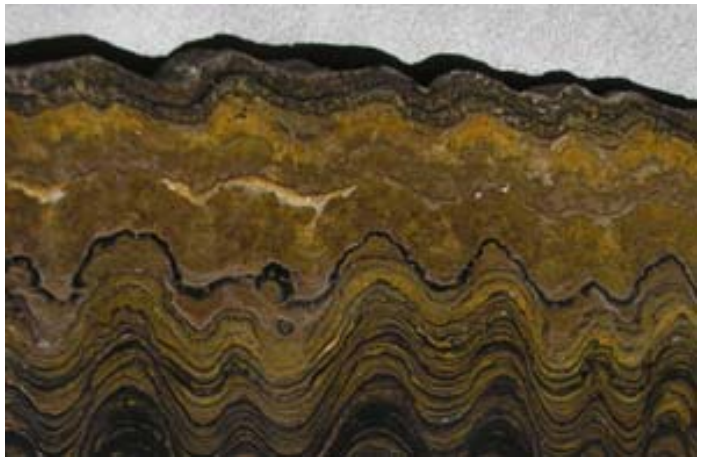
Griffith's experiment

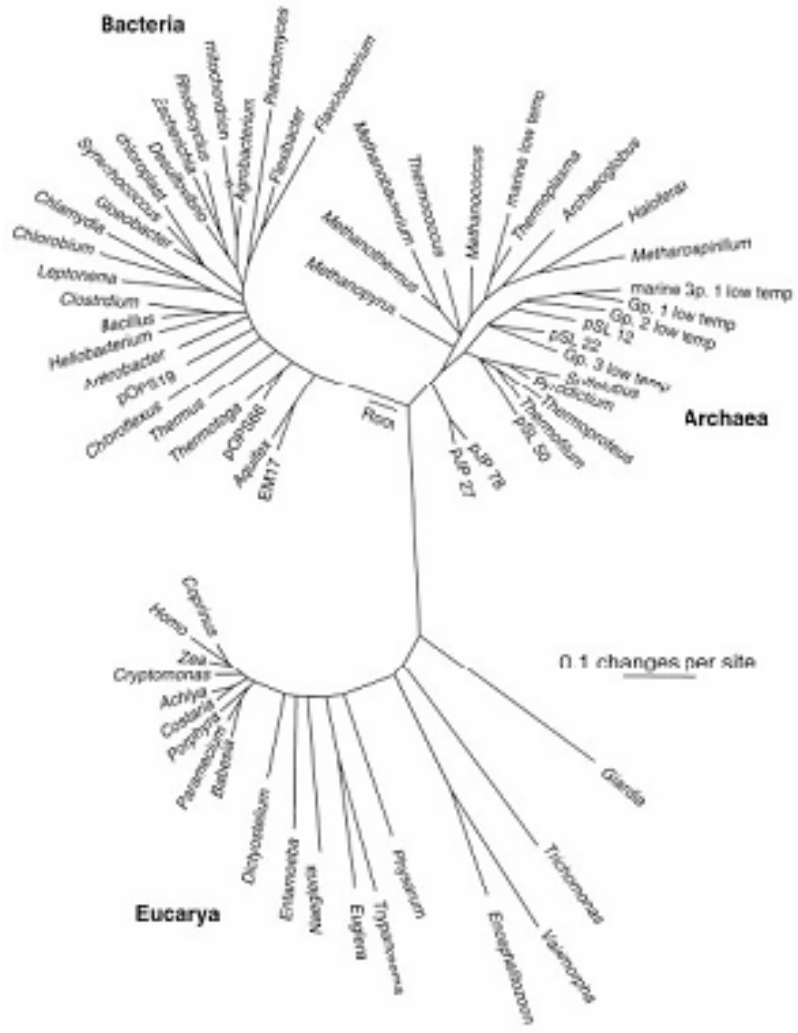
Mouche mutante

Mouche normale

Approaches to studying evolution

Paleontology ... comparative method



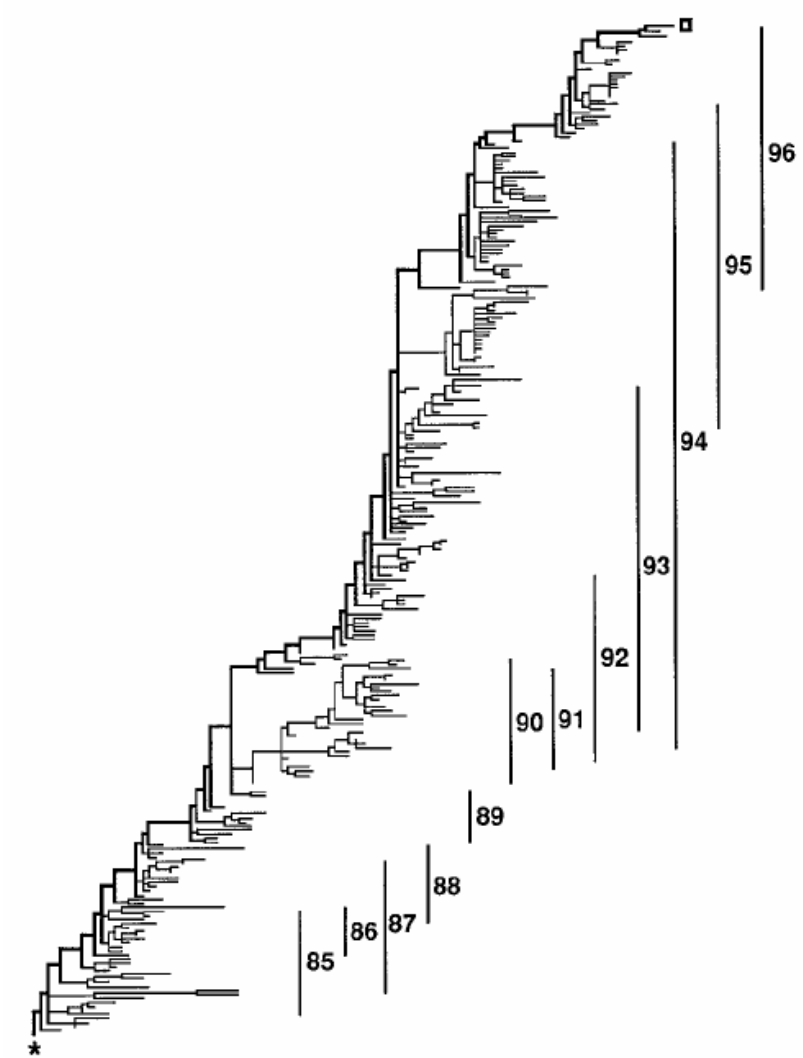


… evolution in action

Great Dane and Chihuahua

Papillon dog

http://www.pbs.org/wgbh/nova/dogs

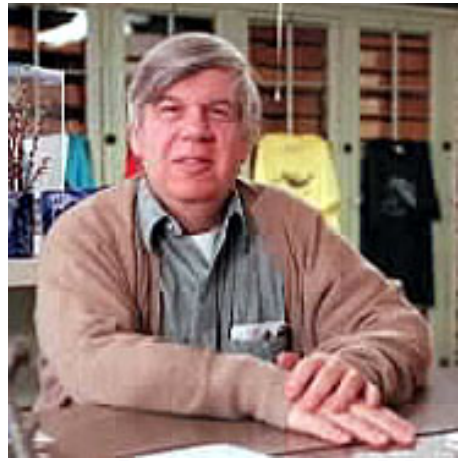


HA1 gene

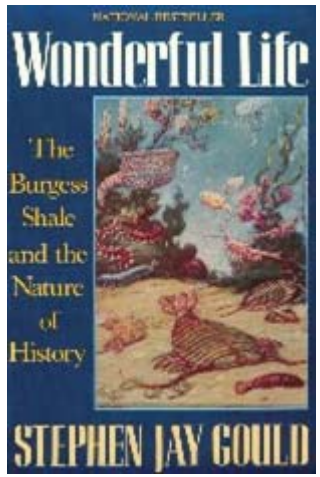
Fitch *et al.*, 1997

Question: How repeatable is evolution?

