Genom evolúció

2010.11.15

Genom programok

- Szekvenálási stratégia
- Genom jellemzése (pl. %CG tartalom…)
- Gének annotációja (ORF – gén – cisztron)
- Géntérkép vs. fizikai térkép
- Új gének – prediktált funkciók (ortológok)
- Funkcionális géncsoportok

Hőskor: 90-es évek

ARTICLE

The Complete Genome Sequence of *Escherichia coli* K-12

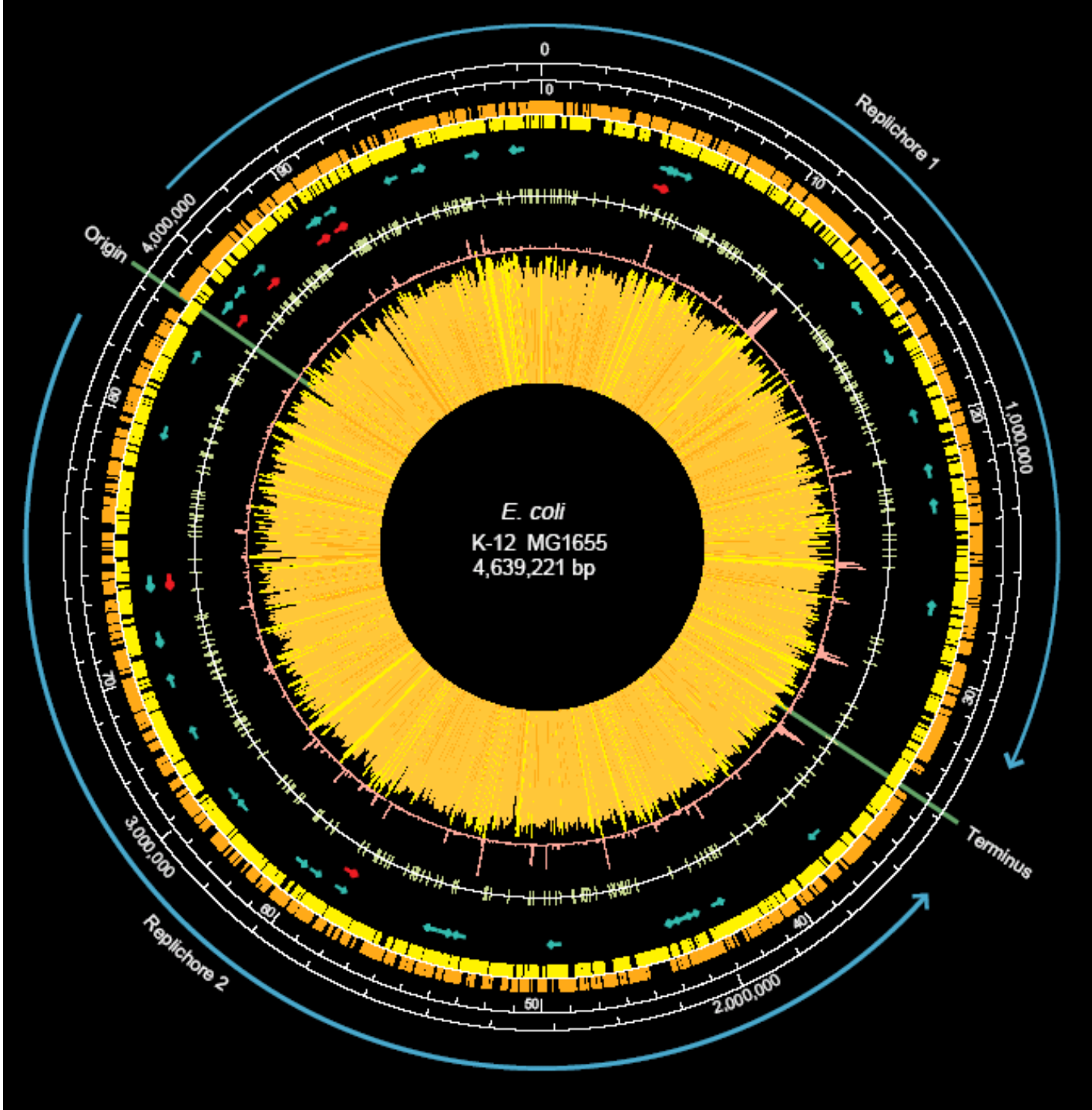
Frederick R. Blattner,* Guy Plunkett III,* Craig A. Bloch, Nicole T. Perna, Valerie Burland, Monica Riley, Julio Collado-Vides, Jeremy D. Glasner, Christopher K. Rode, George F. Mayhew, Jason Gregor, Nelson Wayne Davis, Heather A. Kirkpatrick, Michael A. Goeden, Debra J. Rose, Bob Mau, Ying Shao

The 4,639,221–base pair sequence of *Escherichia coli* K-12 is presented. Of 4288 protein-coding genes annotated, 38 percent have no attributed function. Comparison with five other sequenced microbes reveals ubiquitous as well as narrowly distributed gene families; many families of similar genes within *E. coli* are also evident. The largest family of paralogous proteins contains 80 ABC transporters. The genome as a whole is strikingly organized with respect to the local direction of replication; guanines, oligonucleotides possibly related to replication and recombination, and most genes are so oriented. The genome also contains insertion sequence (IS) elements, phage remnants, and many other patches of unusual composition indicating genome plasticity through horizontal transfer.

The first 1.92 Mb (*13*, *14*), positions 2,686,777 to 4,639,221 [in base pairs (bp)], was sequenced from our overlapping set of 15- to 20-kb MG1655 lambda clones (*15*) by means of radioactive chemistry and was deposited in GenBank between 1992 and 1995. Subsequently, we switched to dye-terminator fluorescence sequencing (Applied Biosystems). In addition to greater speed and lower cost, this new technology avoided electrophoretic compression arti-

SCIENCE • VOL. 277 • 5 SEPTEMBER 1997

E. coli genom szerkezete



The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*

F. Kunst[1], N. Ogasawara[2], I. Moszer[3], A. M. Albertini[4], G. Alloni[4], V. Azevedo[5], M. G. Bertero[3,4], P. Bessières[5], A. Bolotin[5], S. Borchert[6], R. Borriss[7], L. Boursier[3], A. Brans[8], M. Braun[9], S. C. Brignell[10], S. Bron[11], S. Brouillet[3,12], C. V. Bruschi[13], B. Caldwell[14], V. Capuano[5], N. M. Carter[10], S.-K. Choi[15], J.-J. Codani[16], I. F. Connerton[17], N. J. Cummings[17], R. A. Daniel[18], F. Denizot[19], K. M. Devine[20], A. Düsterhöft[9], S. D. Ehrlich[5], P. T. Emmerson[21], K. D. Entian[6], J. Errington[18], C. Fabret[19], E. Ferrari[14], D. Foulger[18], C. Fritz[9], M. Fujita[22], Y. Fujita[23], S. Fuma[24], A. Galizzi[4], N. Galleron[5], S.-Y. Ghim[15], P. Glaser[3], A. Goffeau[25], E. J. Golightly[26], G. Grandi[27], G. Guiseppi[19], B. J. Guy[10], K. Haga[28], J. Haiech[19], C. R. Harwood[10], A. Hénaut[29], H. Hilbert[9], S. Holsappel[11], S. Hosono[30], M.-F. Hullo[3], M. Itaya[31], L. Jones[32], B. Joris[8], D. Karamata[33], Y. Kasahara[2], M. Klaerr-Blanchard[3], C. Klein[6], Y. Kobayashi[30], P. Koetter[6], G. Koningstein[34], S. Krogh[20], M. Kumano[24], K. Kurita[24], A. Lapidus[5], S. Lardinois[8], J. Lauber[9], V. Lazarevic[33], S.-M. Lee[35], A. Levine[36], H. Liu[28], S. Masuda[30], C. Mauël[33], C. Médigue[3,12], N. Medina[36], R. P. Mellado[37], M. Mizuno[30], D. Moestl[9], S. Nakai[2], M. Noback[11], D. Noone[20], M. O'Reilly[20], K. Ogawa[24], A. Ogiwara[38], B. Oudega[34], S.-H. Park[15], V. Parro[37], T. M. Pohl[39], D. Portetelle[40], S. Porwollik[7], A. M. Prescott[18], E. Presecan[3], P. Pujic[5], B. Purnelle[25], G. Rapoport[1], M. Rey[26], S. Reynolds[33], M. Rieger[41], C. Rivolta[33], E. Rocha[3,12], B. Roche[36], M. Rose[6], Y. Sadaie[22], T. Sato[30], E. Scanlan[20], S. Schleich[3], R. Schroeter[7], F. Scoffone[4], J. Sekiguchi[42], A. Sekowska[3], S. J. Seror[36], P. Serror[5], B.-S. Shin[15], B. Soldo[33], A. Sorokin[5], E. Tacconi[4], T. Takagi[43], H. Takahashi[28], K. Takemaru[30], M. Takeuchi[30], A. Tamakoshi[24], T. Tanaka[44], P. Terpstra[11], A. Tognoni[27], V. Tosato[13], S. Uchiyama[42], M. Vandenbol[40], F. Vannier[36], A. Vassarotti[45], A. Viari[12], R. Wambutt[46], E. Wedler[46], H. Wedler[46], T. Weitzenegger[39], P. Winters[14], A. Wipat[10], H. Yamamoto[42], K. Yamane[24], K. Yasumoto[28], K. Yata[22], K. Yoshida[23], H.-F. Yoshikawa[28], E. Zumstein[5], H. Yoshikawa[2] & A. Danchin[3]

[1] *Institut Pasteur, Unité de Biochimie Microbienne, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France*
[2] *Nara Institute of Science and Technology, Graduate School of Biological Sciences, Ikoma, Nara 630-01, Japan*
[3] *Institut Pasteur, Unité de Régulation de l'Expression Génétique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France*
[4] *Dipartimento di Genetica e Microbiologia, Universita di Pavia, Via Abbiategrasso 207, 27100 Pavia, Italy*
[5] *INRA, Génétique Microbienne, Domaine de Vilvert, 78352 Jouy-en-Josas Cedex, France*
[6] *Institut für Mikrobiologie, J. W. Goethe-Universität, Marie Curie Strasse 9, 60439 Frankfurt/Maine, Germany*
[7] *Institut für Genetik und Mikrobiologie, Humboldt Universität, Chausseestrasse 17, D-10115 Berlin, Germany*
[8] *Centre d'Ingénierie des Protéines, Université de Liège, Institut de Chimie B6, Sart Tilman, B-4000 Liège, Belgium*
[9] *QIAGEN GmbH, Max-Volmer-Strasse 4, D-40724 Hilden, Germany*
[10] *Department of Microbiological, Immunological and Virological Sciences, The Medical School, University of Newcastle, Framlington Place, Newcastle upon Tyne NE2 4HH, UK*
[11] *Department of Genetics, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands*
[12] *Atelier de BioInformatique, Université Paris VI, 12 rue Cuvier, 75005 Paris, France*
[13] *ICGEB, AREA Science Park, Padriciano 99, I-34012 Trieste, Italy*
[14] *Genencor International, 925 Page Mill Road, Palo Alto, California 94304-1013, USA*
[15] *Bacterial Molecular Genetics Research Unit, Applied Microbiology Research Division, KRIBB, PO Box 115, Yusong, Taejon 305-600, Korea*
[16] *INRIA, Domaine de Voluceau, PB 105, 78153 Le Chesnay Cedex, France*
[17] *Institute of Food Research, Department of Food Macromolecular Science, Reading Laboratory, Earley Gate, Whiteknights Road, Reading RG6 6BZ, UK*
[18] *Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, UK*
[19] *Laboratoire de Chimie Bactérienne, CNRS BP 71, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 09, France*
[20] *Department of Genetics, Trinity College, Lincoln Place Gate, Dublin 2, Republic of Ireland*
[21] *Department of Biochemistry and Genetics, The Medical School, University of Newcastle, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK*
[22] *Radioisotope Center, National Insitute of Genetics, Mishima, Shizuoka-ken 411, Japan*
[23] *Department of Biotechnology, Faculty of Engineering, Fukuyama University, Higashinnara-cho, Fukuyama-shi, Hiroshima 729-02, Japan*
[24] *Institute of Biological Sciences, Tsukuba University, Tsukuba-shi, Ibaraki 305, Japan*

A nyers genom szekvencia kimenetei:

Table 1. Functional classification of the *Bacillus subtilis* protein-coding genes.

I	CELL ENVELOPE AND CELLULAR PROCESSES 966	
I.1	CELL WALL	93
cwlA	2665	*N*-acetylmuramoyl-L-alanine amidase (minor autolysin)
cwlC	1873	*N*-acetylmuramoyl-L-alanine amidase (sporulation mother cell wall)
cwlD	157	*N*-acetylmuramoyl-L-alanine amidase (germination)
cwlJ	282	cell wall hydrolase (sporulation)
dacA	18	penicillin-binding protein 5 (D-alanyl-D-alanine carboxypeptidase) (peptidoglycan biosynthesis)
dacB	2424	penicillin-binding protein 5* (D-alanyl-D-alanine carboxypeptidase) (peptidoglycan biosynthesis) (spore cortex)
dacF	2445	penicillin-binding protein (D-alanyl-D-alanine carboxypeptidase) (peptidoglycan biosynthesis)
ddlA	508	D-alanyl-D-alanine ligase A (peptidoglycan biosynthesis)
dltA	3961	D-alanyl-D-alanine carrier protein ligase (lipoteichoic acid biosynthesis)
dltB	3963	D-alanine transfer from Dcp to undecaprenol-phosphate (lipoteichoic acid biosynthesis)
dltC	3964	D-alanine carrier protein (lipoteichoic acid biosynthesis)
dltD	3964	D-alanine transfer from undecaprenol-phosphate to the poly(glycerophosphate) chain (lipoteichoic acid biosynthesis)
dltE	3965	involved in lipoteichoic acid biosynthesis
gcaD	56	UDP-*N*-acetylglucosamine pyrophosphorylase (peptidoglycan and lipopolysaccharide biosynthesis)
ggaA	3670	galactosamine-containing minor teichoic acid biosynthesis
ggaB	3669	galactosamine-containing minor teichoic acid biosynthesis
gtaB	3665	UTP-glucose-1-phosphate uridylyltransferase
lytB	3662	modifier protein of major autolysin LytC (CWBP76)
lytC	3660	*N*-acetylmuramoyl-L-alanine amidase (major autolysin) (CWBP49)
lytD	3687	*N*-acetylglucosaminidase (major autolysin) (CWBP90)
lytE	1018	cell wall lytic activity (CWBP33)
mbl	3747	MreB-like protein
mraY	1587	phospho-*N*-acetylmuramoyl-pentapeptide transferase (peptidoglycan biosynthesis)
mreB	2861	cell-shape determining protein
mreBH	1517	cell-shape determining protein
mreC	2860	cell-shape determining protein
mreD	2859	cell-shape determining protein
murA	3778	UDP-*N*-acetylglucosamine 1-carboxyvinyltransferase (peptidoglycan biosynthesis)
murB	1592	UDP-*N*-acetylenolpyruvoylglucosamine reductase (peptidoglycan biosynthesis)
murC	3049	UDP-*N*-acetylmuramate-alanine ligase (peptidoglycan biosynthesis)
murD	1588	UDP-*N*-acetylmuramoylalanine-D-glutamate ligase (peptidoglycan biosynthesis)
murE	1586	UDP-*N*-acetylmuramoylananine-D-glutamate-2,6-diaminopimelate ligase (peptidoglycan biosynthesis)
murF	509	UDP-*N*-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate-D-alanyl-D-alanyl ligase (peptidoglycan biosynthesis)
murG	1581	UDP-*N*-acetylglucosamine-*N*-acetylmuramyl-(pentapeptide)pyrophosphoryl-undecaprenol *N*-acetylglucosamine transferase (peptidoglycan biosynthesis)
murZ	3906	UDP-*N*-acetylglucosamine 1-carboxyvinyltrans-
		prophage-mediated lysis)
xlyB	1317	*N*-acetylmuramoyl-L-alanine amidase (PBSX prophage-mediated lysis)
ybxG	799	CDP-glucose 4,6-dehydratase
yhdD	1013	cell wall-binding protein
ykuA	1487	penicillin-binding protein
ylbI	1569	lipopolysaccharide core biosynthesis
ymaG	1865	cell wall protein
yngB	1946	UTP-glucose-1-phosphate uridylyltransferase
yocH	2093	cell wall-binding protein
yodJ	2135	D-alanyl-D-alanine carboxypeptidase
yojL	2116	cell wall-binding protein
yomC	2263	*N*-acetylmuramoyl-L-alanine amidase
ypdQ	2310	cell wall enzyme
ypfP	2306	cell wall synthesis
ypjH	2357	lipopolysaccharide biosynthesis-related protein
yqeE	2649	*N*-acetylmuramoyl-L-alanine amidase
yqfY	2588	peptidoglycan acetylation
yqiI	2515	*N*-acetylmuramoyl-L-alanine amidase
yrhL	2771	acyltransferase
yrrR	2791	penicillin-binding protein
yrvJ	2818	*N*-acetylmuramoyl-L-alanine amidase
ytcC	3157	lipopolysaccharide *N*-acetylglucosaminyltransferase
ytkC	3135	autolytic amidase
ytxN	3161	lipopolysaccharide *N*-acetylglucosaminyltransferase
yubE	3191	*N*-acetylmuramoyl-L-alanine amidase
yvcE	3575	cell wall-binding protein
ywhE	3849	penicillin-binding protein
ywtD	3697	murein hydrolase
I.2	TRANSPORT/BINDING PROTEINS AND LIPOPROTEINS	381
aapA	2766	amino acid permease
alsT	1938	amino acid carrier protein
amyC	3099	maltose transport protein
amyD	3098	sugar transport
appA	1213	oligopeptide ABC transporter (oligopeptide-binding protein)
appB	1215	oligopeptide ABC transporter (permease)
appC	1216	oligopeptide ABC transporter (permease)
appD	1211	oligopeptide ABC transporter (ATP-binding protein)
appF	1212	oligopeptide ABC transporter (ATP-binding protein)
araE	3485	L-arabinose transport (permease)
araN	2942	L-arabinose transport (sugar-binding protein)
araP	2941	L-arabinose transport (integral membrane protein)
araQ	2940	L-arabinose transport (integral membrane protein)
azlC	2729	branched-chain amino acid transport
azlD	2728	branched-chain amino acid transport
bglP	4084	phosphotransferase system (PTS) β-glucoside-specific enzyme IIABC component
blt	2716	multidrug-efflux transporter
bmr	2494	multidrug-efflux transporter
braB	3027	branched-chain amino acid transporter
brnQ	2728	branched-chain amino acid transporter
citM	834	secondary transporter of the Mg^{2+}/citrate complex
csbX	2838	α-ketoglutarate permease
cydC	3976	ABC transporter required for expression of cytochrome *bd* (ATP-binding protein)
cydD	3974	ABC transporter required for expression of cytochrome *bd* (ATP-binding protein)
czcD	2724	cation-efflux system membrane protein
dppA	1360	dipeptide ABC transporter (sporulation)
dppB	1361	dipeptide ABC transporter (permease) (sporulation)
dppC	1362	dipeptide ABC transporter (permease) (sporula-
		specific enzyme IIC component
lmrB	290	lincomycin-resistance protein
lplA	779	lipoprotein
lplB	781	transmembrane lipoprotein
lplC	782	transmembrane lipoprotein
mdr	334	multidrug-efflux transporter (puromycin, norfloxacin, tosufloxacin)
msmE	3097	multiple sugar-binding protein
msmX	3984	multiple sugar-binding transport ATP-binding protein
mtlA	449	phosphotransferase system (PTS) mannitol-specific enzyme IIABC component
narK	3833	nitrite extrusion protein
nasA	363	nitrate transporter
natA	296	Na^{+} ABC transporter (extrusion) (ATP-binding protein)
natB	297	Na^{+} ABC transporter (extrusion) (membrane protein)
nrgA	3756	ammonium transporter
nupC	4050	pyrimidine-nucleoside transport protein
oppA	1219	oligopeptide ABC transporter (binding protein) (initiation of sporulation, competence development)
oppB	1221	oligopeptide ABC transporter (permease) (initiation of sporulation, competence development)
oppC	1222	oligopeptide ABC transporter (permease) (initiation of sporulation, competence development)
oppD	1223	oligopeptide ABC transporter (ATP-binding protein) (initiation of sporulation, competence development)
oppF	1224	oligopeptide ABC transporter (ATP-binding protein) (initiation of sporulation, competence development)
opuAA	321	glycine betaine ABC transporter (ATP-binding protein) (osmoprotection)
opuAB	322	glycine betaine ABC transporter (permease) (osmoprotection)
opuAC	323	glycine betaine ABC transporter (glycine betaine-binding protein) (osmoprotection)
opuBA	3462	choline ABC transporter (ATP-binding protein) (osmoprotection)
opuBB	3461	choline ABC transporter (membrane protein) (osmoprotection)
opuBC	3460	choline ABC transporter (choline-binding protein) (osmoprotection)
opuBD	3460	choline ABC transporter (membrane protein) (osmoprotection)
opuCA	3470	glycine betaine/carnitine/choline ABC transporter (ATP-binding protein) (osmoprotection)
opuCB	3469	glycine betaine/carnitine/choline ABC transporter (membrane protein) (osmoprotection)
opuCC	3468	glycine betaine/carnitine/choline ABC transporter (osmoprotectant-binding protein) (osmoprotection)
opuCD	3467	glycine betaine/carnitine/choline ABC transporter (membrane protein) (osmoprotection)
opuD	3076	glycine betaine transporter (osmoprotection)
opuE	728	proline transporter (osmoprotection)
pbuX	2319	xanthine permease
ptsG	1457	phosphotransferase system (PTS) glucose-specific enzyme IIABC component
ptsI	1459	phosphotransferase system (PTS) enzyme I (general energy coupling protein of the PTS)
pyrP	1618	uracil permease (pyrimidine biosynthesis)
rbsA	3703	ribose ABC transporter (ATP-binding protein)
rbsB	3705	ribose ABC transporter (ribose-binding protein)
rbsC	3704	ribose ABC transporter (permease)
rbsD	3702	ribose ABC transporter (membrane protein)
rocC	3976	amino acid permease (arginine and ornithine utilization)
rocE	4143	amino acid permease (arginine and ornithine utilization)