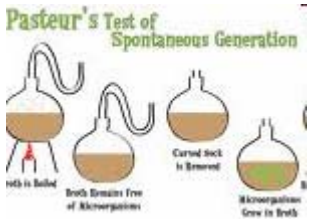
"I call this experiment '**replaying life's tape**.' You press the rewind button and, making sure you thoroughly erase everything that actually happened, go back to any time and place in the past. Then let the tape run again and see if the repetition looks at all like the original ... **The bad news is that we can't possibly perform the experiment.**"

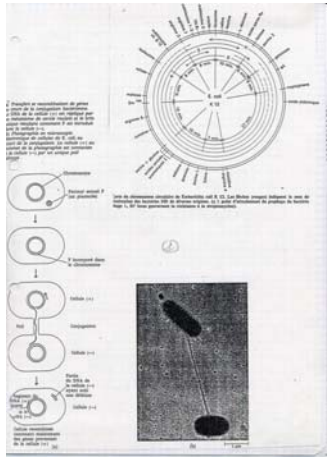


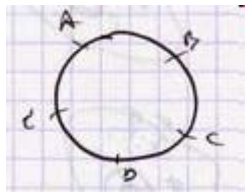
Stephen Jay Gould,
1989, *Wonderful Life*

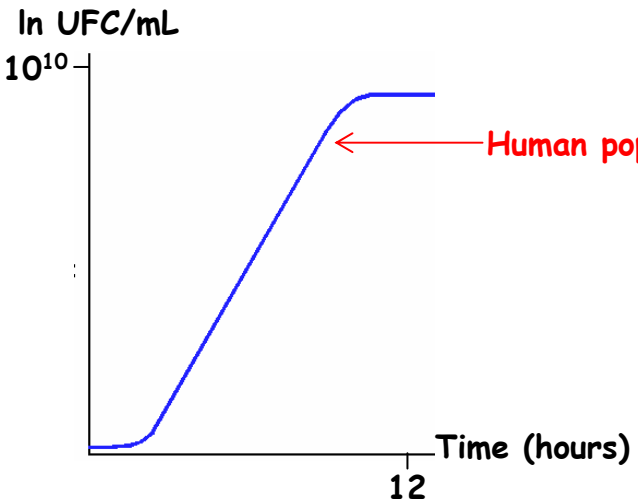


Evolution theory: use of microorganisms





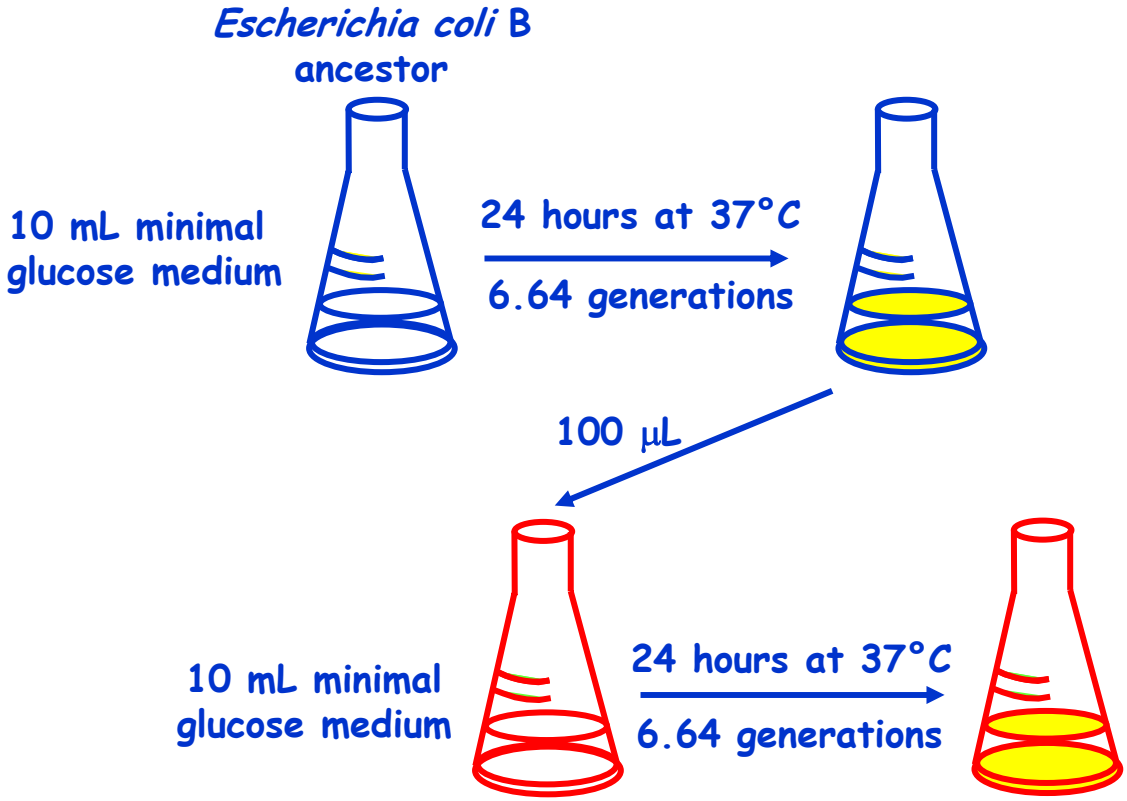


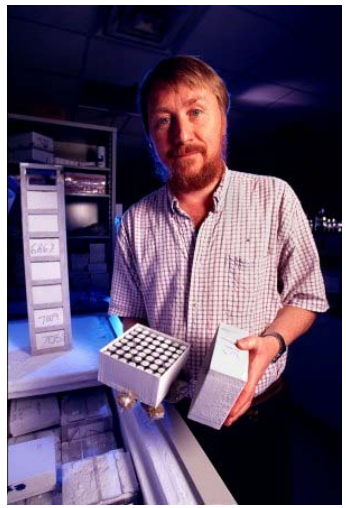


Extremely rapid growth.
Gigantic populations.

Storage at -80°C.
Revivification.

Long-term experimental evolution using *Escherichia coli*.





Richard Lenski
Michigan State Univ.

1988…

E. coli B ancestor

…2010

>50,000 generations

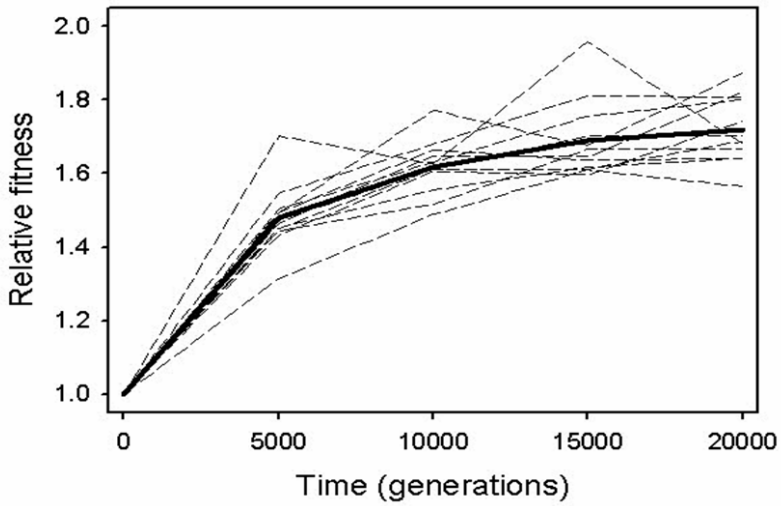
A billion mutations in each evolving population

- Effective population size adjusted for serial dilution — $N = 3 \times 10^7$
- Number of generations — $G = 4 \times 10^4$
- Base-pair mutation rate — $M = 4 \times 10^{-10}$
- Genome size, base pairs — $S = 5 \times 10^6$
- Total effective number of mutations per population — $N \times G \times M \times S > 10^9$

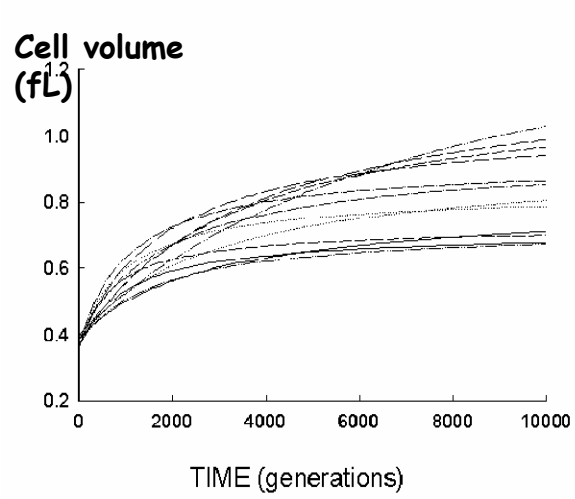
How many mutations?

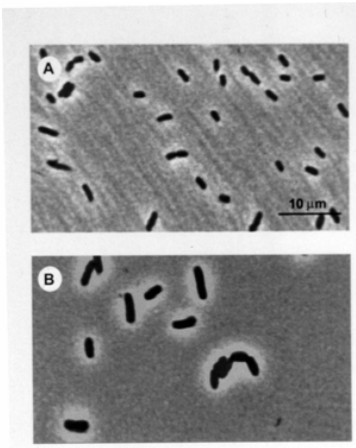
Phenotypic changes

Parallel fitness increases: decelerating rates of improvement

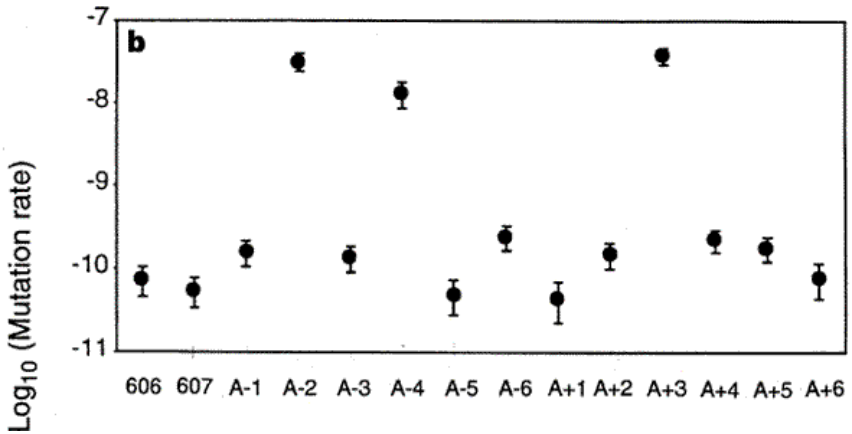


Parallel cell volume increases: decelerating rates of improvement

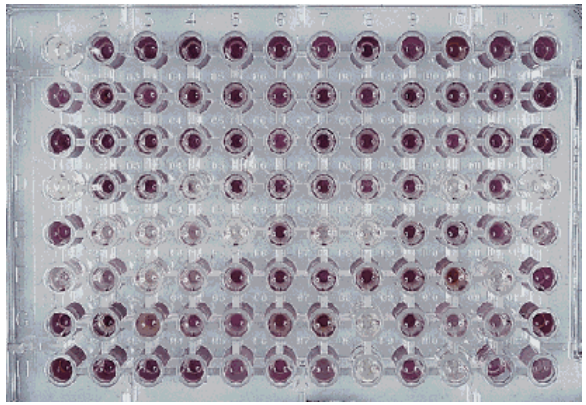




Increases in mutation rates



Decrease of total catabolic activity



Genetic changes

How?

- Sequence 36 random genes → almost none.
- RLFP analysis using native IS elements as probes.

Mutations in *pbpA*, *nadR*, *pykF*, *hok* in one population.

Sequencing in all 12 populations: genetic parallelism.

Non-synonymous mutations.

PNAS 1999, 2006
Genetics 2000

Genetic changes and fitness

- Sequenced ~500 bp for each of 36 genes in 4 clones from all 12 lines plus the ancestor.
  - **27** genes: no mutations in any evolved clones.
  - **9** genes: mutations in clones from a single population; in all cases, mutators.
  - **No** gene has mutations in two or more populations.
  - Almost all synonymous.

→ Cannot explain the parallel changes as neutral mutations substituted by drift or hitchhiking.

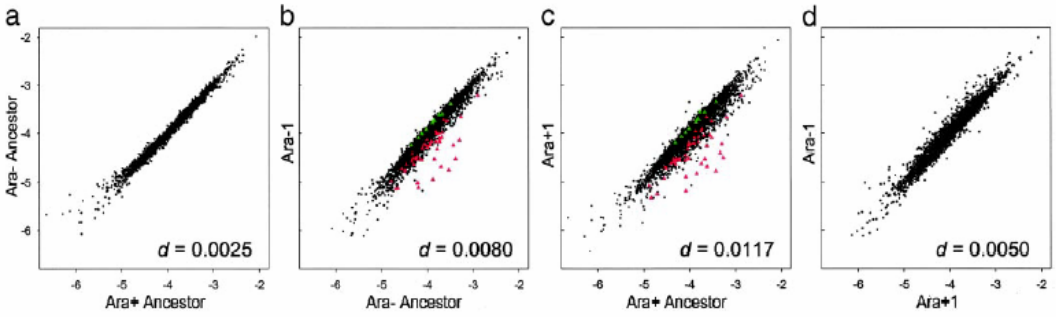
→ Instead, these changes are adaptations produced by natural selection.

PNAS 2006

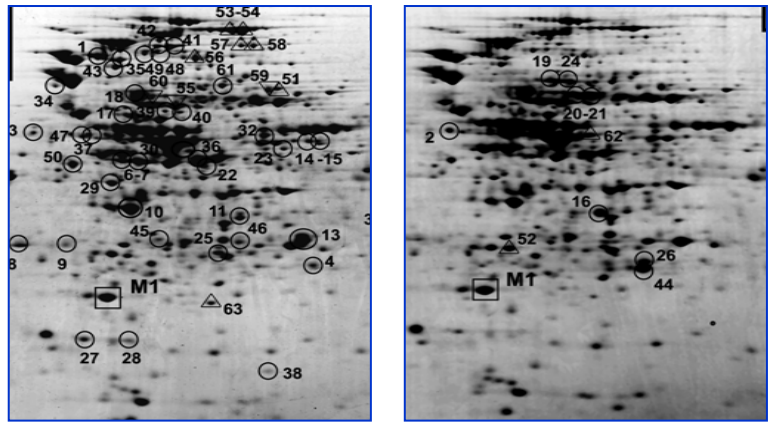
Genetic changes

How?

- Global expression profiles: ancestor versus two independent evolved clones from 20,000 generations



DNA arrays



Proteomics

Look for changes in both evolved clones and in the same direction:

- ppGpp regulon (stringent response): mutation in *spoT*
- loss of ribose genes and proteins: deletions of *rbs*
- loss of maltose genes and proteins: mutations in *malT*.

Fitness: beneficial alleles in the focal population.

Genetic parallelism:

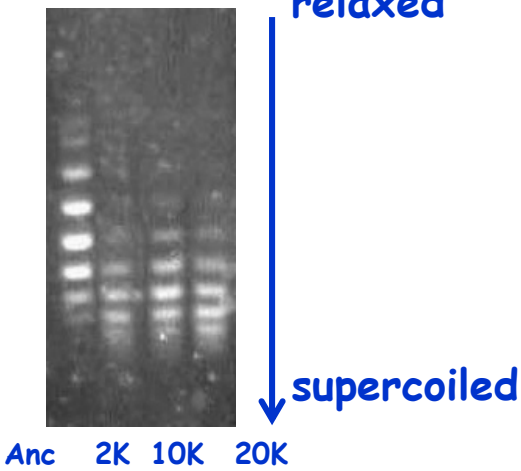
- *spoT*: 8/12 populations
- *rbs*: 12/12 populations
- *malT*: 8/12 populations.