Funkcionális géncsoportok

Table 4. Distribution of *E. coli* proteins among 22 functional groups (simplified schema).

Functional class	Number	Percent of total
Regulatory function	45	1.05
Putative regulatory proteins	133	3.10
Cell structure	182	4.24
Putative membrane proteins	13	0.30
Putative structural proteins	42	0.98
Phage, transposons, plasmids	87	2.03
Transport and binding proteins	281	6.55
Putative transport proteins	146	3.40
Energy metabolism	243	5.67
DNA replication, recombination, modification, and repair	115	2.68
Transcription, RNA synthesis, metabolism, and modification	55	1.28
Translation, posttranslational protein modification	182	4.24
Cell processes (including adaptation, protection)	188	4.38
Biosynthesis of cofactors, prosthetic groups, and carriers	103	2.40
Putative chaperones	9	0.21
Nucleotide biosynthesis and metabolism	58	1.35
Amino acid biosynthesis and metabolism	131	3.06
Fatty acid and phospholipid metabolism	48	1.12
Carbon compound catabolism	130	3.03
Central intermediary metabolism	188	4.38
Putative enzymes	251	5.85
Other known genes (gene product or phenotype known)	26	0.61
Hypothetical, unclassified, unknown	1632	38.06
Total	4288	100.00*

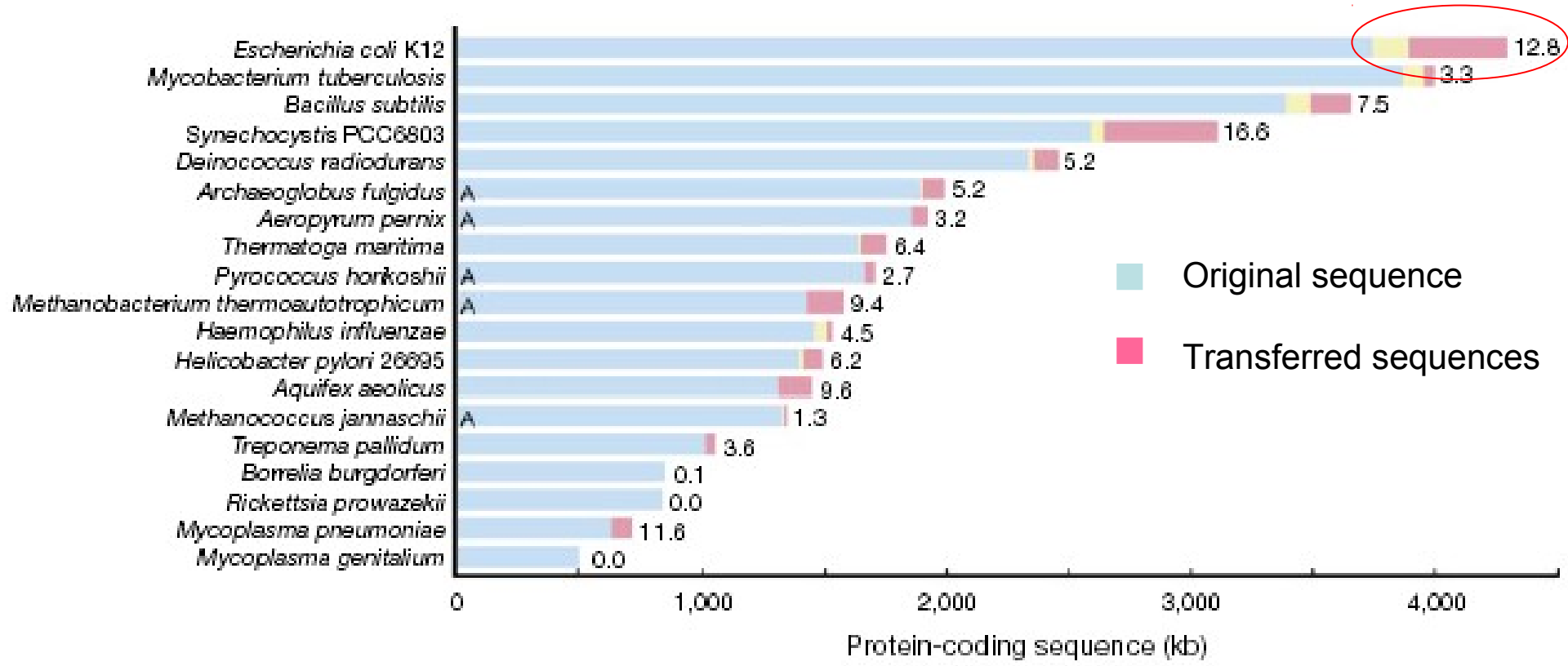
*Total of these rounded values is 99.97%.

Funkcionális *E. coli* géncsoportok

Functional class	Number	Percent of total
Regulatory function	45	1.05
Putative regulatory proteins	133	3.10
Cell structure	182	4.24
Putative membrane proteins	13	0.30
Putative structural proteins	42	0.98
Phage, transposons, plasmids	87	2.03
Transport and binding proteins	281	6.55
Putative transport proteins	146	3.40
Energy metabolism	243	5.67
DNA replication, recombination, modification, and repair	115	2.68
Transcription, RNA synthesis, metabolism, and modification	55	1.28
Translation, posttranslational protein modification	182	4.24
Cell processes (including adaptation, protection)	188	4.38
Biosynthesis of cofactors, prosthetic groups, and carriers	103	2.40
Putative chaperones	9	0.21
Nucleotide biosynthesis and metabolism	58	1.35
Amino acid biosynthesis and metabolism	131	3.06
Fatty acid and phospholipid metabolism	48	1.12
Carbon compound catabolism	130	3.03
Central intermediary metabolism	188	4.38
Putative enzymes	251	5.85
Other known genes (gene product or phenotype known)	26	0.61
Hypothetical, unclassified, unknown	1632	38.06
Total	4288	100.00*

*Total of these rounded values is 99.97%.

Horizontális transzferrel felvett gének aránya a baktérium genomokban



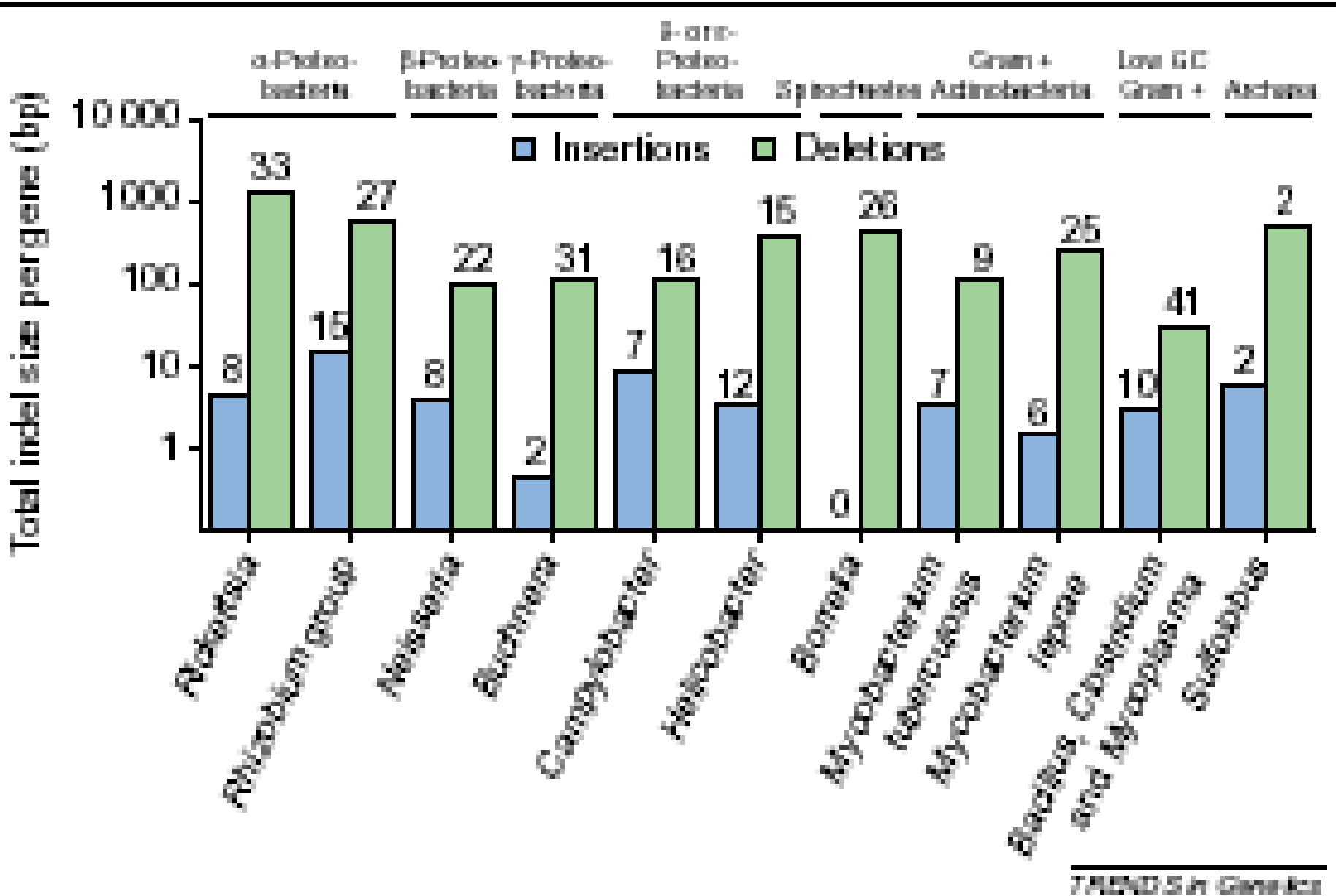
review article

NATURE | VOL 405 | 18 MAY 2000 |

Lateral gene transfer and the nature of bacterial innovation

Howard Ochman*, Jeffrey G. Lawrence† & Eduardo A. Groisman‡

Deléciók és inszerciók frekvenciája a baktérium genomokban



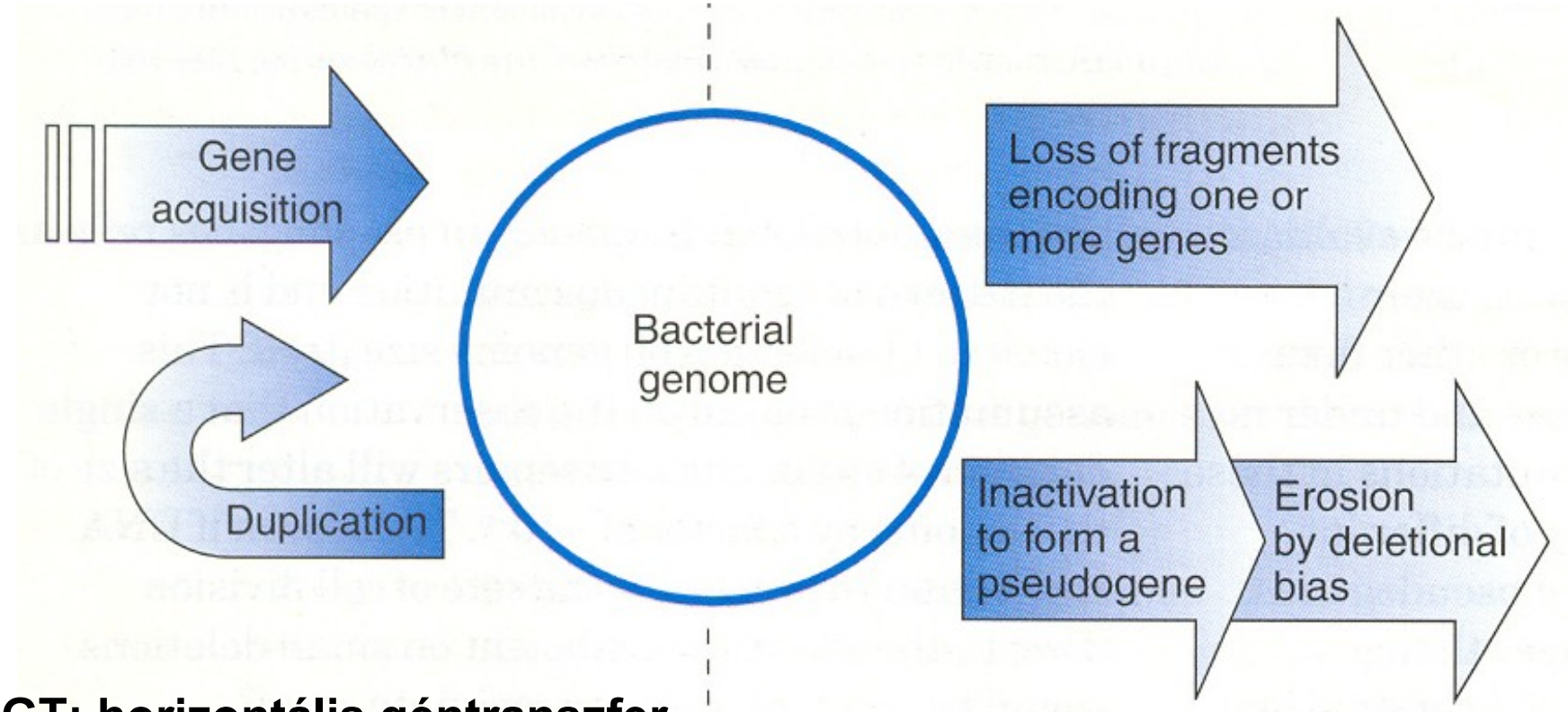
A bakteriális genomok evolúciós dinamikája

A DNS tartalom növekedése

gén felvétellel (HGT) és duplikációval

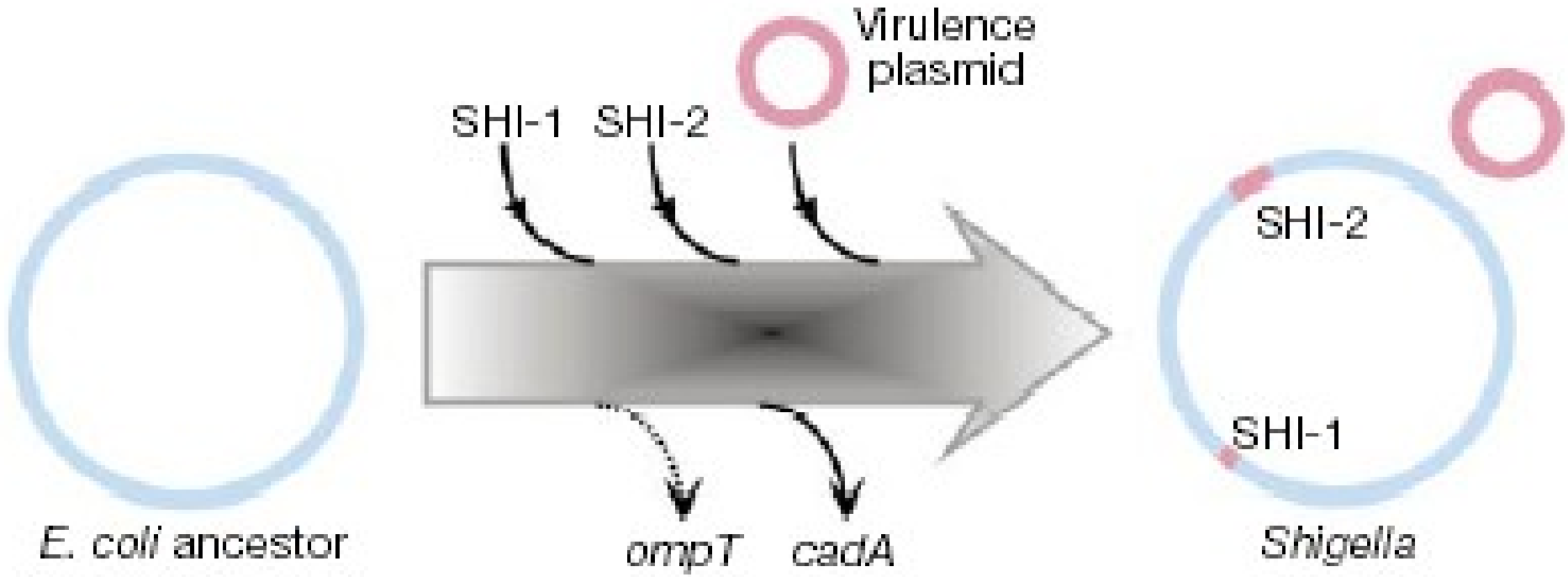
A DNS tartalom csökkenése

deléciókkal



HGT: horizontális géntranszfer

Példa: a patogén *Shigella* evolúciója *az E. coli* genomból

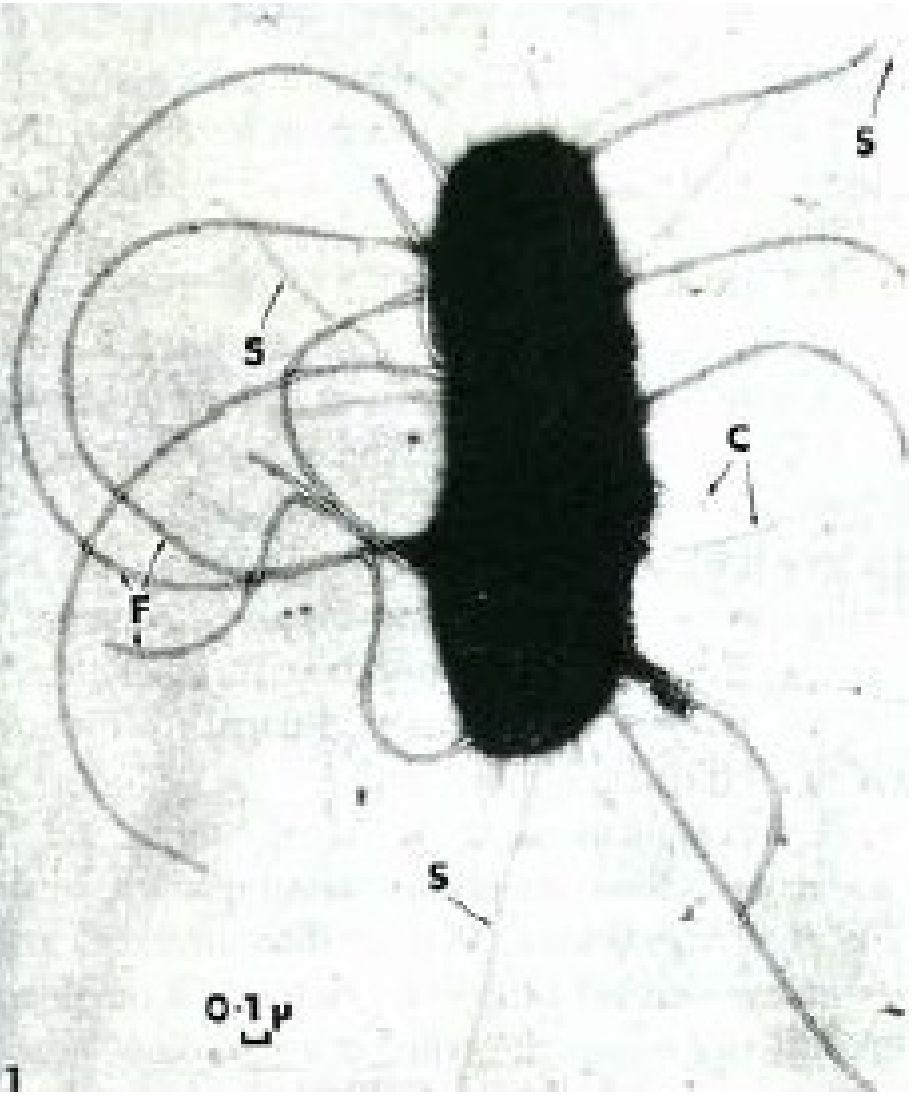


~100 millió év

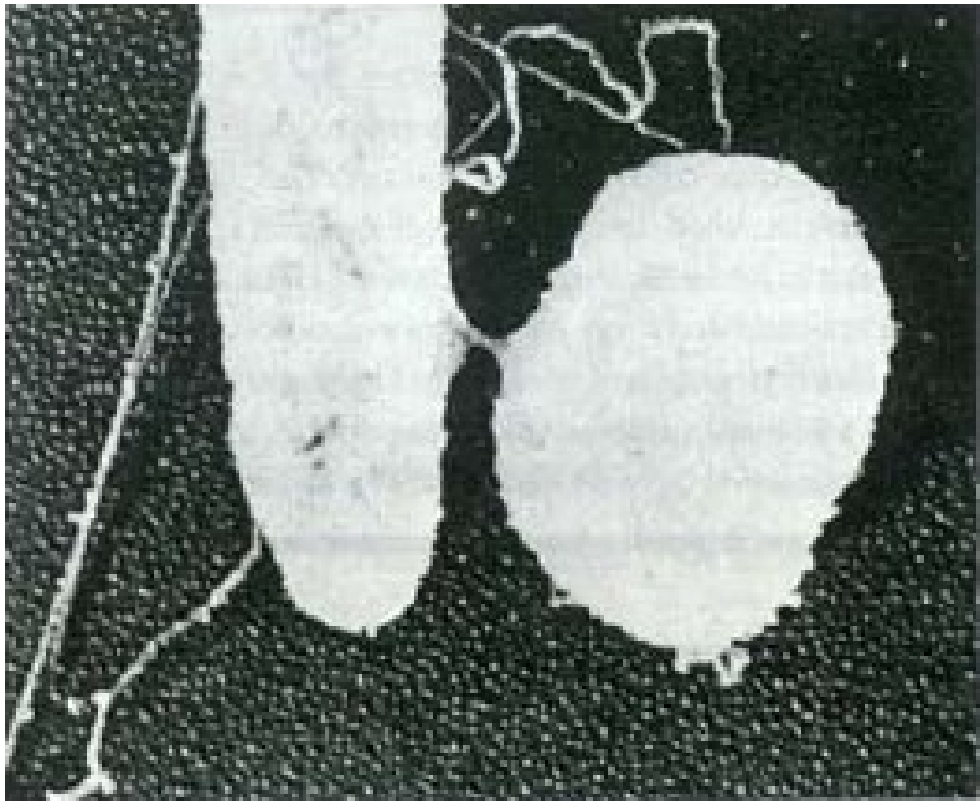
DNS átvitel mechanizmusai baktériumokban

1. Konjugáció (plazmidok)
2. Transzdukció (fágok)
3. Természetes genetikai transzformáció (külső DNS aktív felvétele)

1. Konjugáció



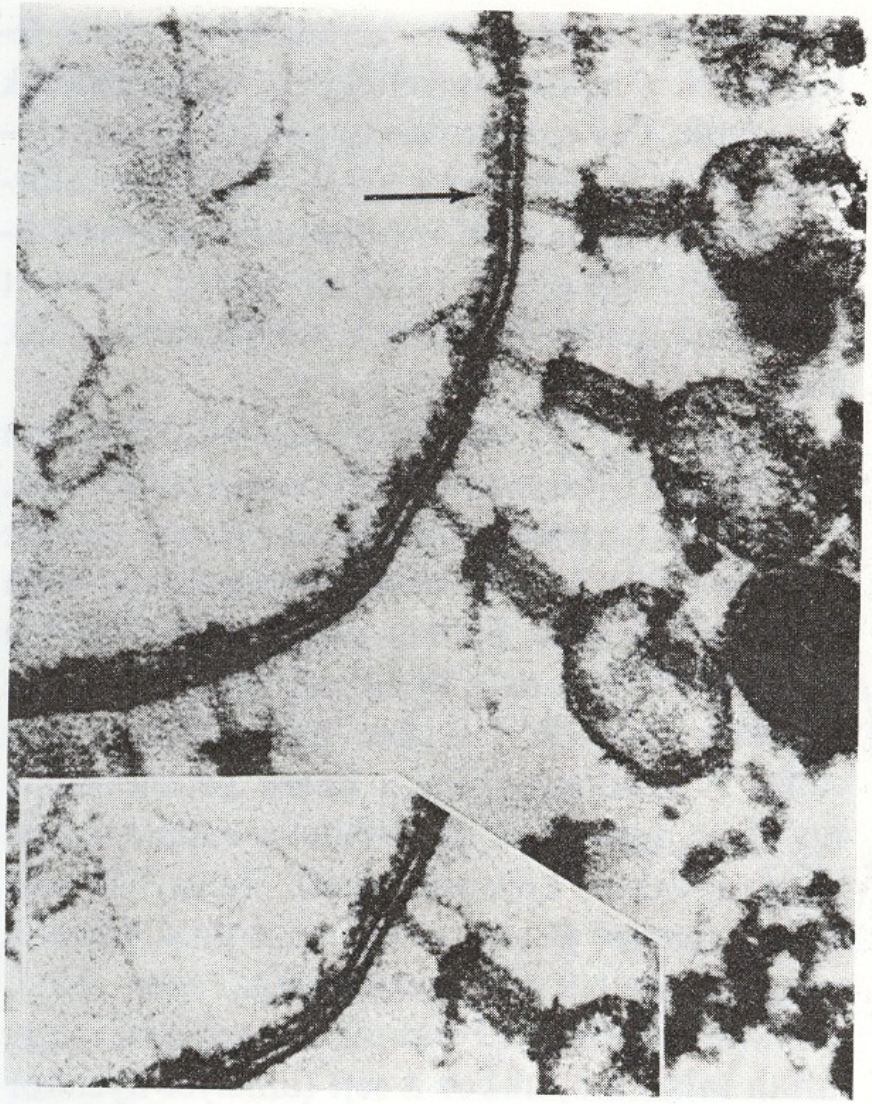
Szex (F) pílusok (*E. coli*)

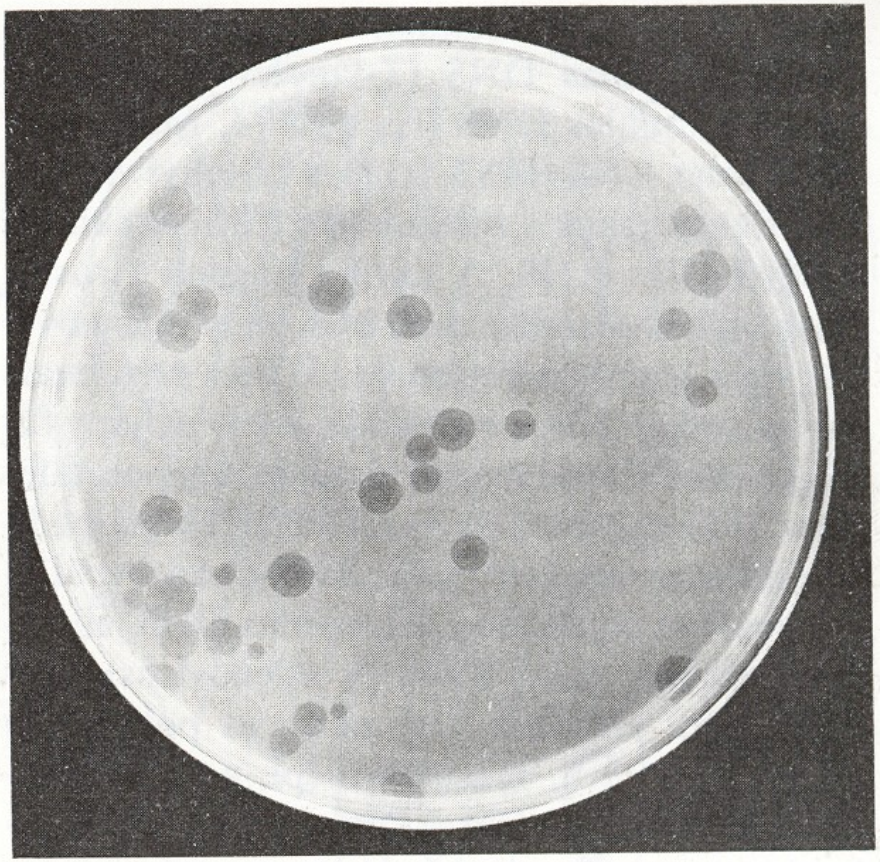


Konjugáló *E. coli* sejtek

3. DNS transzfer baktérium fágok segítségével - transzdukció

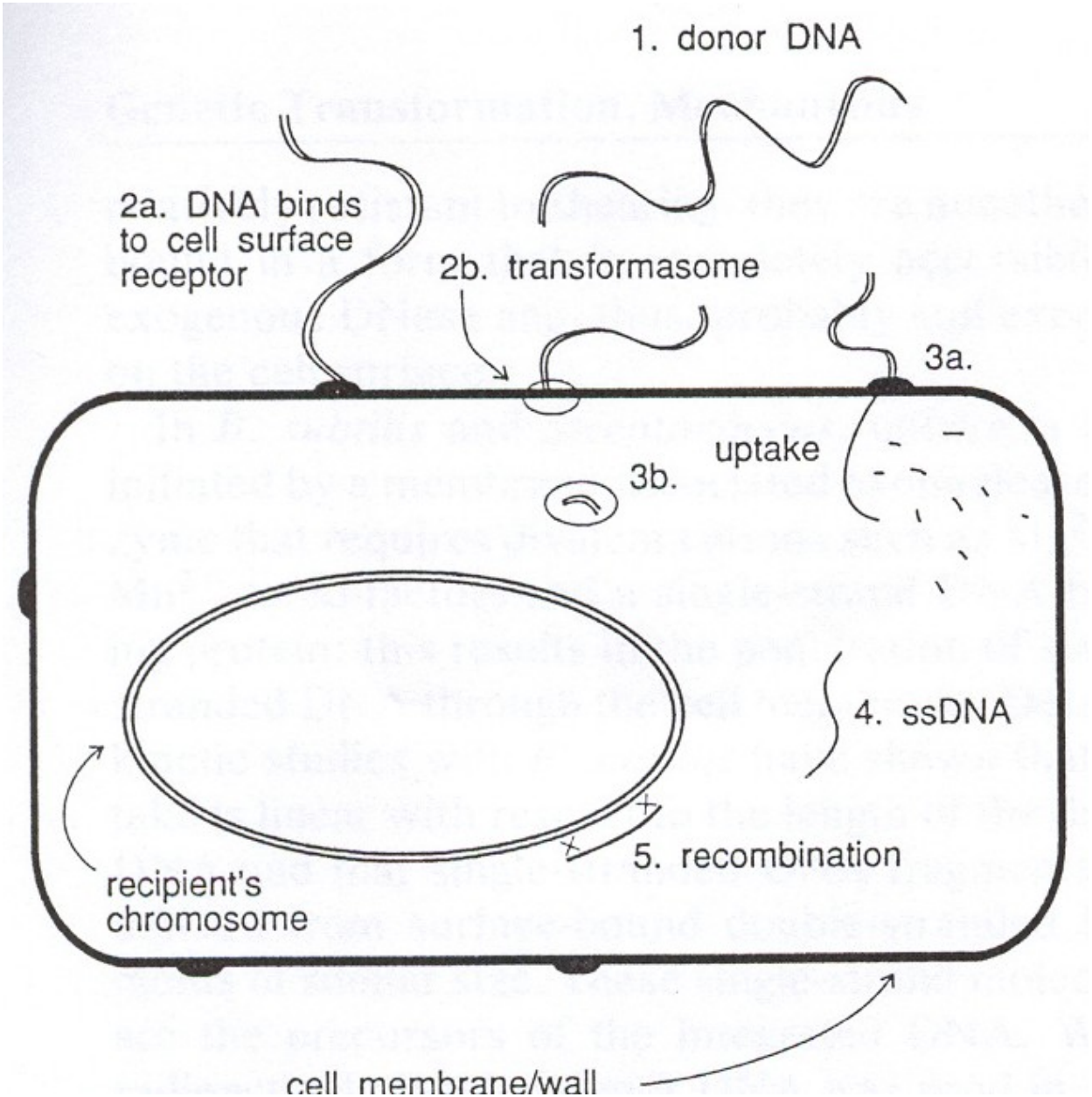
T4 fágok egy E. *coli* sejt felszínén





3–28. ábra. *T3 bakteriofág okozta tarfoltok (plakkok) baktériummal előzőleg sűrűn benőtt agarlemezen, petricsészében [C. S. Gowans szívességéből]*

Természetes genetikai transzformáció



Kompetencia

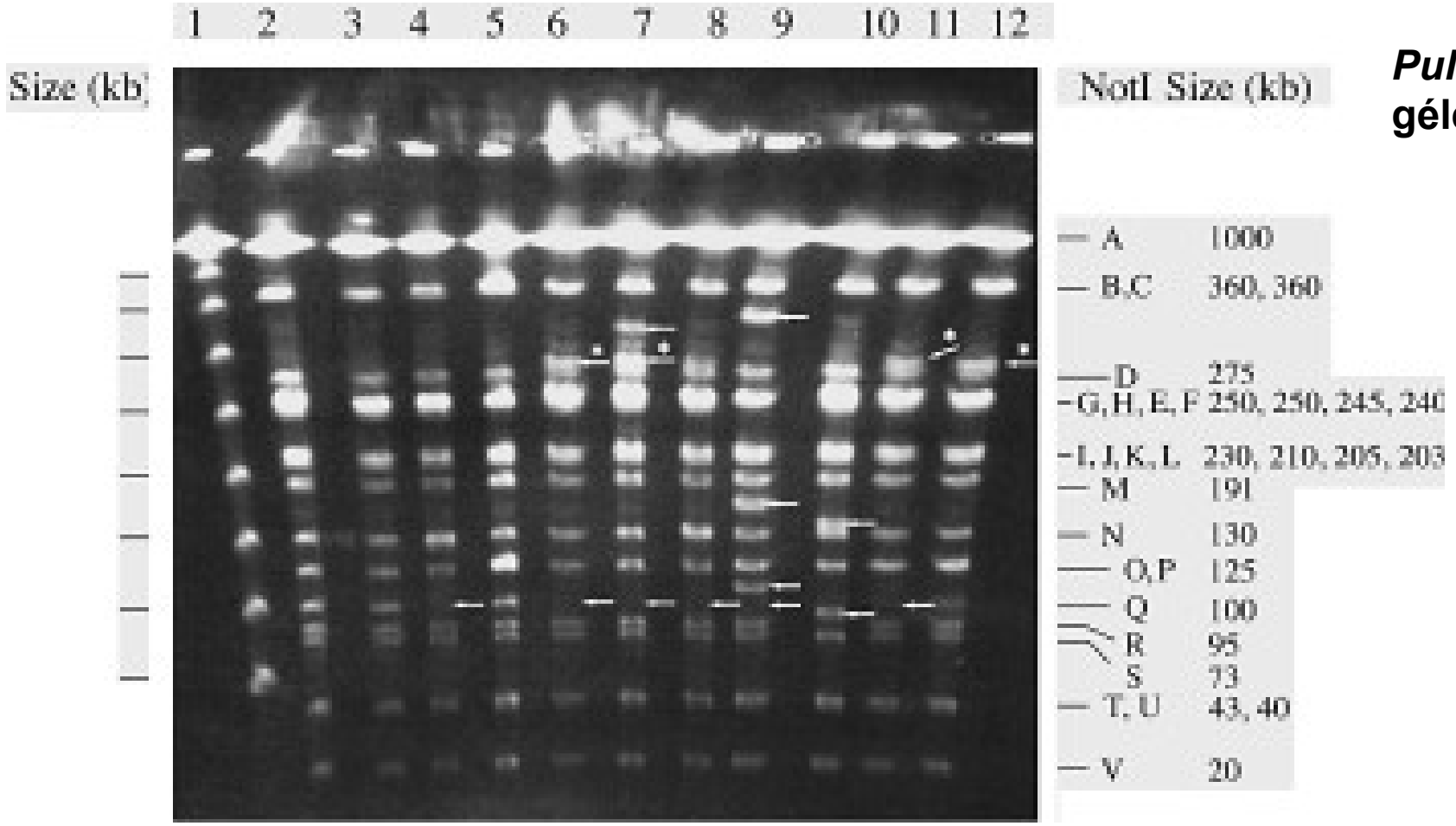
(fiziológiás/mesterséges)