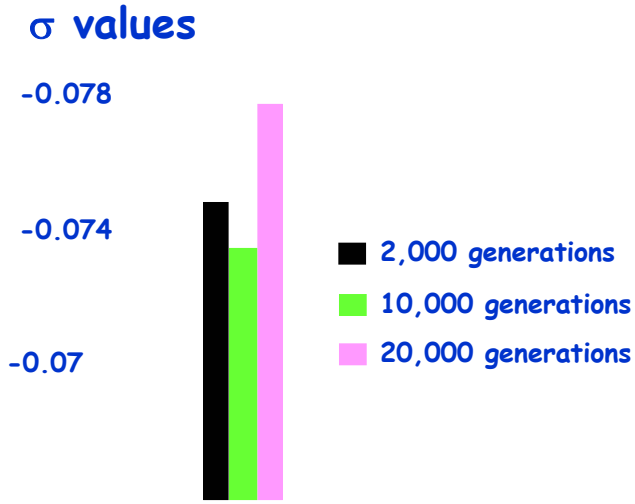
J. Bacteriol. 2001
Genetics 2006
PNAS 2003

Genetic changes

How?

- Analyses of DNA superhelicity: increases in all 12 populations
- Model population: Ara-1, with two successive changes





Two mutations:

- *topA*, topoisomerase I
- *fis*, histone-like proteins.

Introduction of each mutation, alone and in combination, in the ancestral chromosome:

- responsible for all DNA superhelicity changes
- beneficial mutations (ancestral background).

Genetics 2005

Genetic changes

Sequencing of 17 genes known to be involved in DNA superhelicity control in all 12 populations.
Comparison with 36 control genes.

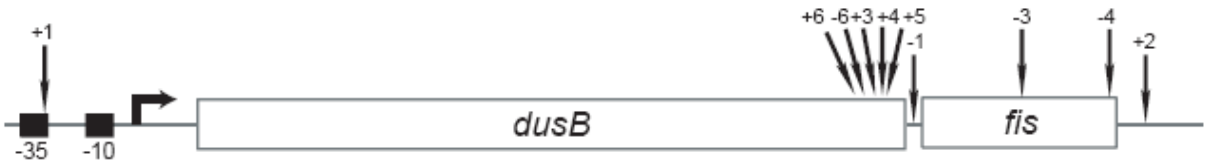
Control genes:

- 27 genes: no mutations in any evolved clones.
- 9 genes: mutations in clones from a single population;
  in all cases, mutators.
- No gene has mutations in two or more populations.
- Almost all mutations are synonymous.

17 superhelicity genes:

- 14: same profiles as control genes
- 3: mutations in almost all 12 populations, mutators and non-mutators, precisely one mutation in each of the 3 in each population, non-syn.

topA, *fis*, *dusB*: likely to be beneficial.



Mol. Biol. Evol. 2010.

Parallel changes in gene regulatory networks

Stringent response: adaptation to nutrient starvation *(spoT)*.

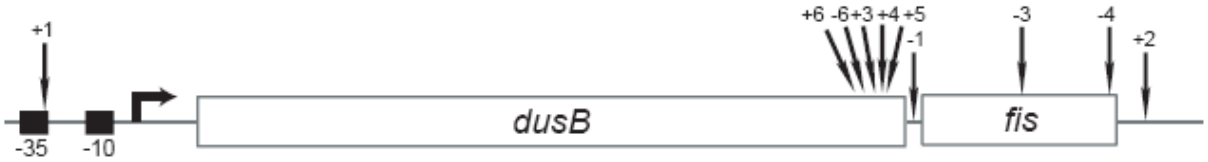
DNA topology (topoisomerase I, histone-like protein Fis).

Central metabolism (*pykF*, *nadR*).

Identification of a new global regulator: DusB = dihydrouridine synthase

Genetic parallel changes versus allelic variation?

Example of *fis*.



Ara-1: mutation in RBS
Ara+1: mutation in -35 promoter box
Ara+2: mutation in terminator
Ara-3, -4: mutations in coding sequence
Ara-6, +3, +4, +5, +6: mutations in *dusB*.

Mol. Biol. Evol. 2010.

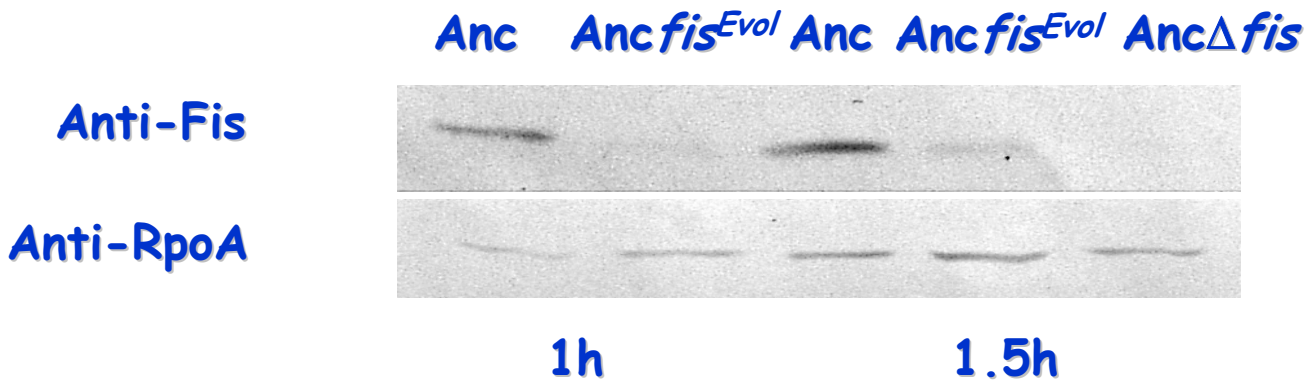
Genetic parallel changes versus allelic variation?

Ara-1 allele (RBS):

C

5'…AUAAAGAGCUGACAGAACUAUGUUC…3' *fis* 5'UTR

3'AUUCCUCCACUAGGUUGG…5' 16S rRNA

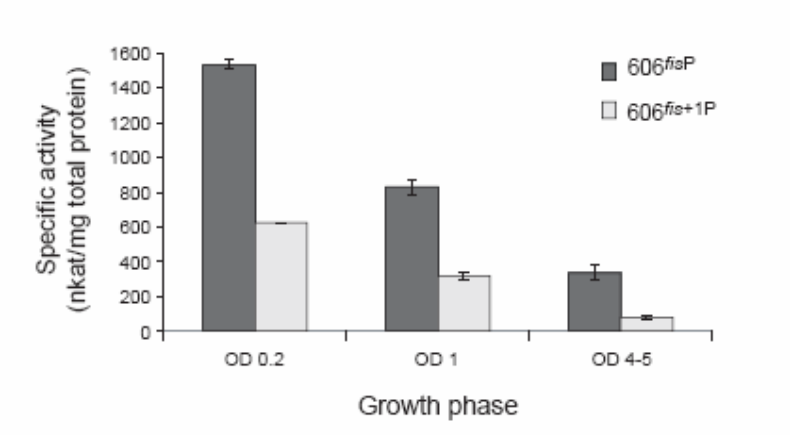


3 fold decrease of the quantity of Fis.

Genetics. 2005.

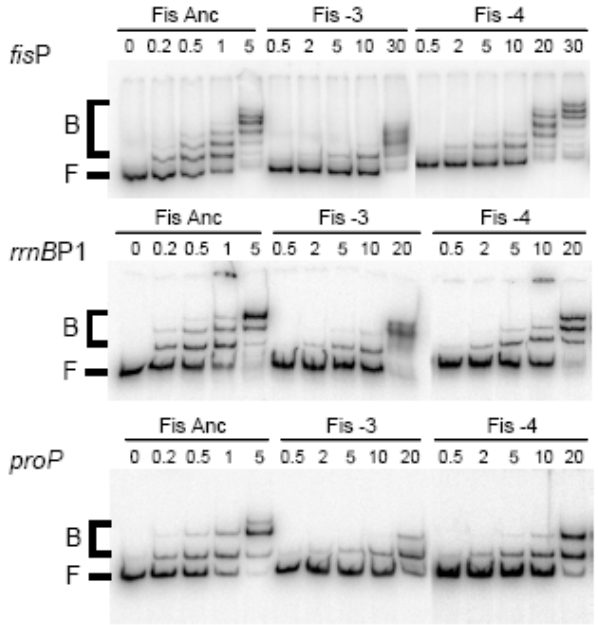
Genetic parallel changes versus allelic variation?