Természetes Genetikai Transzformáció

- **Kompetencia**
- **Aktív (energiaigényes) folyamat**
- **Patogenitási faktorok (*Streptococcus pneumoniae*)**
- **Szorosan kapcsolt gének együtt transzformálódnak - térképezés**
- **Kapcsoltság mértéke: kettős transzformációk / összes transzformáns**

Faj	Élőhely	DNS felvétel	Szabályozás
Streptococcus pneumoniae fázis	Respirációs traktus	Nem specifikus	Nem stacioner
Bacillus subtilis	Talaj	Nem specifikus	Tápanyag jelenléte
Neisseria gonorrhoeae	Genitáliák	Szekvencia specifikus	Konstitutív
Haemophilus influenzae	Respirációs traktus	Szekvencia specifikus	Éhezés indukálja

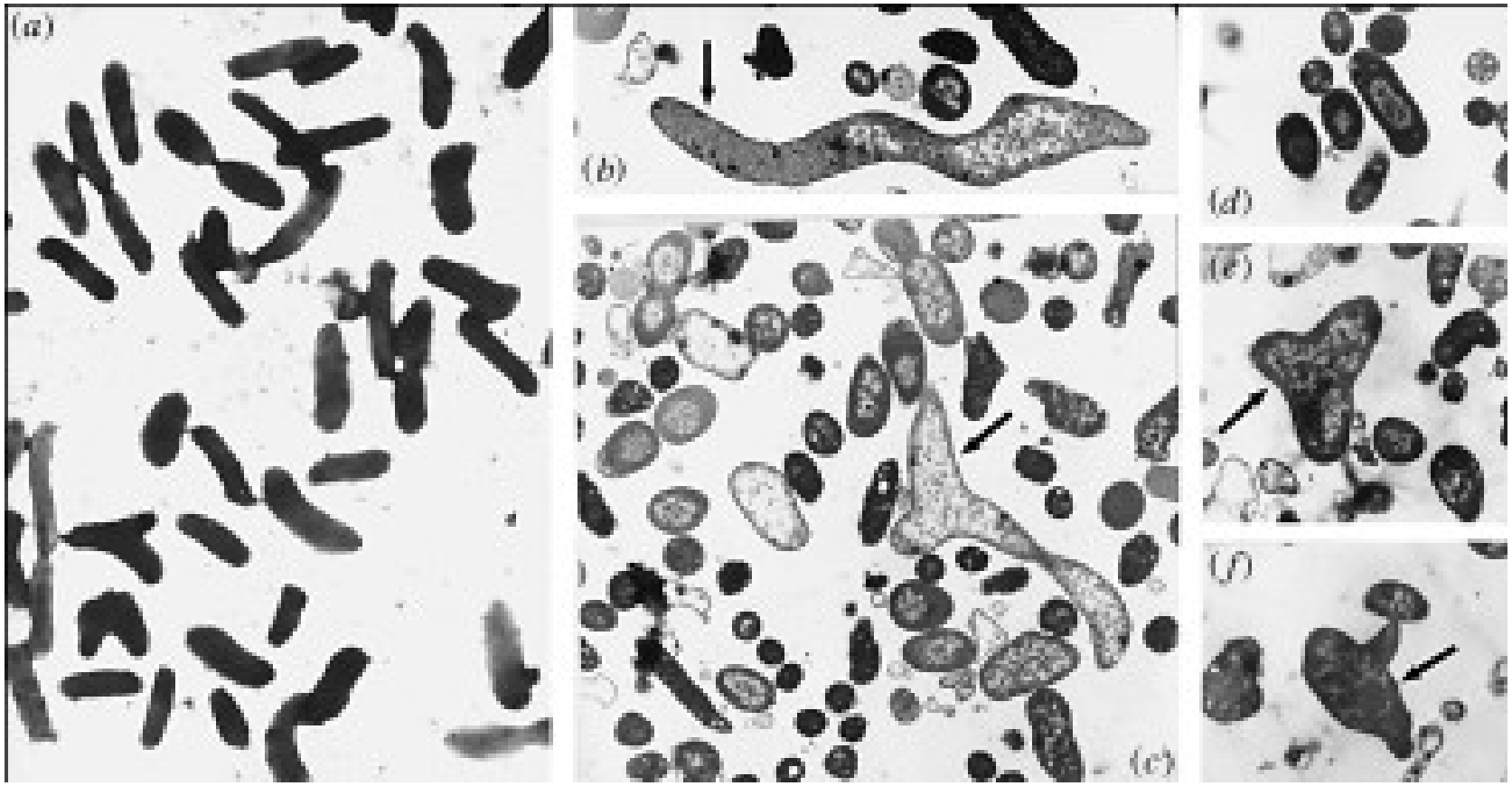
Bakteriális genom: „közösségi” genom szerveződés



Pulsed field **gélelfo.**

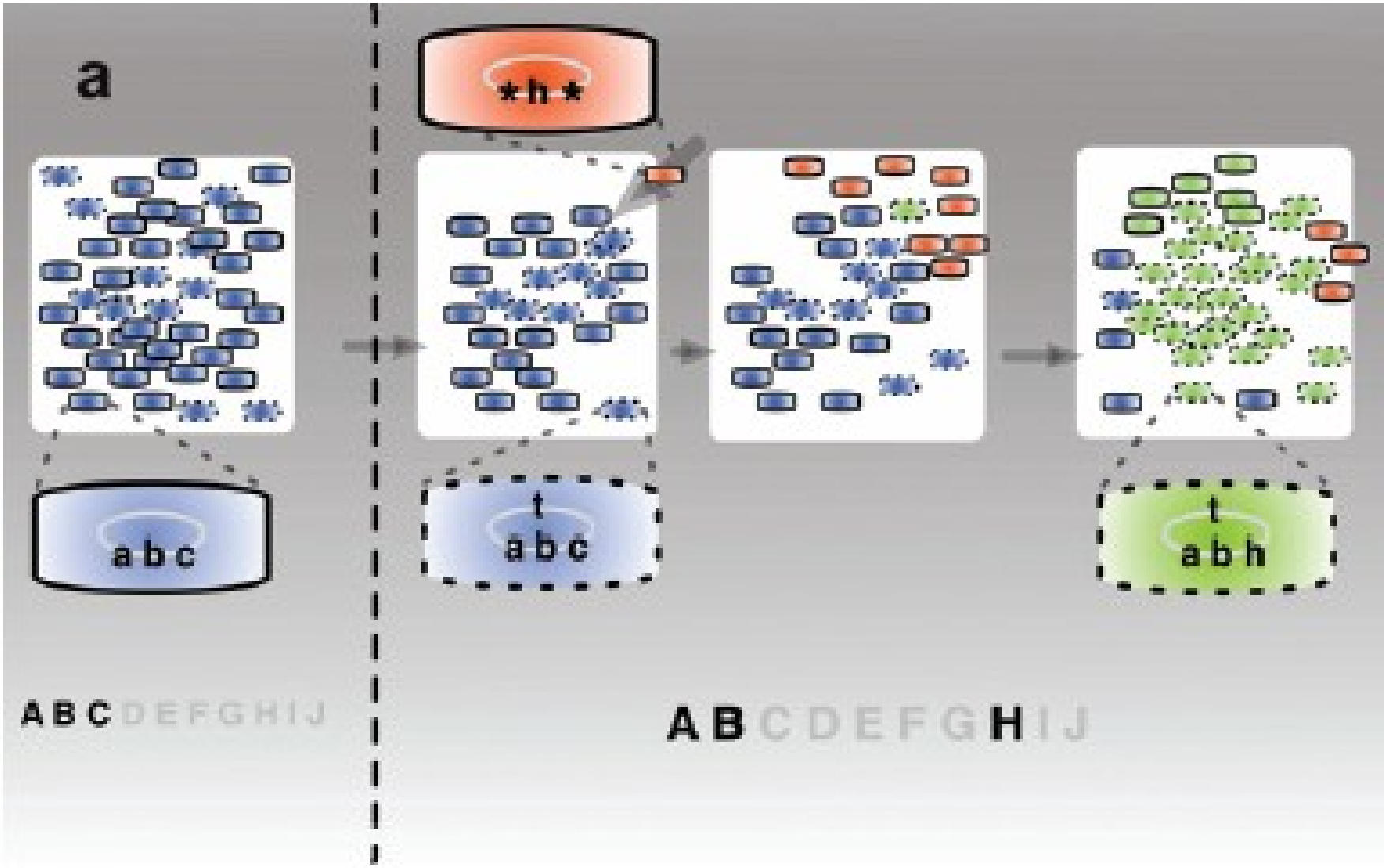
E. coli **tenyészet (egyedei sejtek) 1 hónapos éheztetés után.**

A természetes baktérium populációk morfológiailag is diverzek

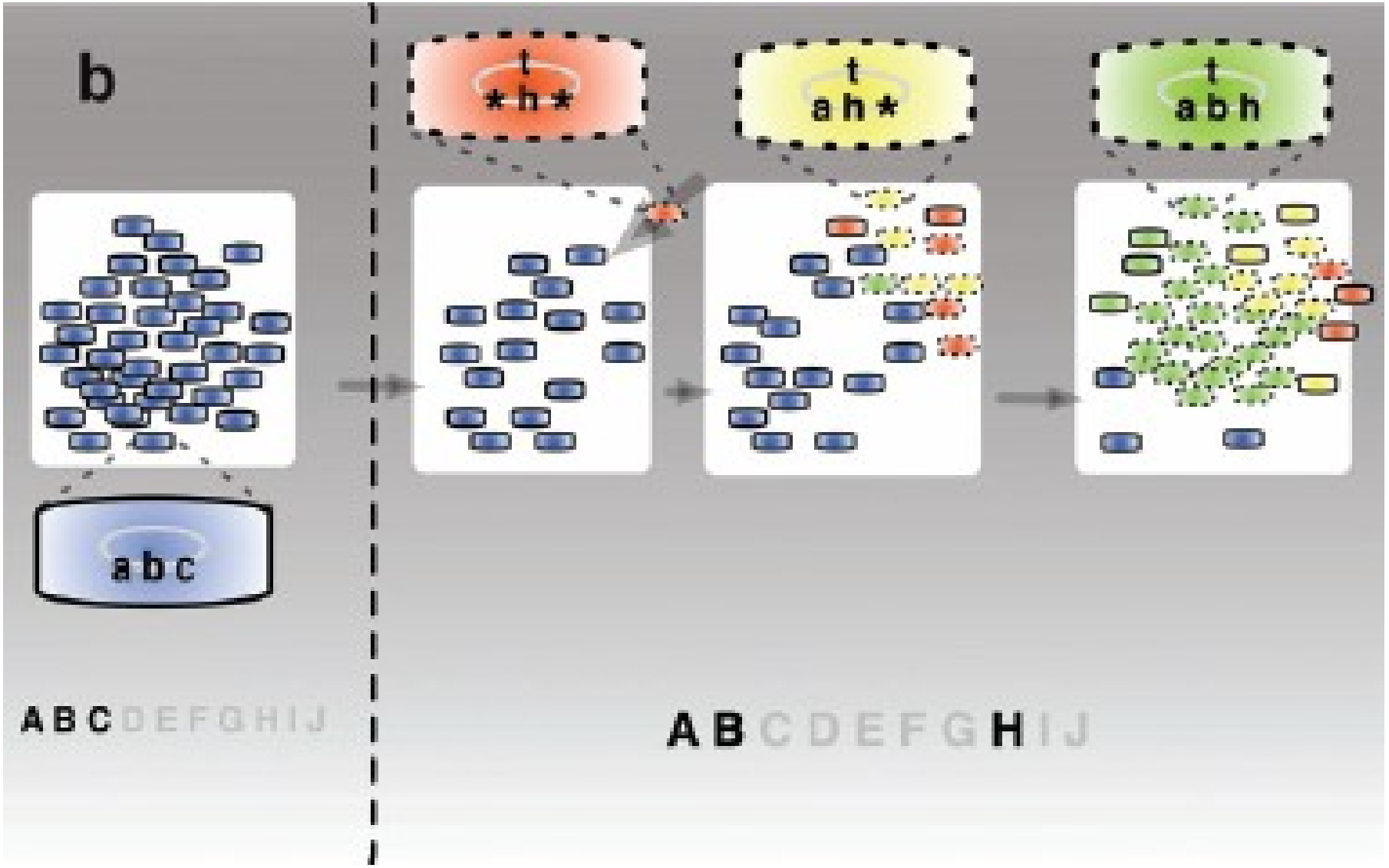


E. coli **tenyészet 1 hónapos éheztetés után.**

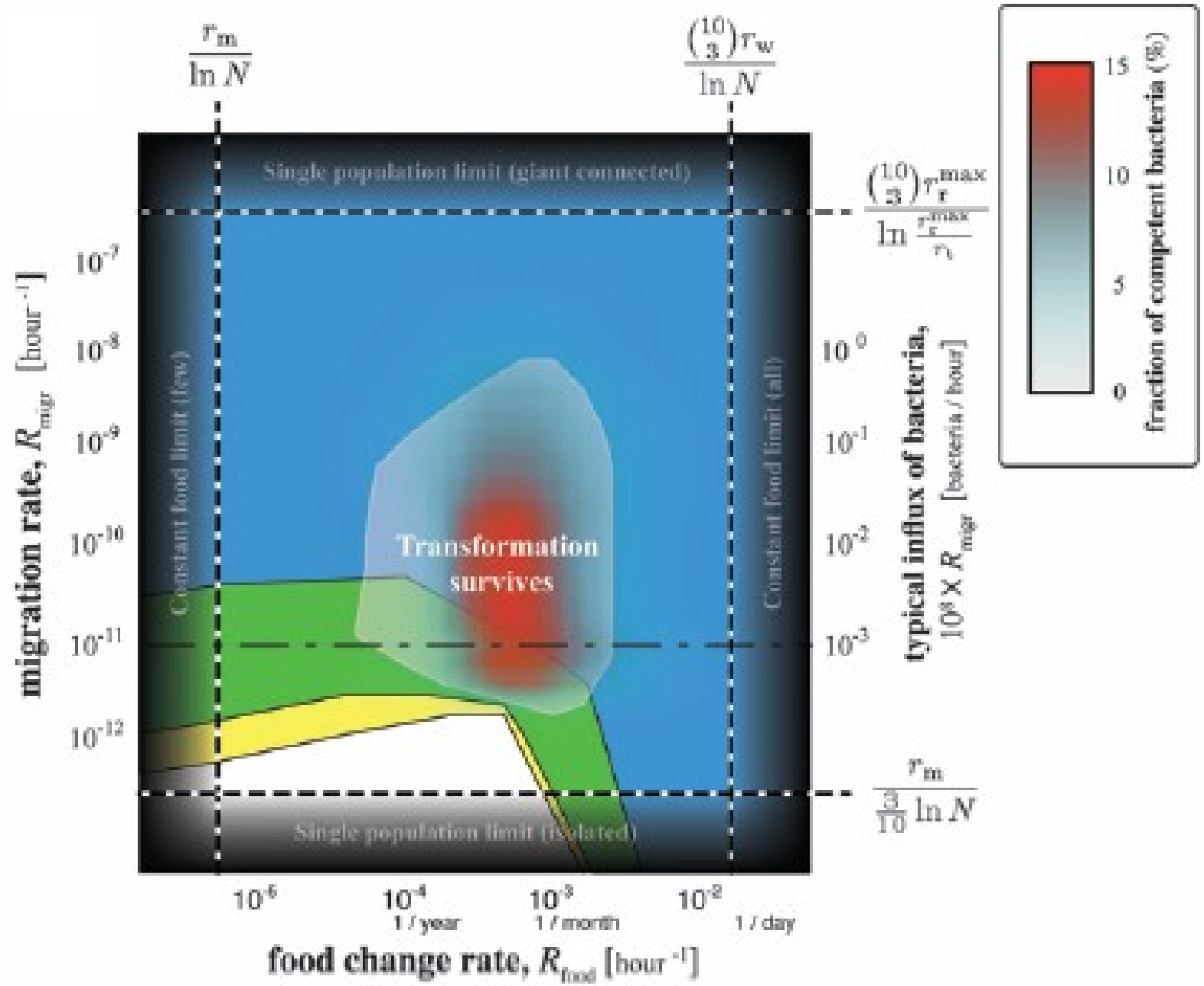
Transzformáció: új (adaptív) szekvenciák felvétele más populációkból



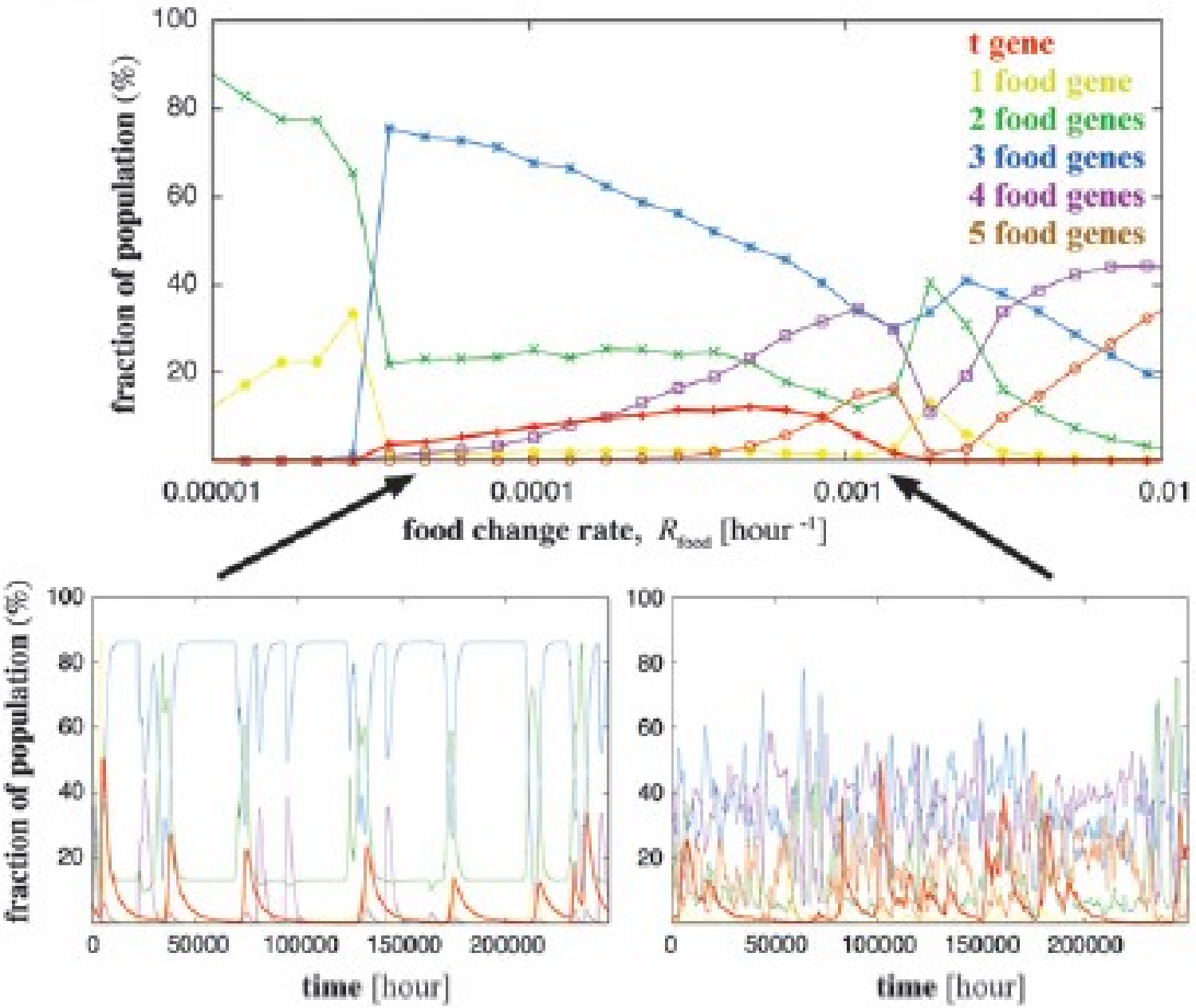
Transzformáció: új (adaptív) szekvenciák felvétele más populációkból



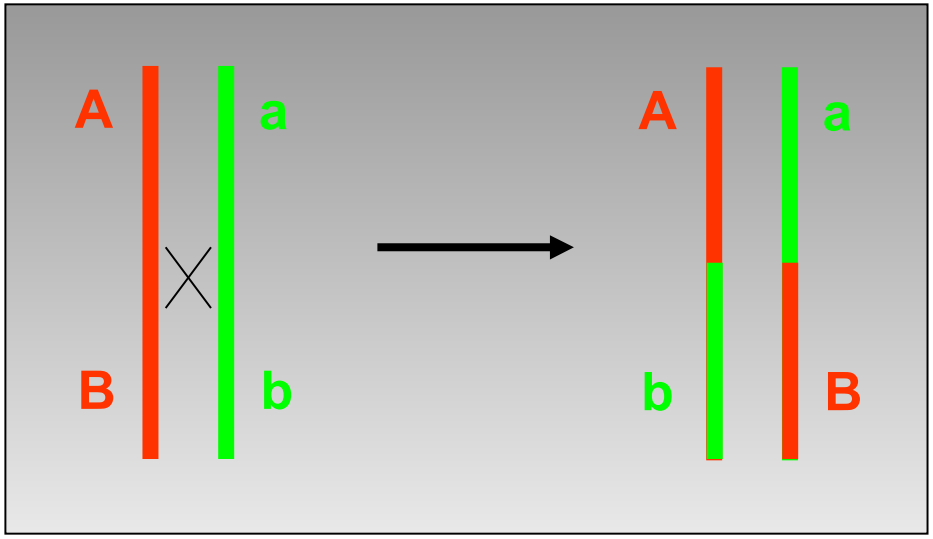
Transzformáció: bakteriális szex



A transzformációs rendszer stabilan fennmaradhat!



Genome evolution in eukaryotes: Recombination - meiotic sex



(incorrect model!)

Two important questions related to recombination:

1, how can non-functional sequences (e.g., microsatellites) evolve?

2, why does meiotic sex (syngamy and meiotic recombination) exist at all?

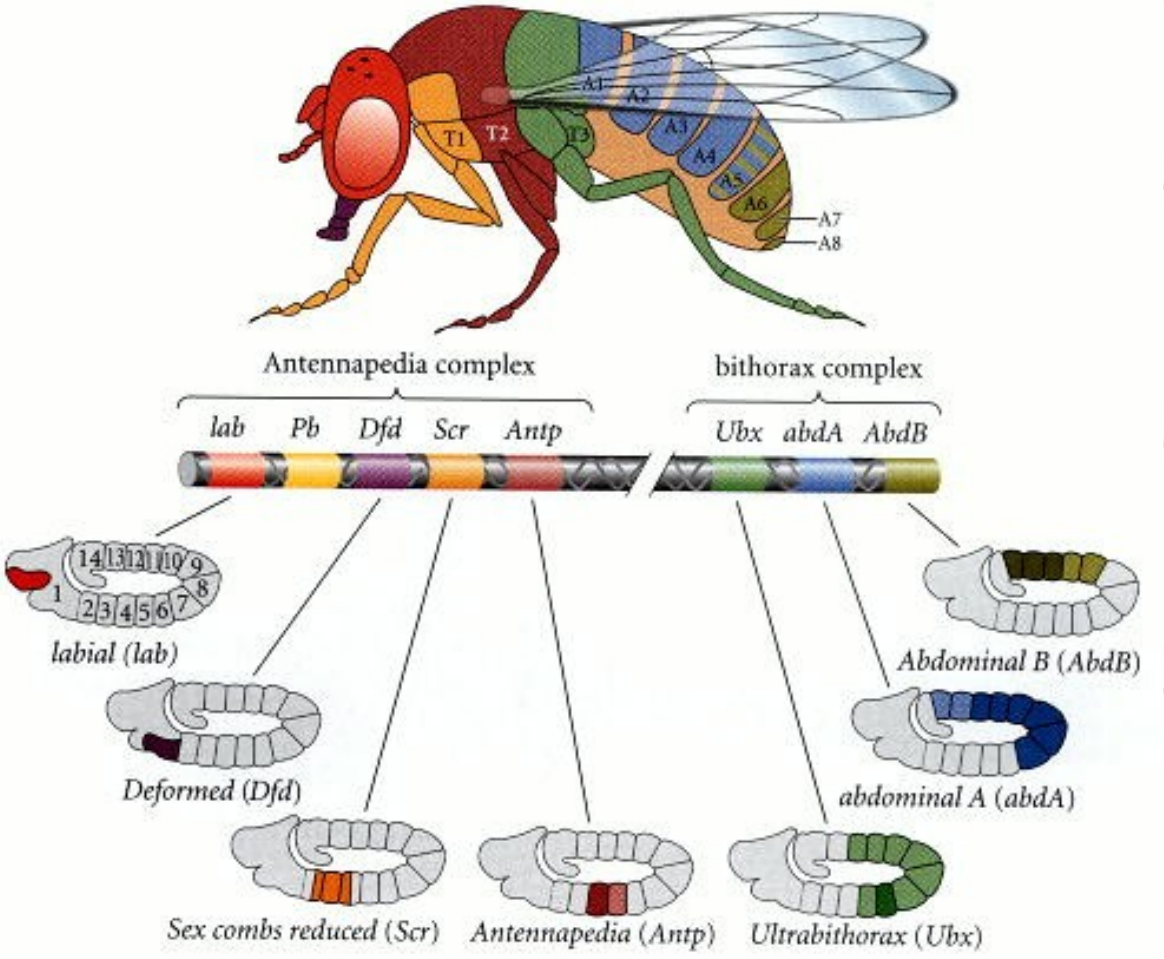
1, Remember: individual (beneficial) mutations arise primarily in individual genomes

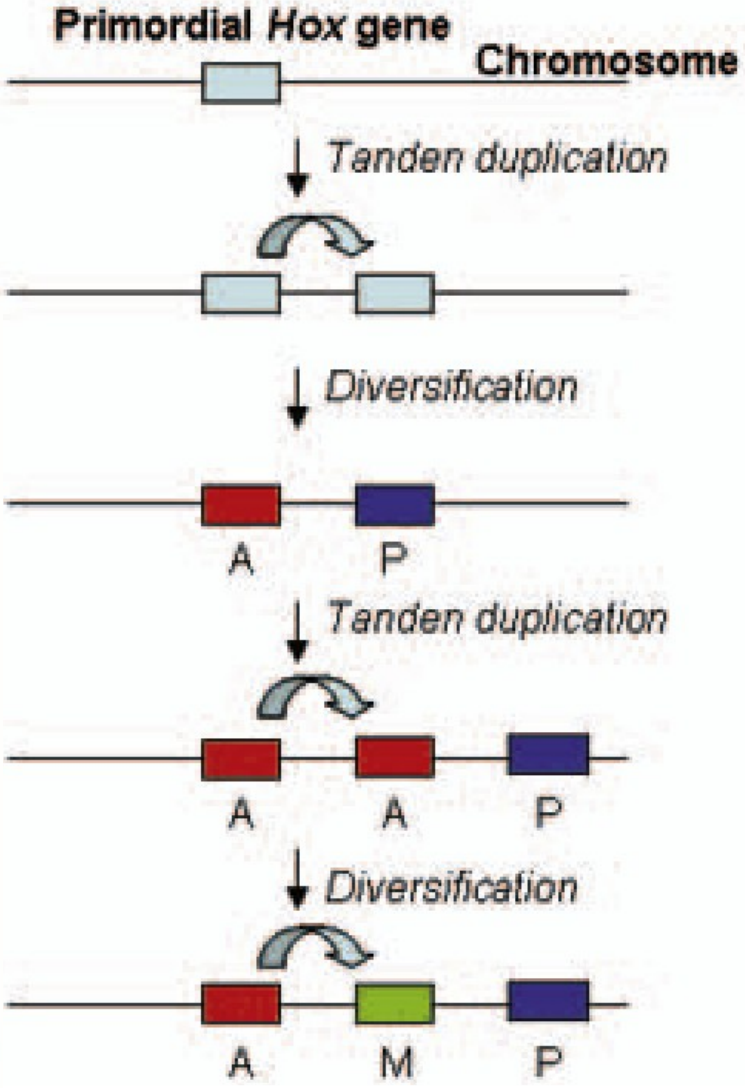
how do they spread within species?

(genome evolution)

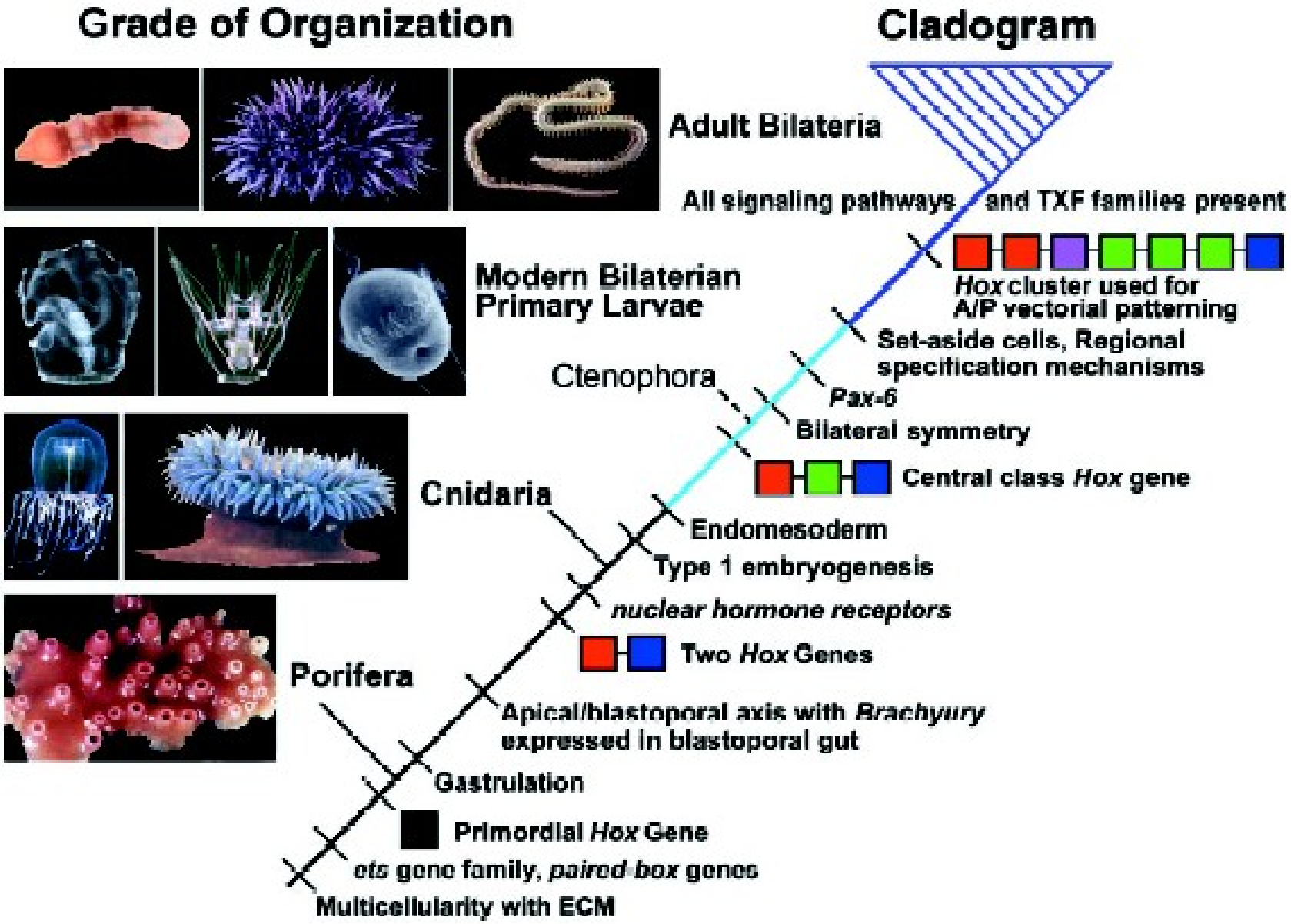
But what about non-functional sequences? (e.g., $AATGCA)_n$

An example:
The evolution of *Hox* clusters





Evolution of functional sequences (e.g., *Hox* genes)



Why does sex exist in nature?

this is a fascinating, fundamental and long-standing problem in biology

- **every week a new paper on the origin of sex**
- **are there asexual organisms at all?**
- **can we recognize sex every time?**
- **twofold cost of sex**

The origin of sexual reproduction

…

The most accepted hypothesis for the origin of sex:

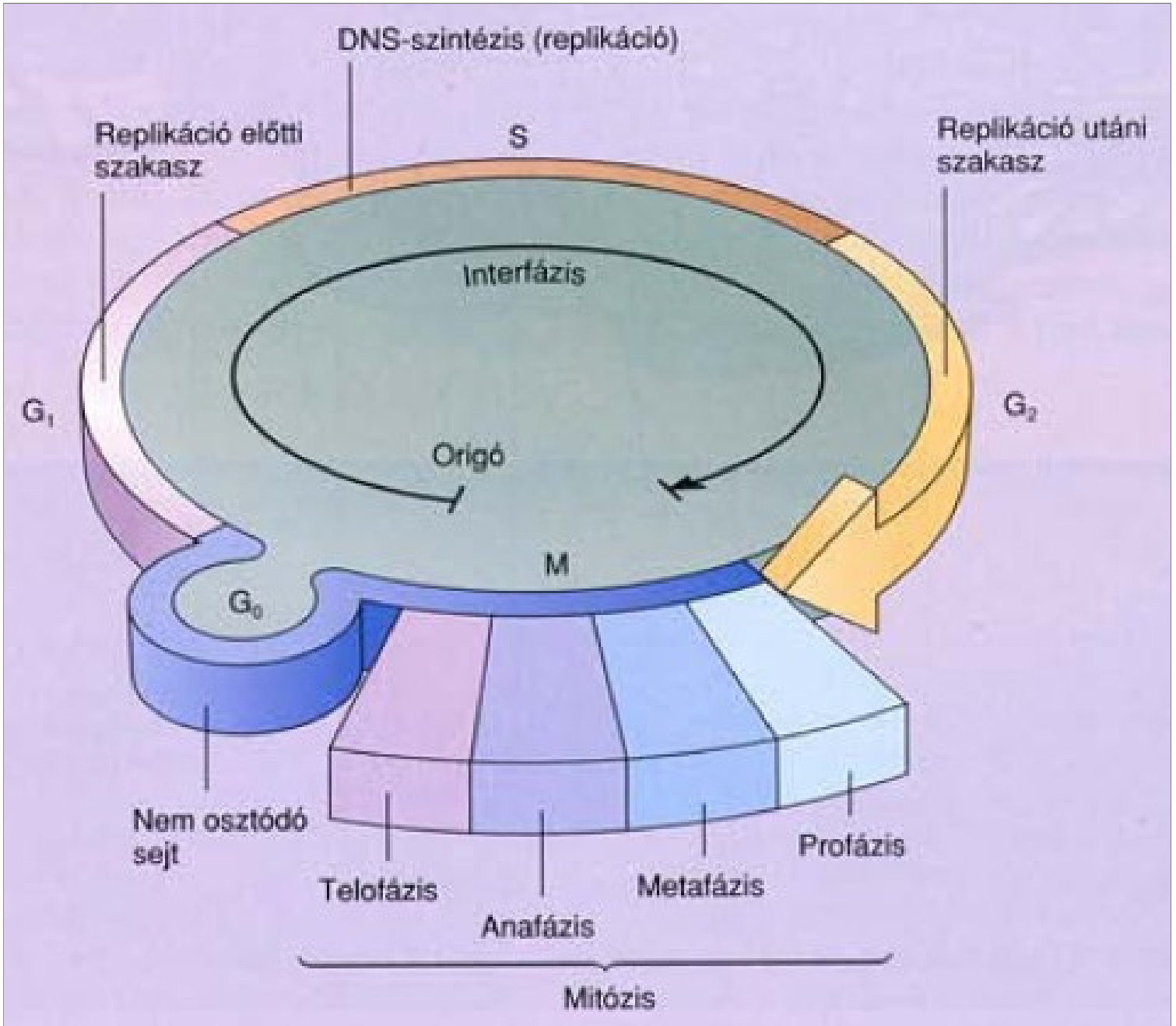
Mutational deterministic hypothesis

An increased efficiency of selection against synergestically interacting deleterious mutations (Kondrashow 1982; Crow 1983)