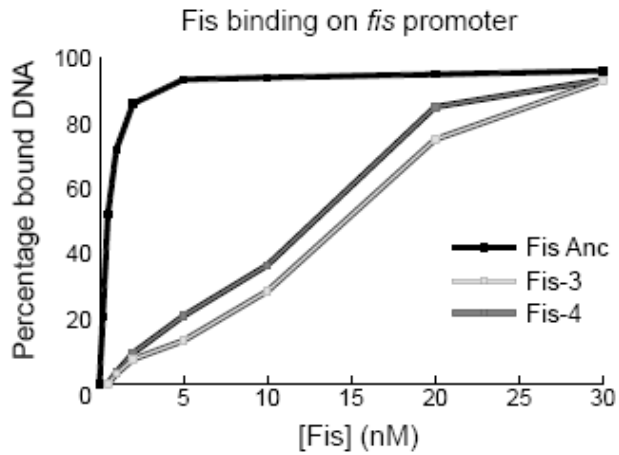
Ara+1 allele (-35):



3 fold decrease in transcription.

Ara-3, -4 alleles (coding sequence):



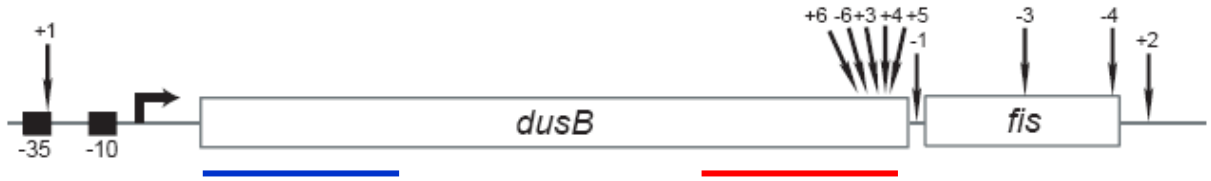


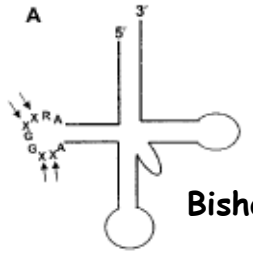
Decrease of Fis activity.

Mol. Biol. Evol. 2010.

Genetic parallel changes versus allelic variation?

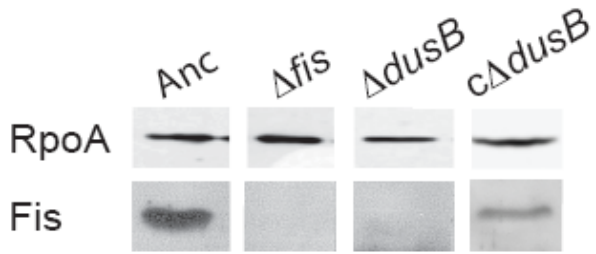
Ara+5 allele *(dusB):*





Bishop *et al.*, 2002.

Dihydrouridine synthase



New function: global regulation.
Activator of *fis* at the post-transcriptional level.

Mol. Biol. Evol. 2010.

Genetic parallel changes versus allelic variation?

Strong genetic parallelism

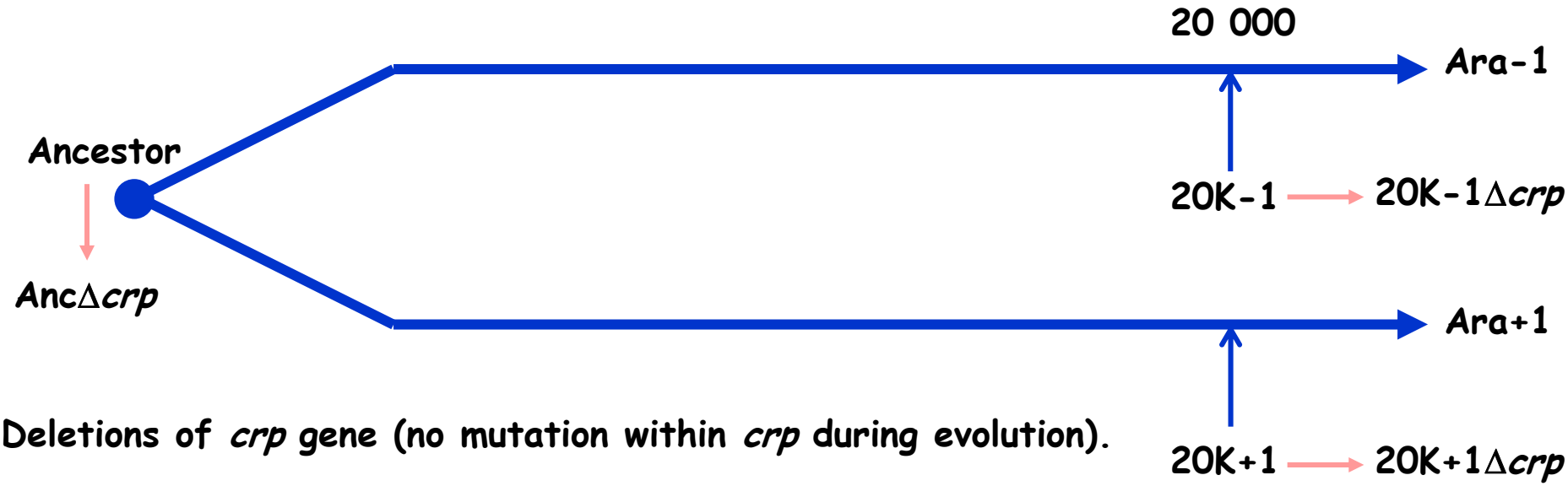
Allelic divergence: all different, all levels of regulation

Fitness effects: far from predictable.

-1: 3% benefit (ancestral background)
+1, -3: 5% deleterious (ancestral background)
+5: neutral (ancestral AND evolved backgrounds !!!!).

Fitness assays have to be carefully considered.

Parallel changes in gene regulatory networks



Deletions of *crp* gene (no mutation within *crp* during evolution).

Effect of *crp* deletion at 0 and 20 000 generations:

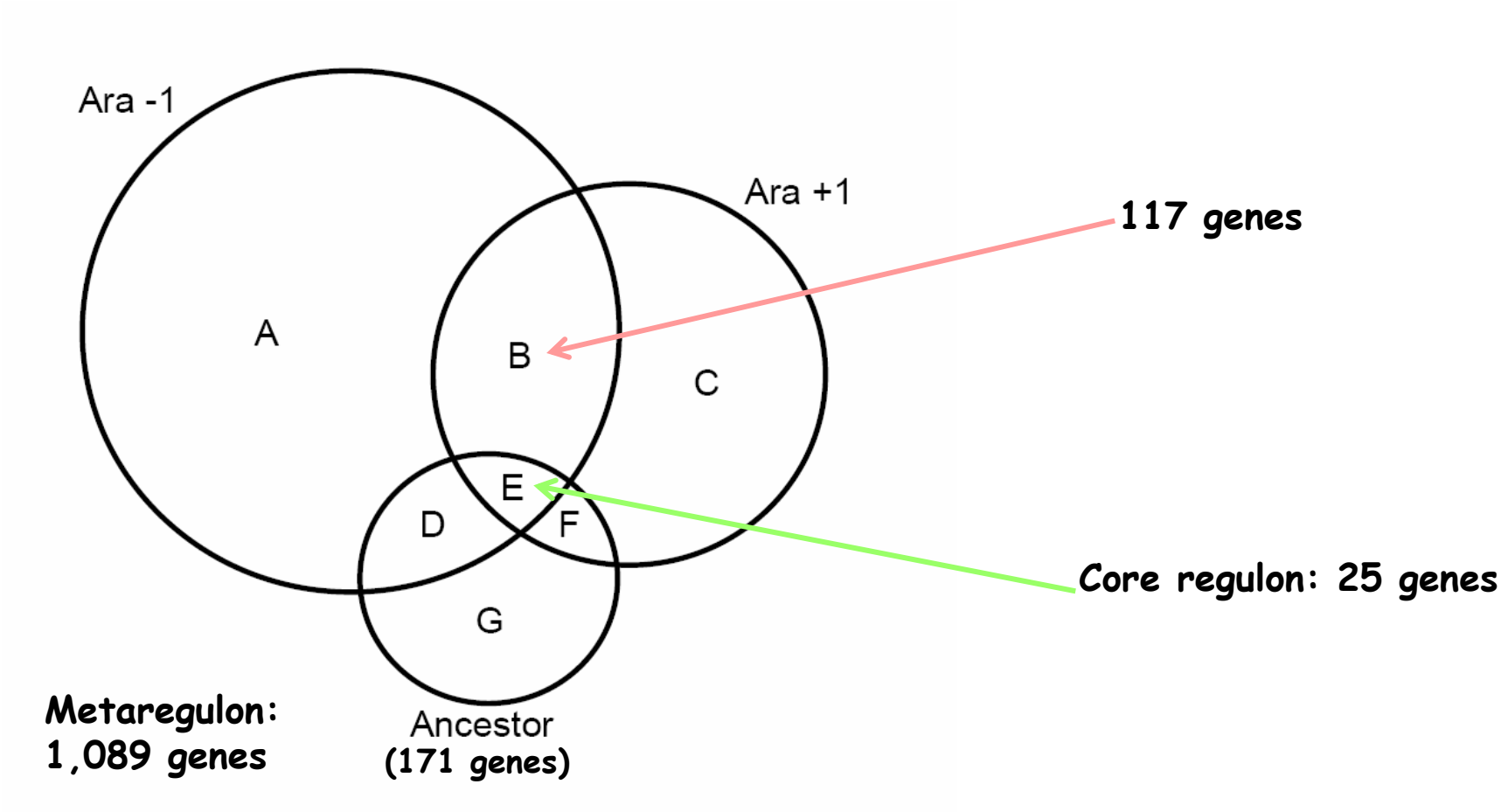
- On growth rates:

	Growth rate crp^+ / crp^-
Ancestor	1.35
20K-1	13.01
20K+1	5.92

- On global transcription profiles.

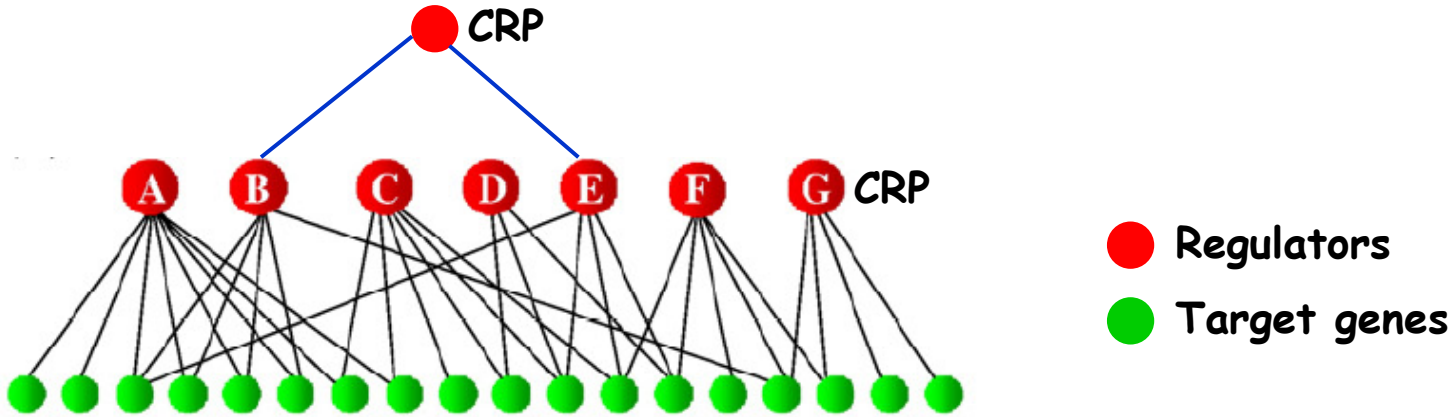
PLoS Genetics 2008

Parallel changes in gene regulatory networks



Parallel extension of the CRP regulon.
Same directional expression changes.

Parallel changes in gene regulatory networks



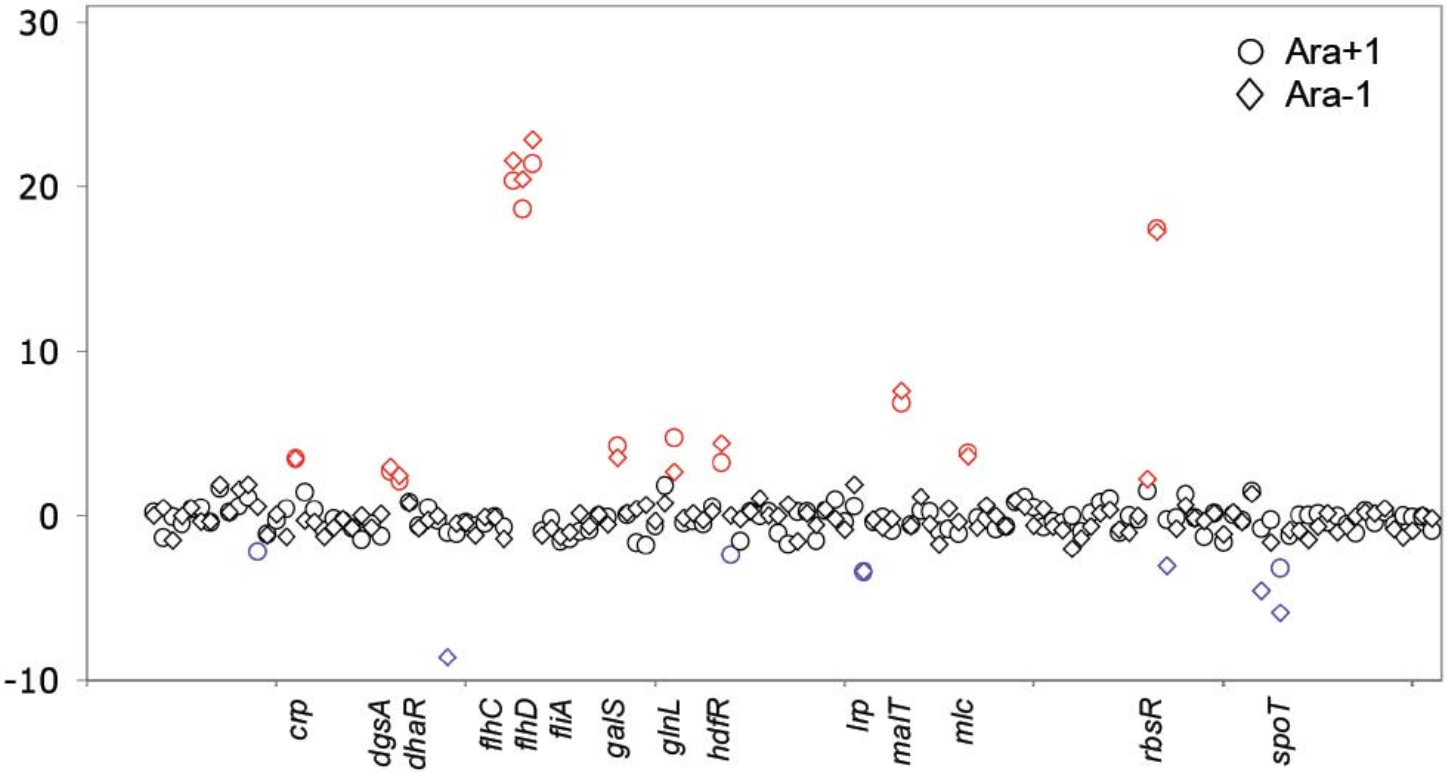
Strong modification of CRP regulon during evolution:

- mutation within *crp* or *cya*? NO
- mutation within each target gene? NO
- mutations within global regulators interacting with CRP?

Modifications of global architecture of regulatory networks?
Epistatic interactions between *crp* and beneficial mutations substituted during evolution?

Parallel changes in gene regulatory networks

Significance of regulon expression change (difference from ancestor double Z-score)



14 regulons modified in both evolved lines, AND in the same direction.

Parallel epistatic changes with modifications of global regulatory genes.