„It is postulated that sex provides the advantage of generating novel, adaptive gene combinations and/or preventing the accumulation of deleterious mutations.
However, these advantages rely upon selection between populations (group selection) to maintain sex and do not address the short-term advantage of sex to individuals, an advantage that appears necessary if sex is to be maintained long enough for such group selection to operate…"

Goodwin *et al*. 2003 *Science*

Cell cycle


DNS-szintézis (replikáció)
Replikáció előtti szakasz
S
Replikáció utáni szakasz
Interfázis
G1
G2
Origó
M
G0
Nem osztódó sejt
Telofázis
Anafázis
Metafázis
Profázis
Mitózis

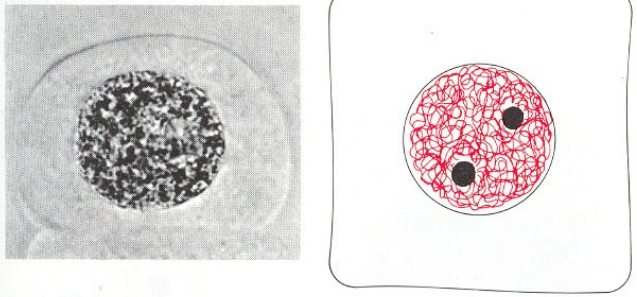
Mitosis

Prophase

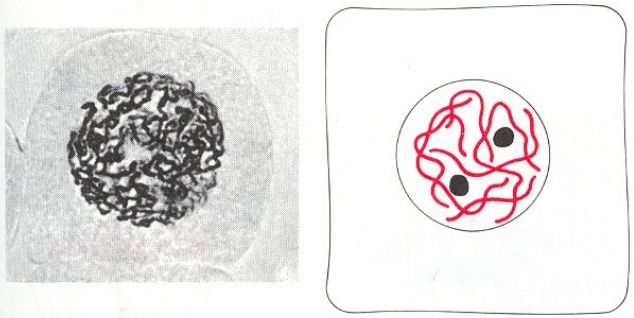
Metaphase

Anaphase

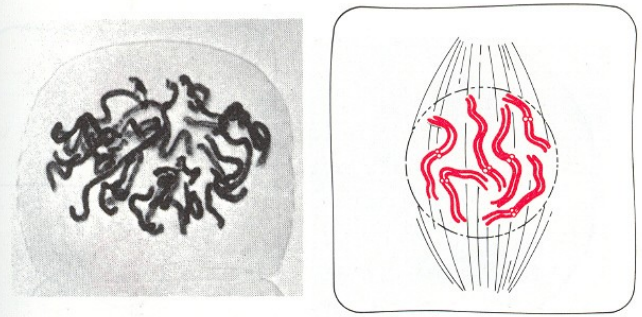
Telophase



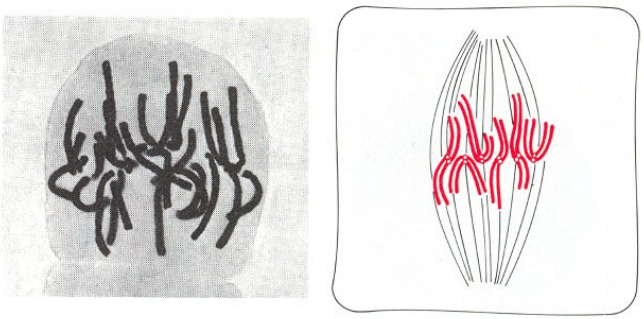
interphase



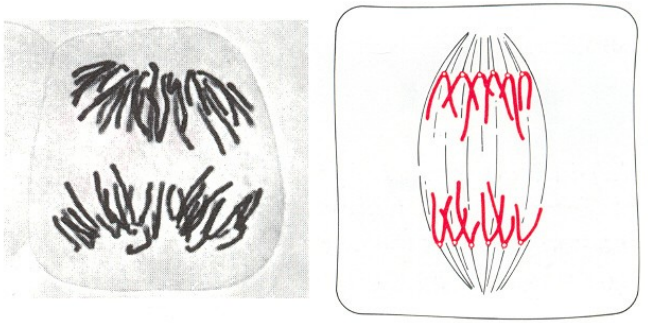
early prophase



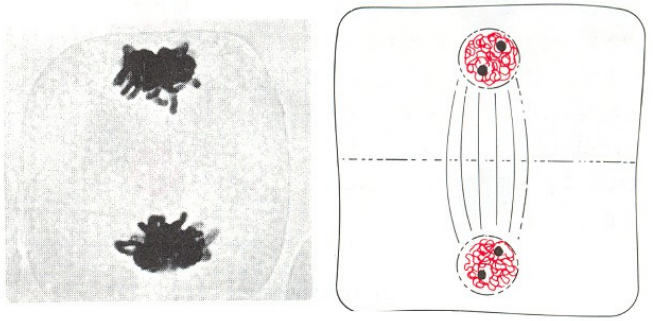
late prophase



metaphase



anaphase



telophase

Meiosis

Prophase I.

- Leptotene
- Zygotene

(homologous pairs, synaptonemal complex)

- Pachytene

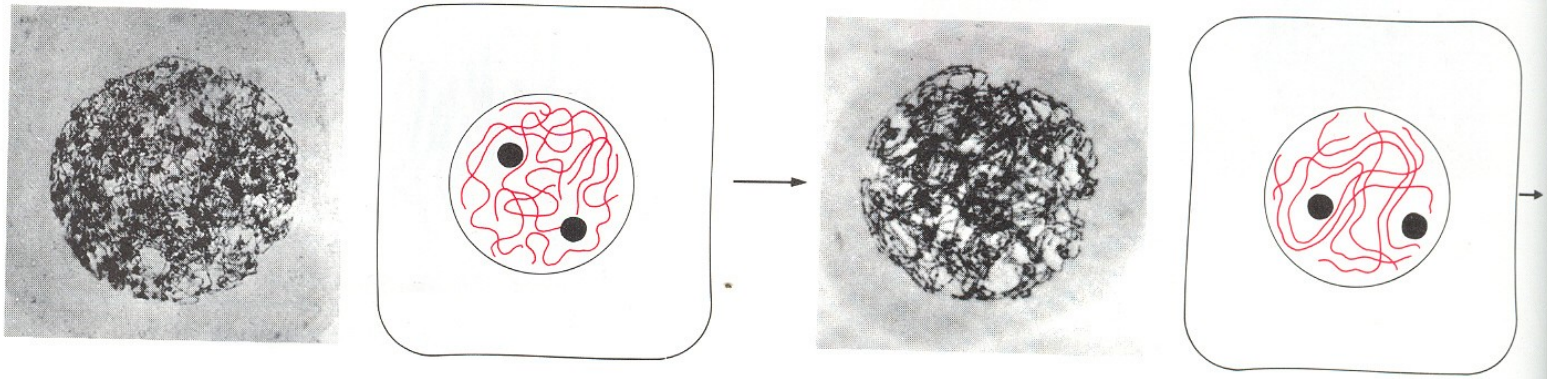
(chiasmata, crossovers)

- Diplotene
- Diakinesis

Metaphase

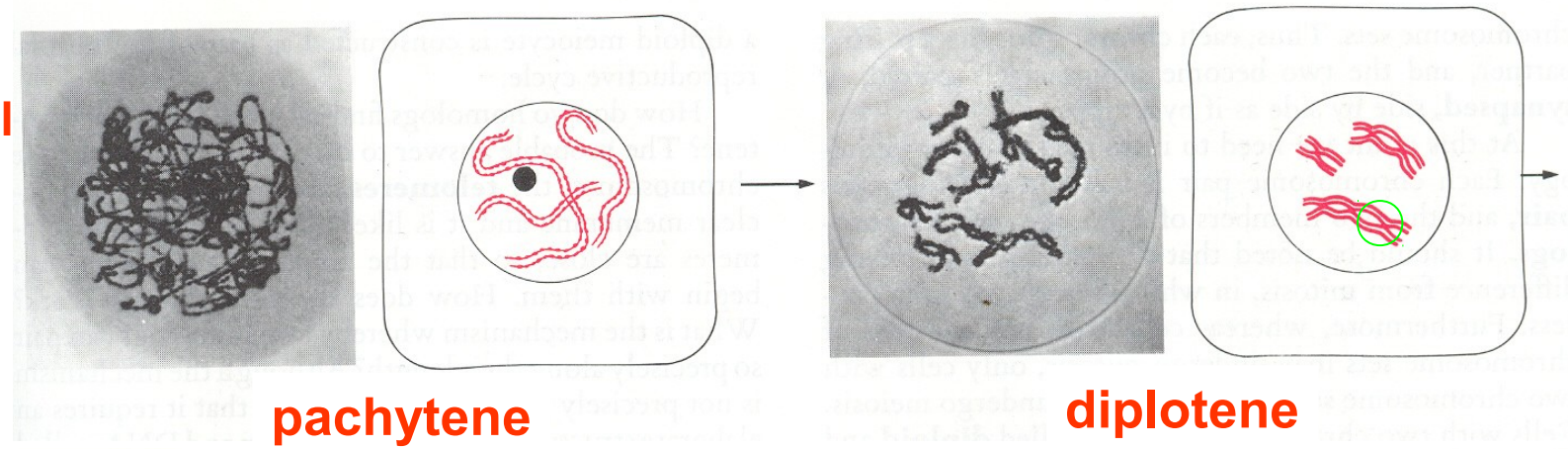
Anaphase

Telophase



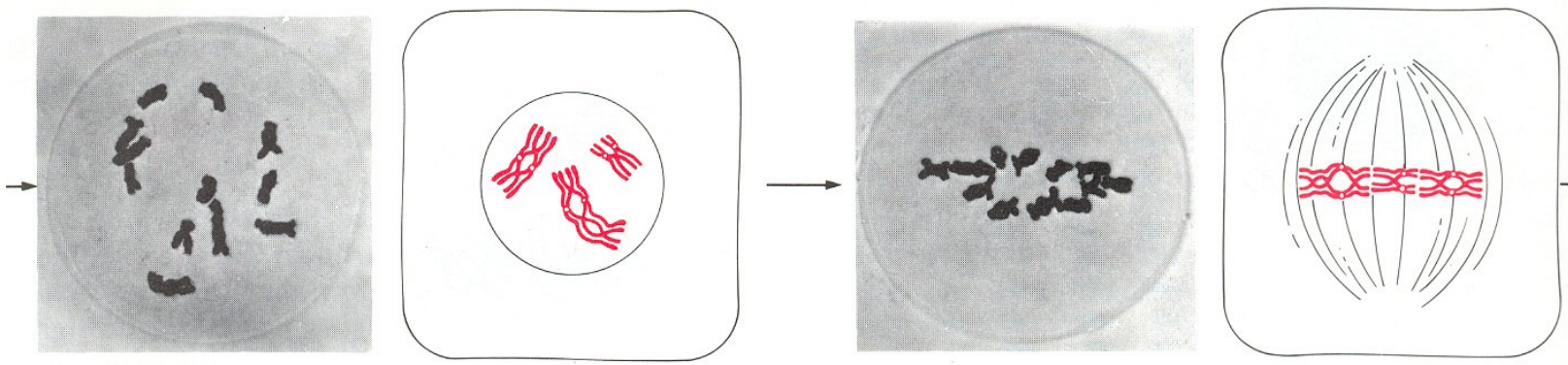
leptotene

zygoten



pachytene

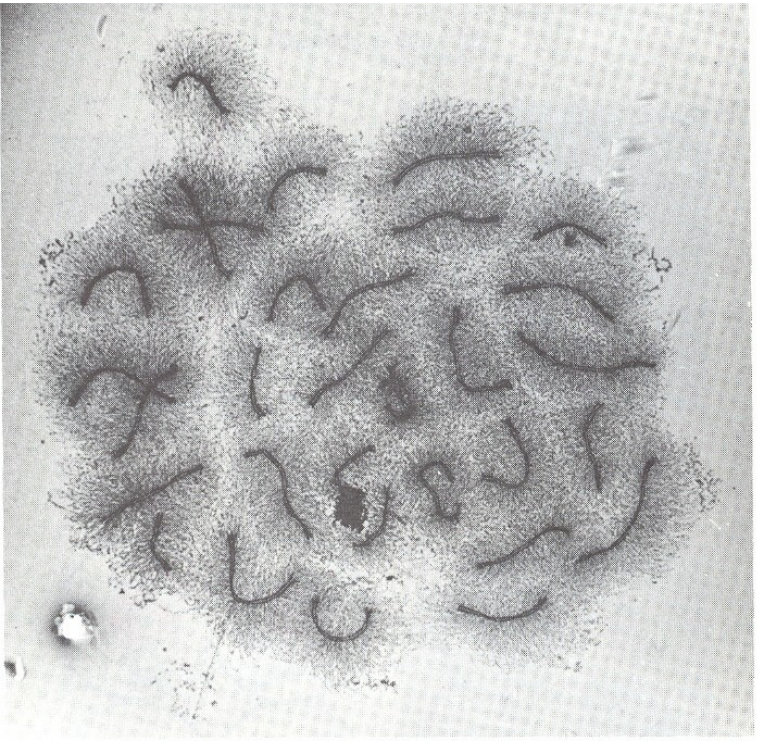
diplotene



diakineses

(f) Metaphase I

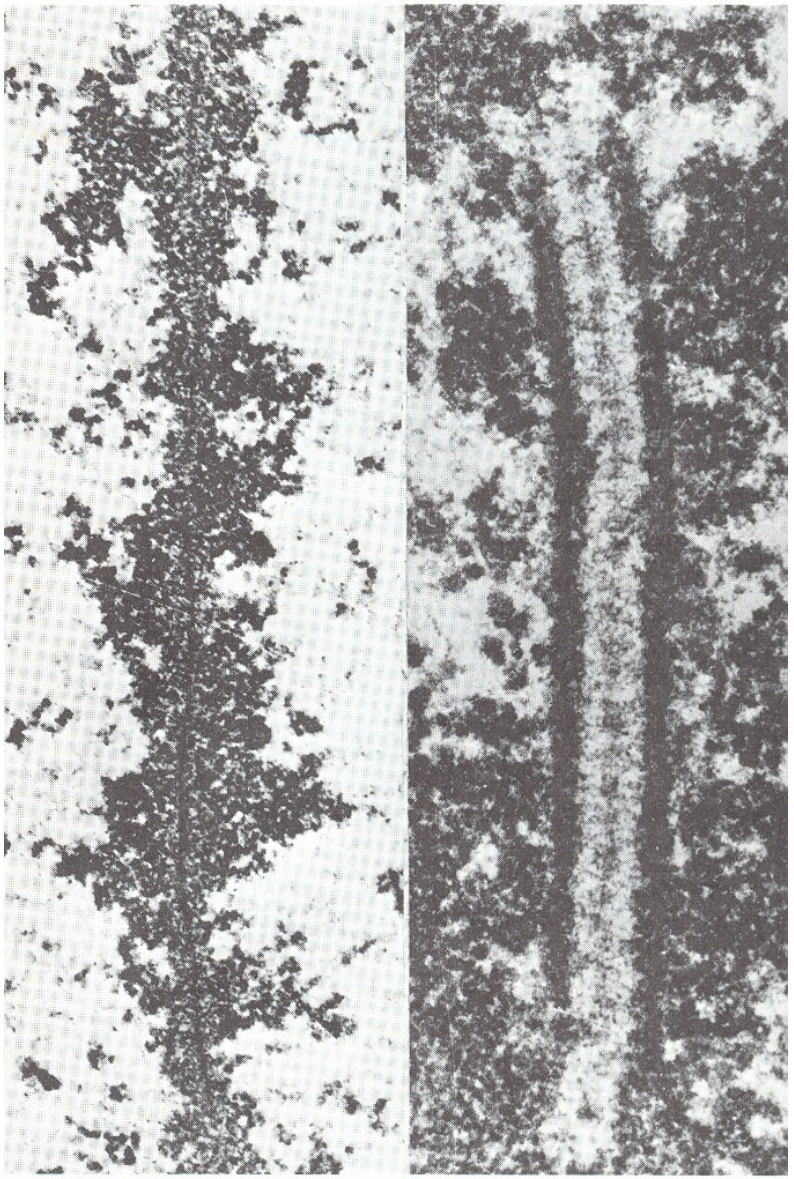
Synaptonemal complex



Hyalophora cecropia

2n = 62, SC = 31

Synaptonemal complex



Lilium tyrinum

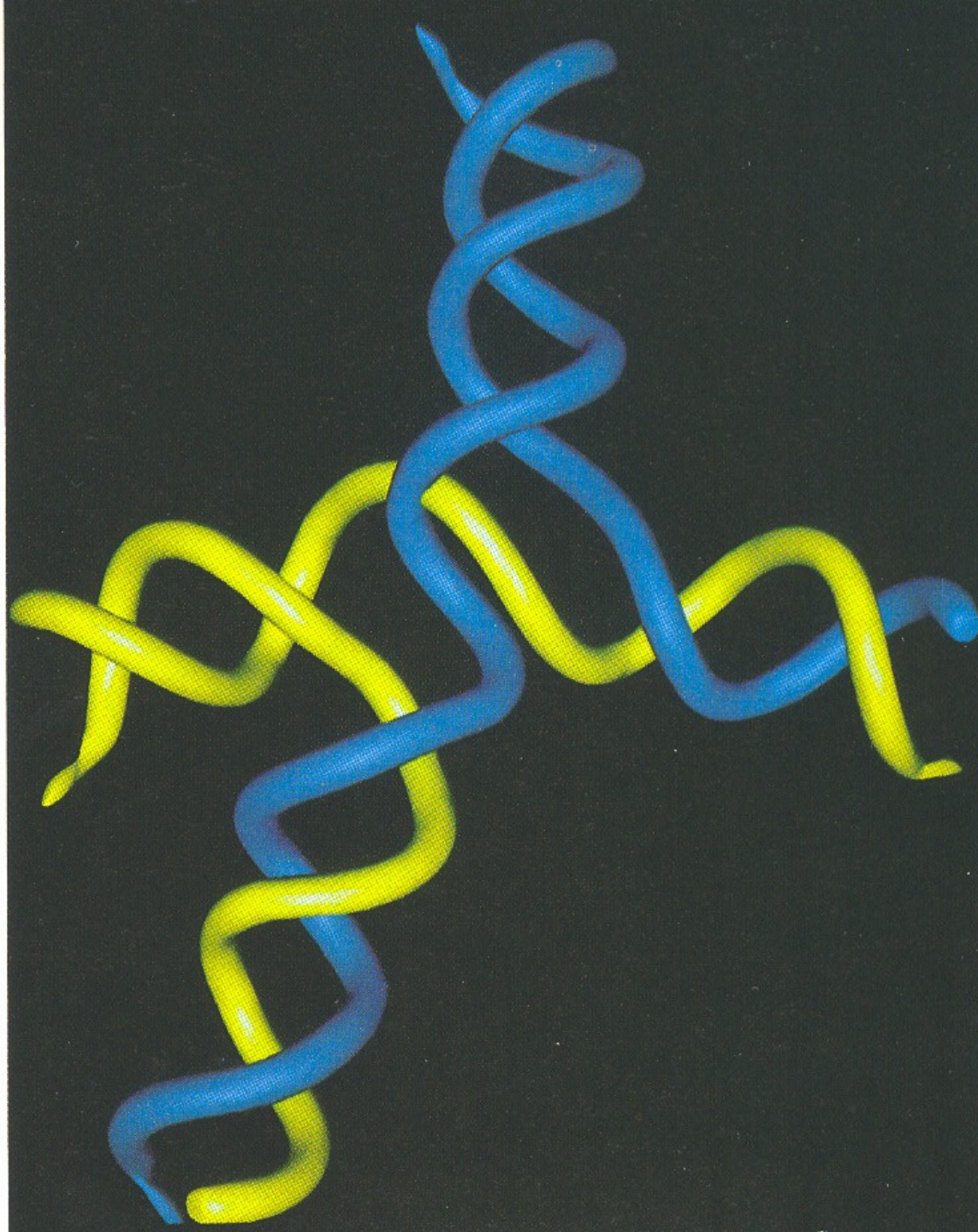
2n = 62, SC = 31

Revolution in our understanding how meiotic recombination occurs (1980-2007…)

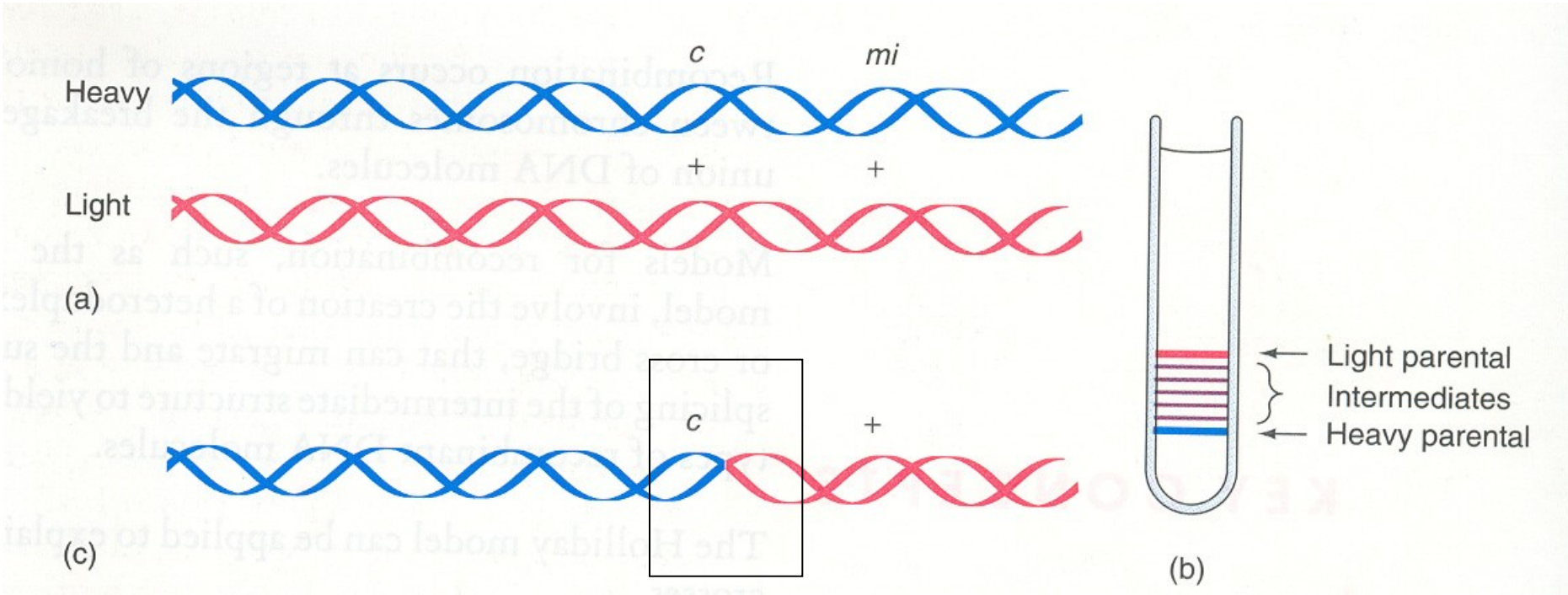
homology search precedes double-strand DNA break that induces meiotic recombination!!!

Recombination

(sex occurs as a result of fertilization and meiotic recombination)



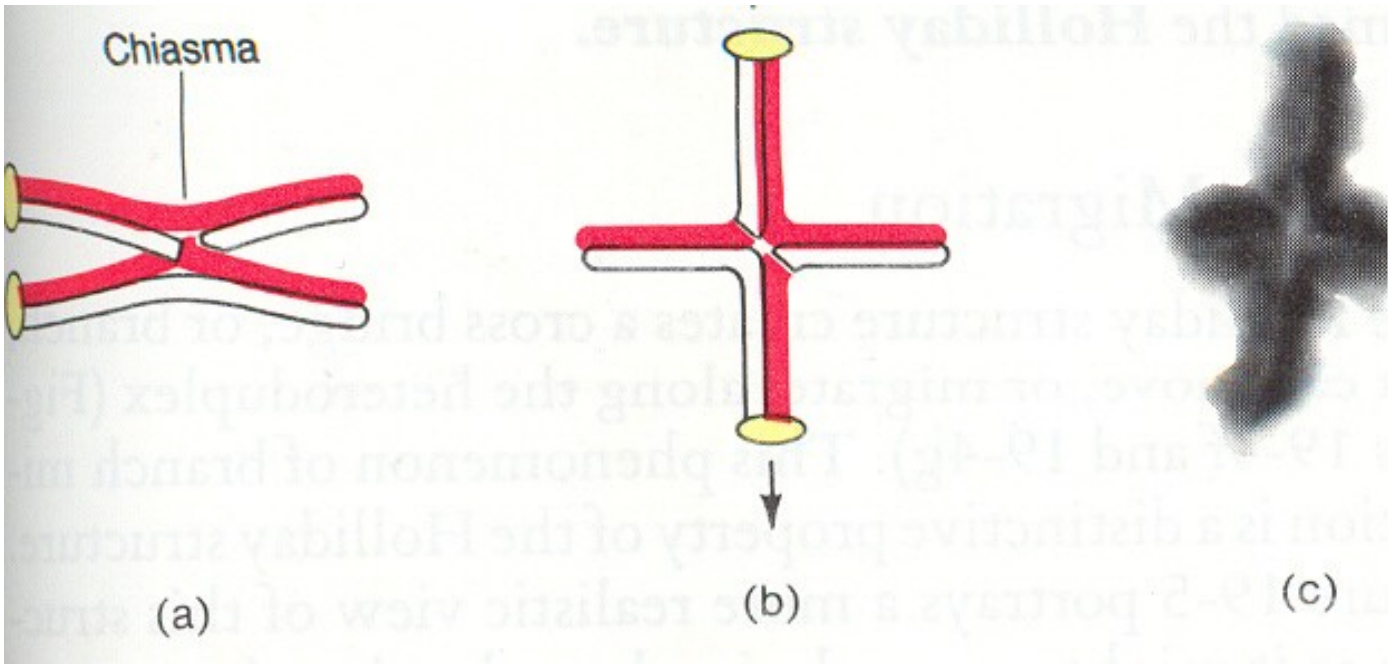
Chromosome breakage and reunion in λ phage



Point-like event

Meselson and Weigle, 1961

Chiasmata: the crossover points



Crossing-over between dark- and light-stained non-sister chromatids

chromatid

chromatid

The Holliday model:

Enzymatic cleavage

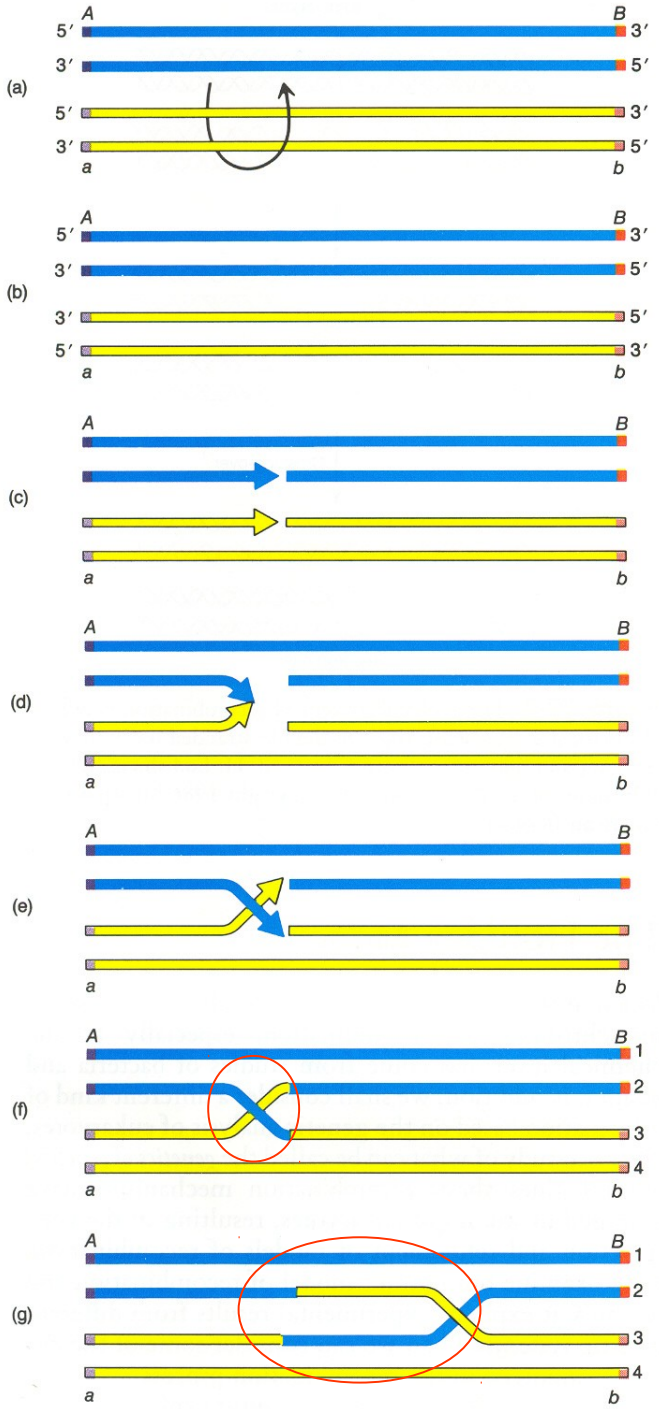
Creation of heteroduplex DNA

/Holliday structure/

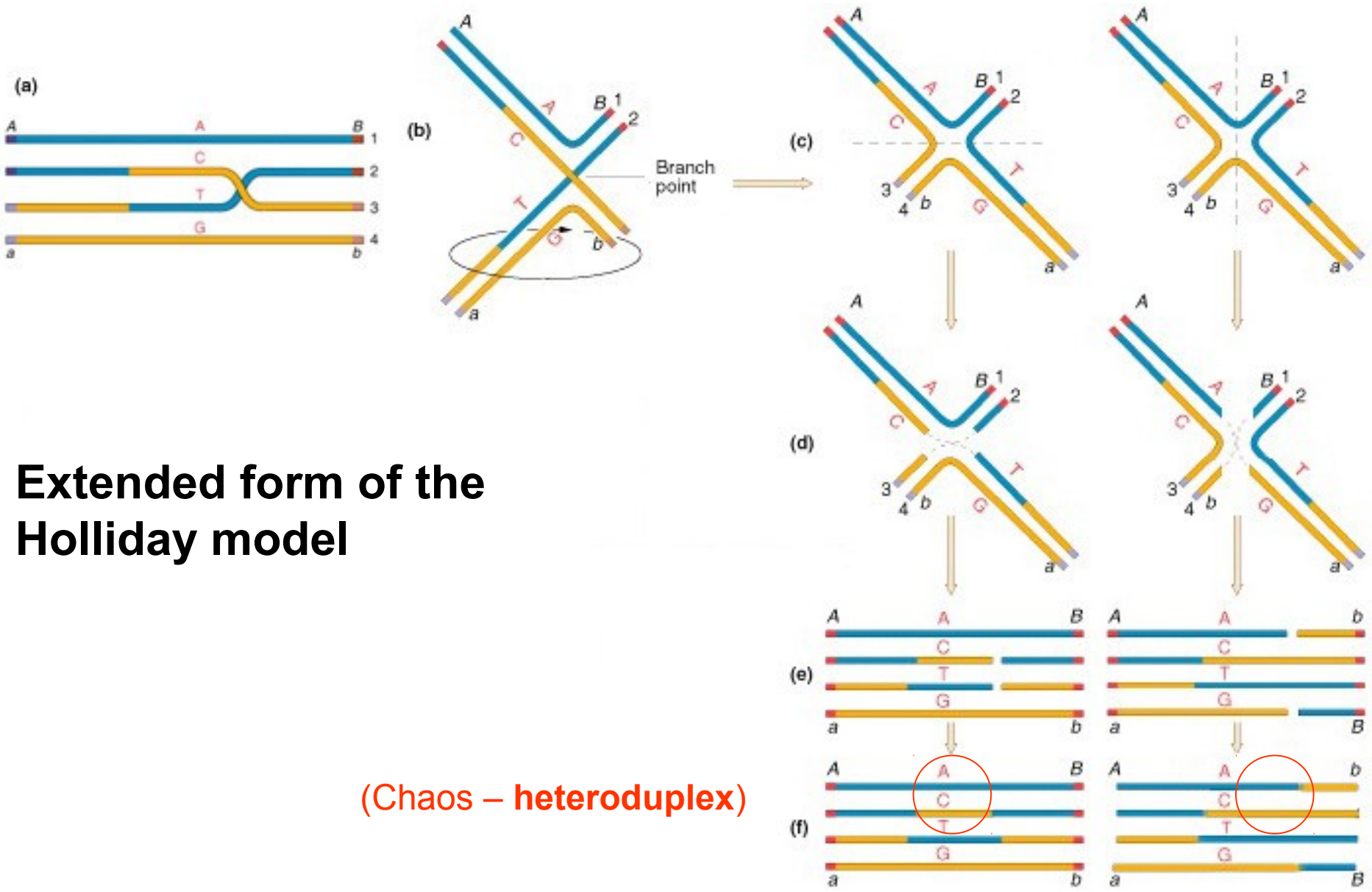
Branch migration

Resolution of Holliday structure

Recombination is not a point-like event



Extended form of the Holliday model

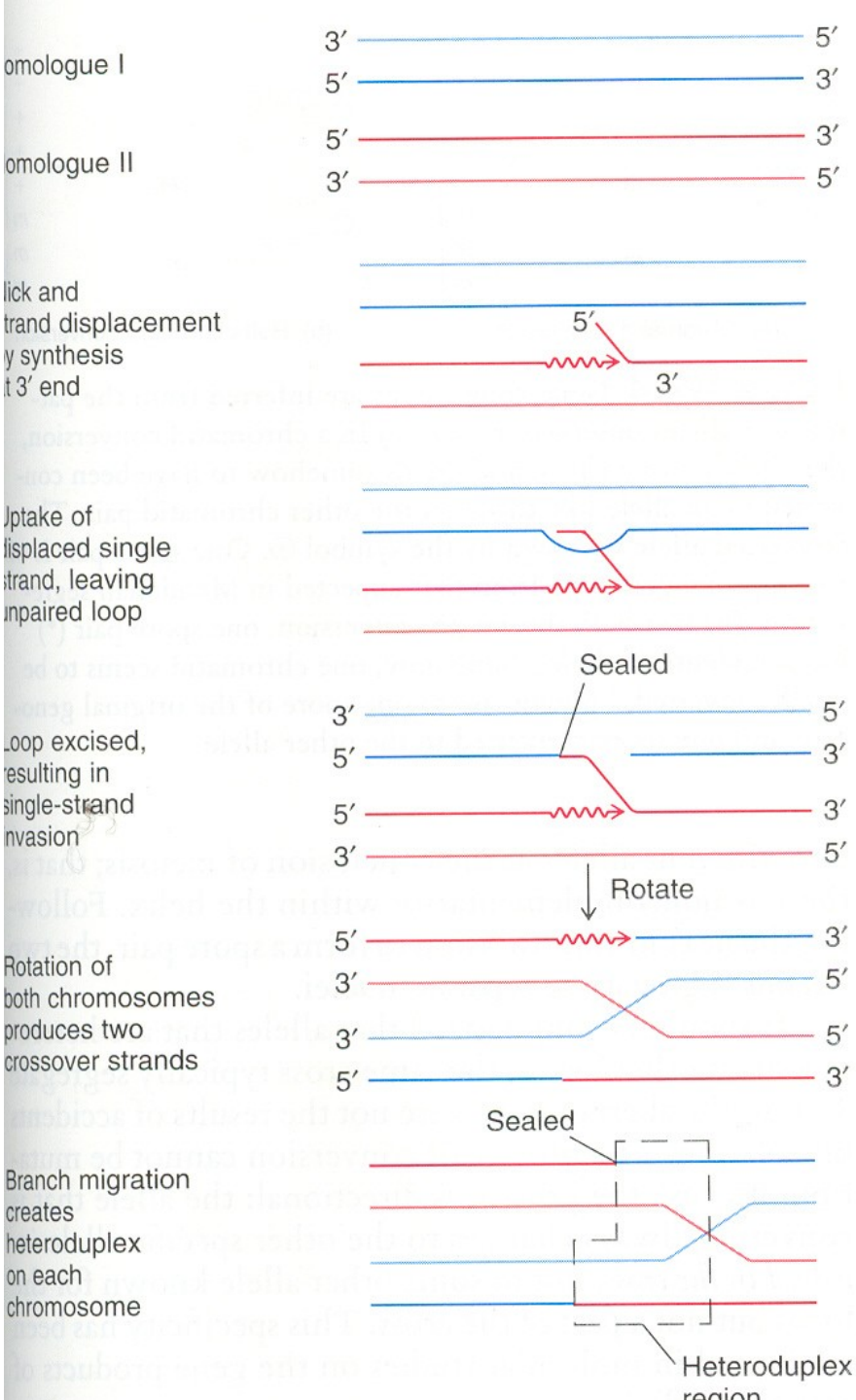


(Chaos – **heteroduplex**)