Three already identified: *malT*, *rbsR*, et *spoT*!!

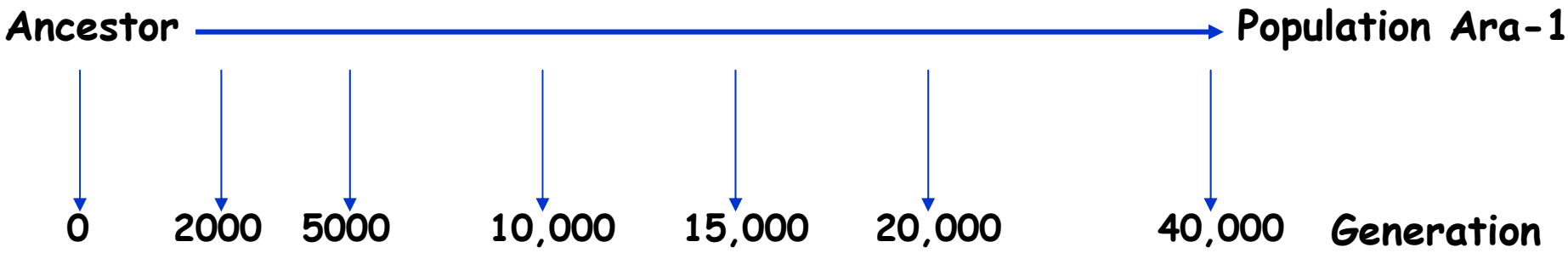
Genome sequences after long-term evolution

Questions:

- how many mutations?
- rates of phenotypic versus genomic evolution?
- mutation rates?
- genomic parallelism?
- rewiring of global regulatory networks? Metabolic networks?
- epistatic interactions?
- pleiotropic interactions?
- historical contingency?

Genome sequences after long-term evolution

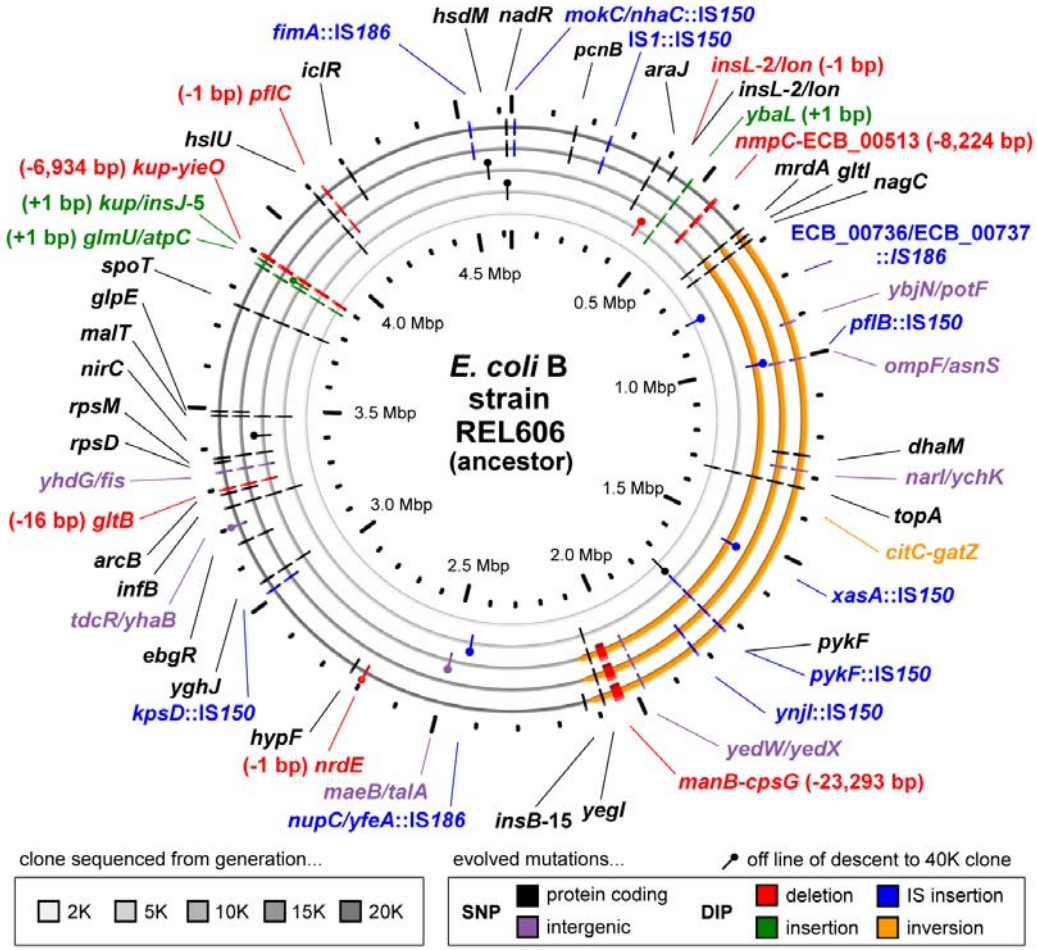
« We see nothing of these slow changes in progress,
until the hand of time has marked the long lapse of ages… »
Darwin, 1859, *The Origin of Species*.



Genome sequence of one evolved clone at each generation.

Nature 2009

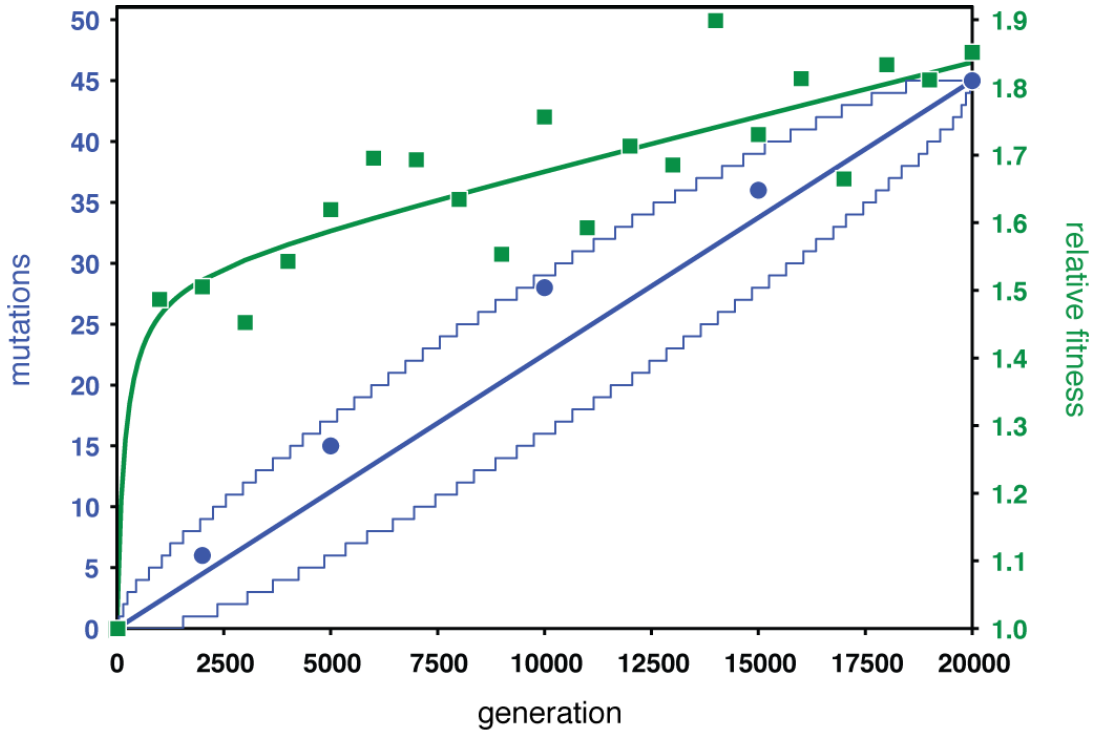
Genome sequences after long-term evolution: How many mutations?



After 20,000 generations:
45 mutations including 29 SNPs and 16 others.
29 SNPs: 26 in coding regions.

About 2 / 1000 generations.

Genome sequences after long-term evolution: Trajectories for phenotypic and genomic evolution?



Trajectory for genomic evolution: near-linear.
Fitness trajectory: non linear.

Small fraction of beneficial mutations, early in the experiment?
Constant pace of neutral mutations?

Genome sequences after long-term evolution: Trajectories for phenotypic and genomic evolution?

Characteristics of 45 mutations:

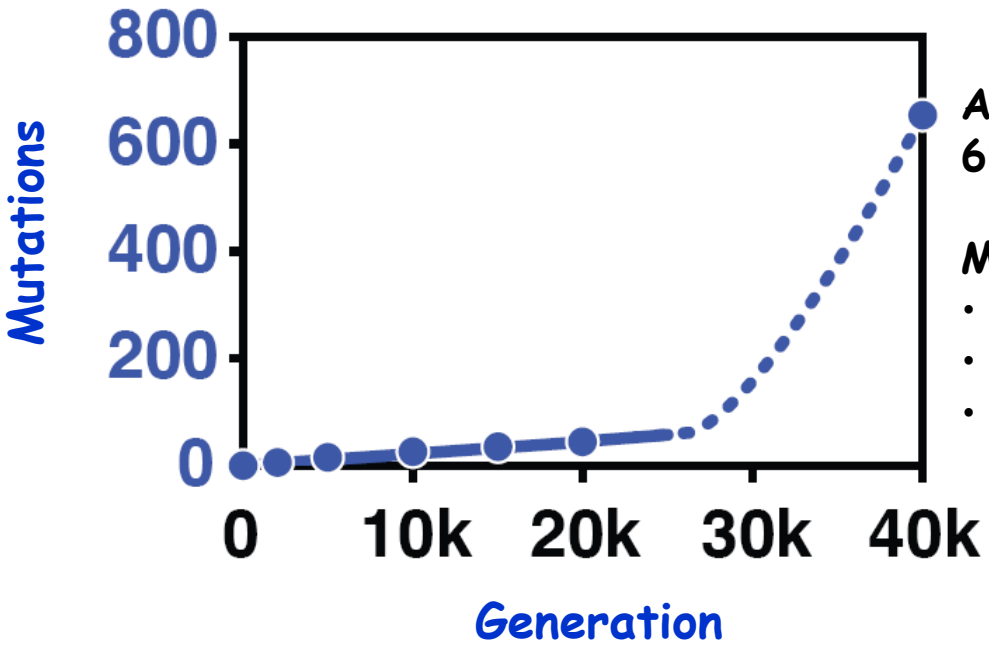
- 26 SNPs in coding regions: all non-synonymous
- substantial genetic parallelism for 14 analyzed genes
- substitution of almost all mutations
- fitness checked for 9 mutations: 8 beneficial.

Gene or Region	Function	% Parallel
nadR	transcriptional regulator	100%
pykF	pyruvate kinase	100%
rbs operon	ribose catabolism	100%
malT	transcriptional regulator	64%
spoT	stringent response regulator	64%
mrdA	cell-wall biosynthesis	45%
infB	translation initiation factor 2	45%*
fis	nucleoid-associated protein	27%
topA	DNA topoisomerase I	27%
pcnB	poly(A) polymerase	27%
ompF	porin	18%*
rpsD	30S ribosomal protein	18%*
rpsM	30S ribosomal protein	0%
glmU promoter	cell-wall biosynthesis	0%

Gene or Region	Fitness Effect
topA	13.3%
pykF	11.1%
spoT	9.4%
nadR	8.1%
glmU promoter	4.9%
fis	2.9%
rbs operon	2.1%
malT	0.4%
ompF	-9.7%

Predominance of beneficial mutations.

Genome sequences after long-term evolution: Mutation rates?



After 40,000 generations:
627 SNPs and 26 others.

Mutation in *mutT*:
- 92.3% of A·T to C · G
- detected by generation 26,500
- 50- to 100-fold mutation rate increase.

Number of generations since mutator: 40,000 – 25,000
Genome size at 40K: 4.57×10^6 bp
Number of synonymous substitutions: 83
Proportion of *mutT* synonymous SNPs: 11.3%

Mutation rate = 1.1×10^{-8} mutation / bp / generation
Estimation for the ancestor: 1.6×10^{-10} mutation / bp / generation.

Genome sequences after long-term evolution: 15 additional clones from the same population

Project « Genome Evolution »: more genome sequences.

- **Génoscope:**

Claudine Médigue

Stéphane Cruveiller

Béatrice Chane-Woon-Ming

- **Laboratoire Ecologie et Evolution, Paris 7:**

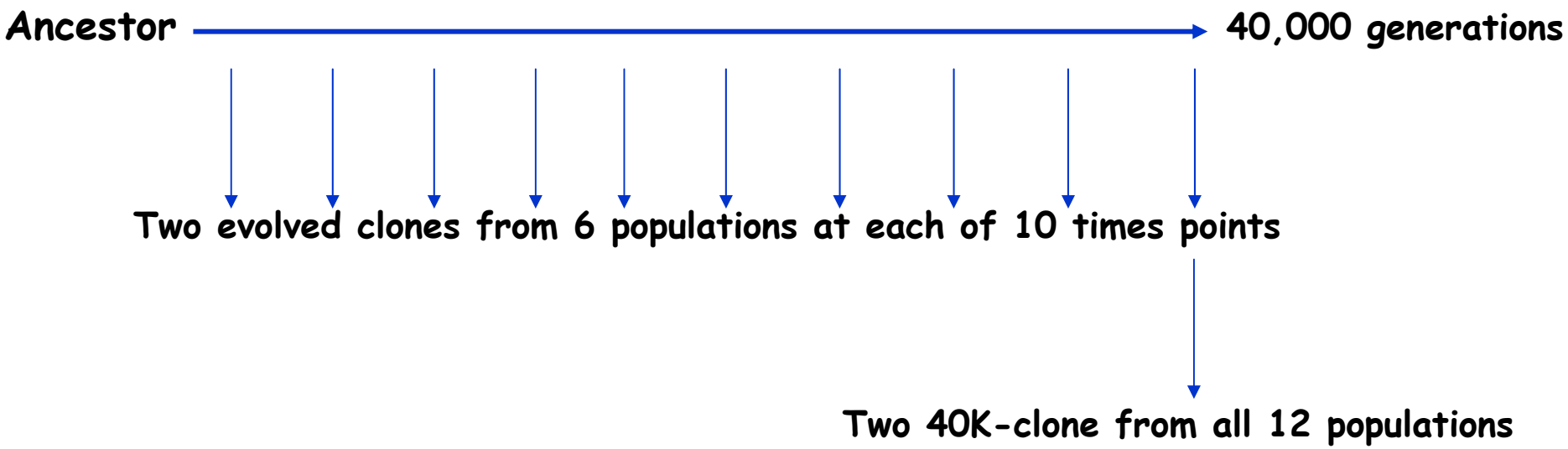
Olivier Tenaillon

Genome sequences after long-term evolution: Even more genomes

Project « Genome Evolution »: even more genome sequences:

two evolved clones from each of 6 populations at each of 10 time points

two evolved clones at 40,000 generations from all 12 populations.



And now?

Questions:

- mutation rates?
- genomic parallelism? Trajectories for genome evolution?
- rewiring of global regulatory networks? Metabolic networks?
- long-term co-existence of several ecotypes?
- epistatic interactions?
- pleiotropic interactions?
- historical contingency?

LAPM (Grenoble):

Evelyne Coursange, Joël Gaffé, Thomas Hindré, Mickaël Le Gac, Jessica Plucain, Colin Raeside, Sébastien Wielgoss.

Elisabeth Kay, Mike Vergnes, Xavier Charpentier, Emmanuelle Le Pihive, Brieuc Gestin, Max Maurin.

UNIVERSITE JOSEPH FOURIER
GRENOBLE 1
SCIENCES.TECHNOLOGIE.SANTÉ

CNRS
CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

Agence Nationale de la Recherche
ANR

Many thanks…

Michigan State University:
Jeffrey Barrick, Richard Lenski.

University of Houston:
Tim Cooper.

Korea Research Institute of Bioscience and Biotechnology (Kribb):
Haeyoung Jeong, Jihyun F. Kim.