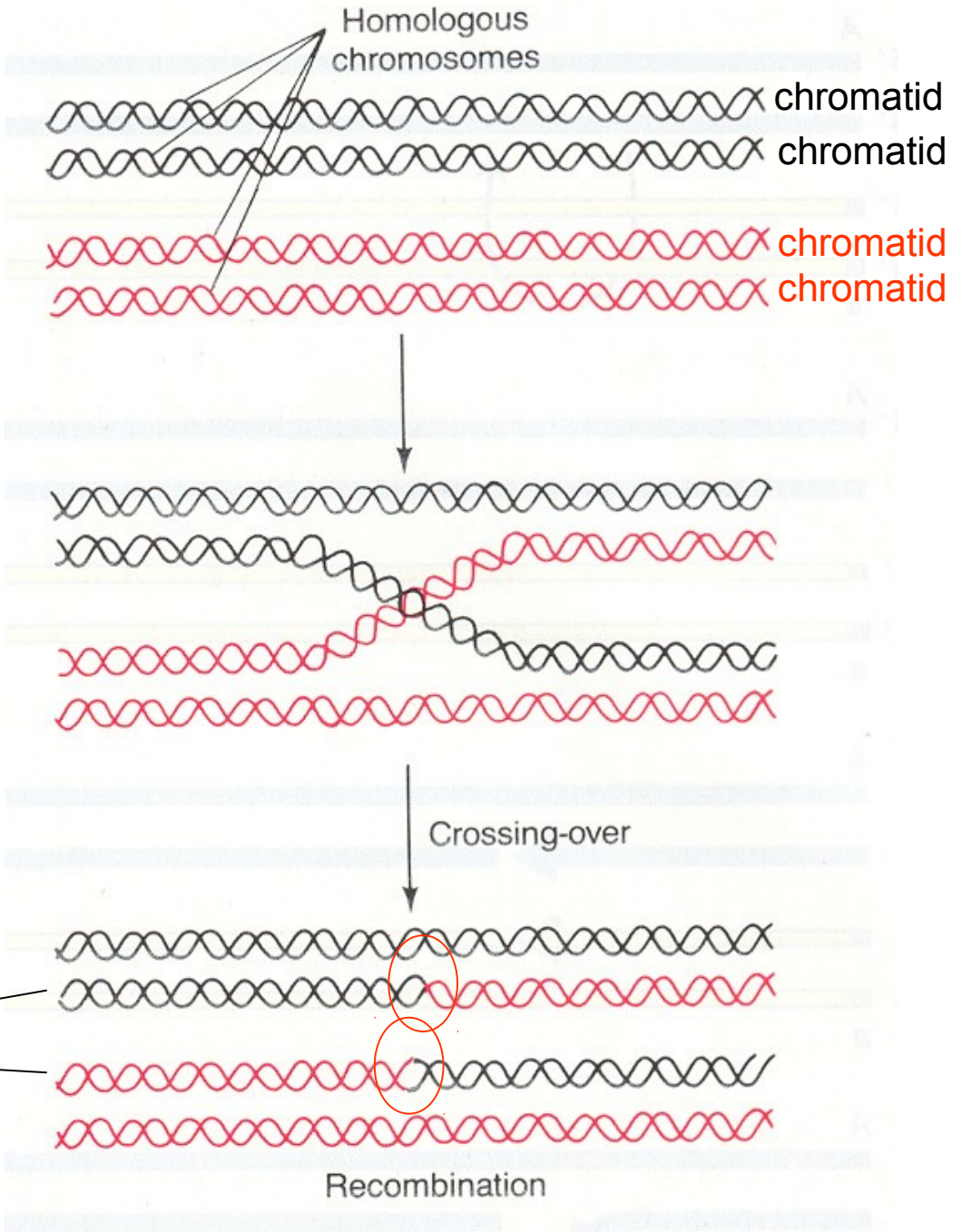
The Meselson-Radding model

Branch migration

(heteroduplex)



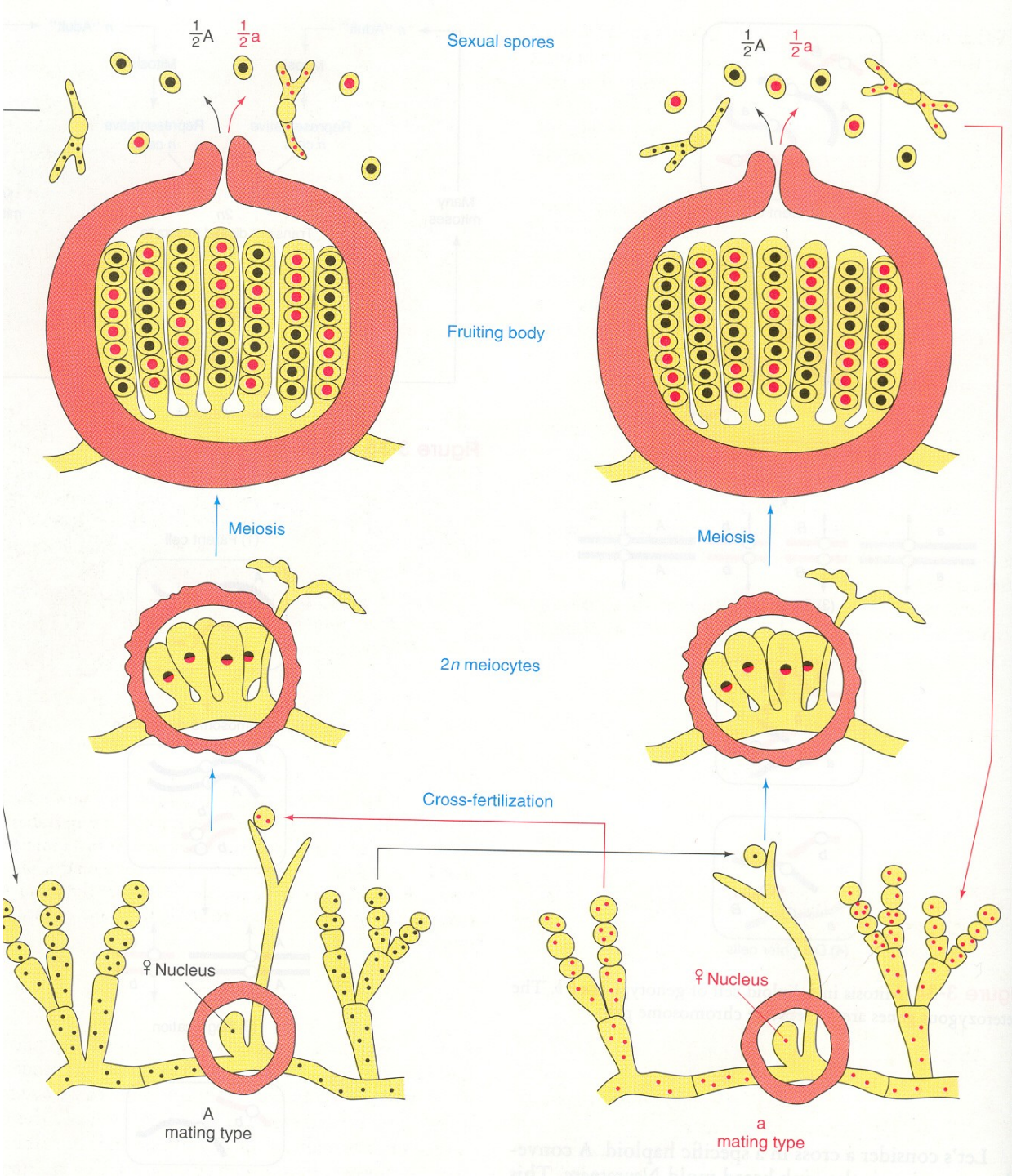
Recombination: double-strand DNA break and rejoining



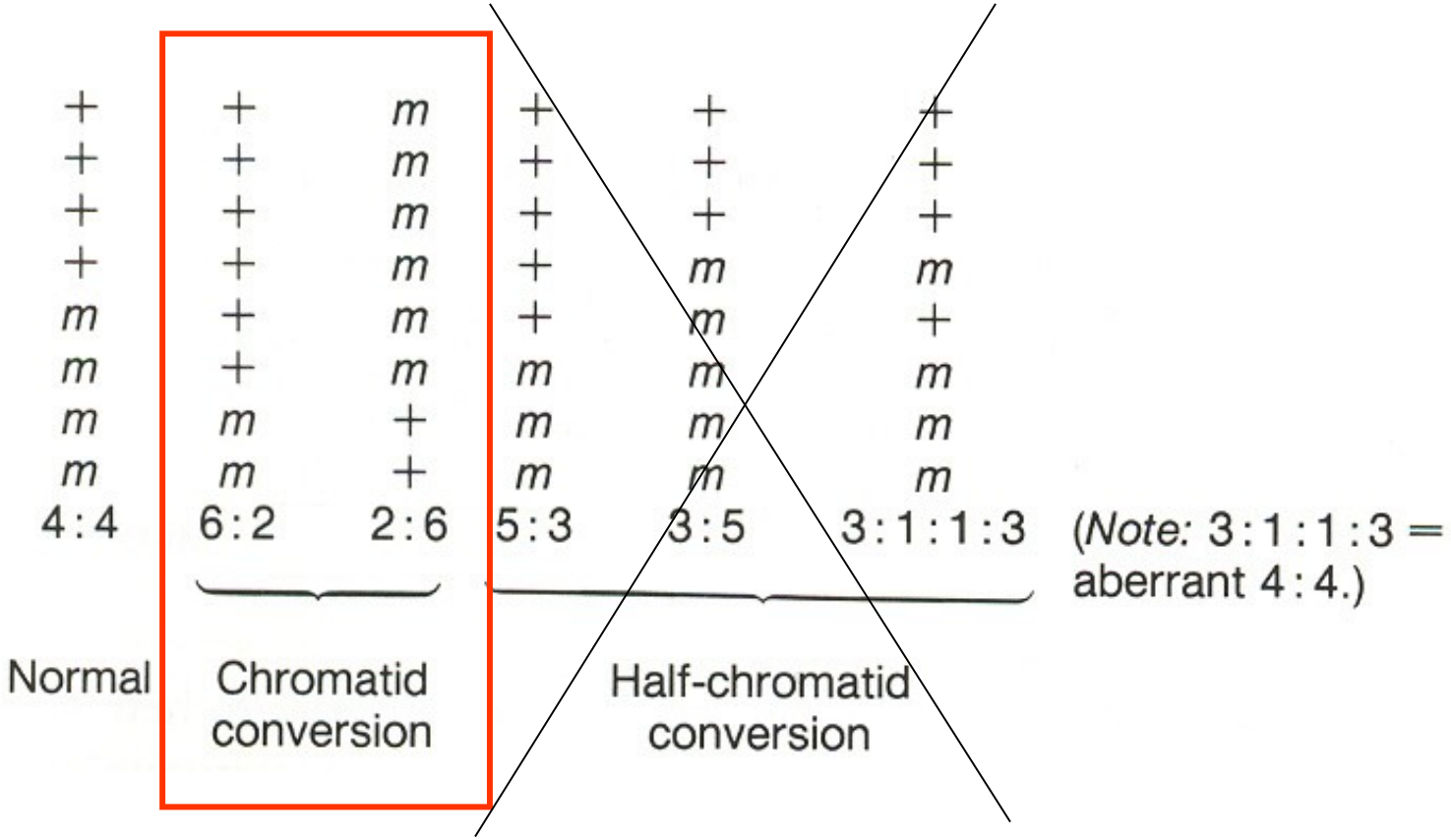
Tetrad analysis

Neurospora crassa

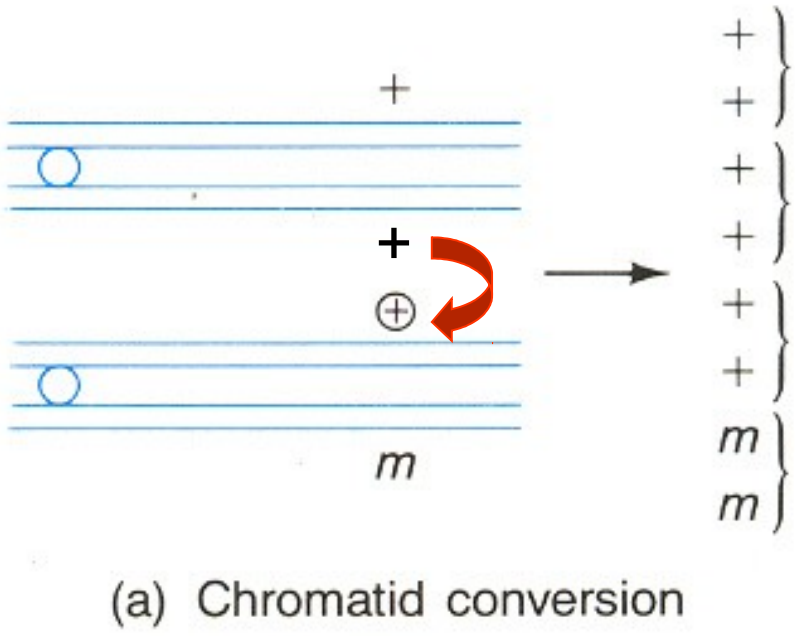
Gene conversion



Gene conversion!!!!!

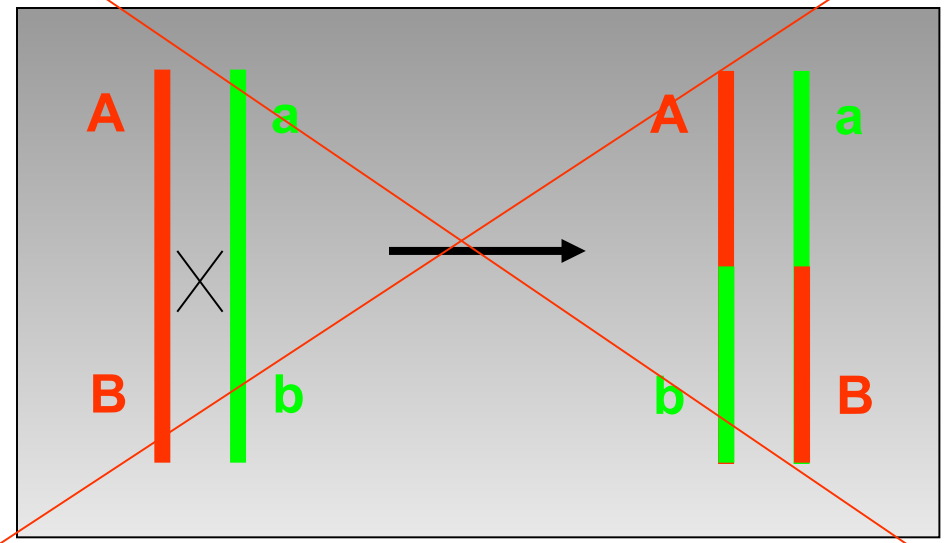


Gene conversion

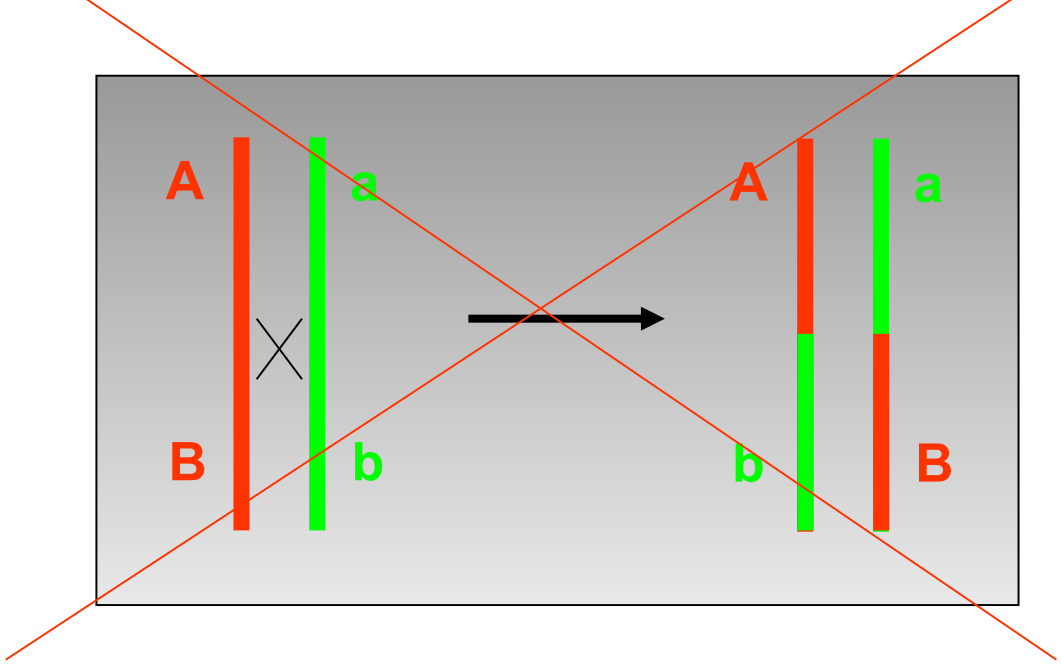


(a) Chromatid conversion

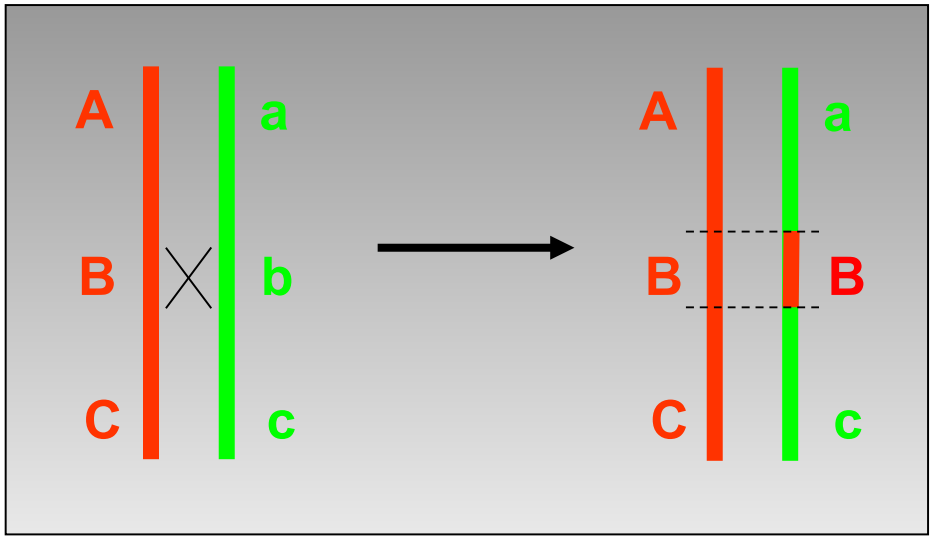
Non-reciprocal!!!



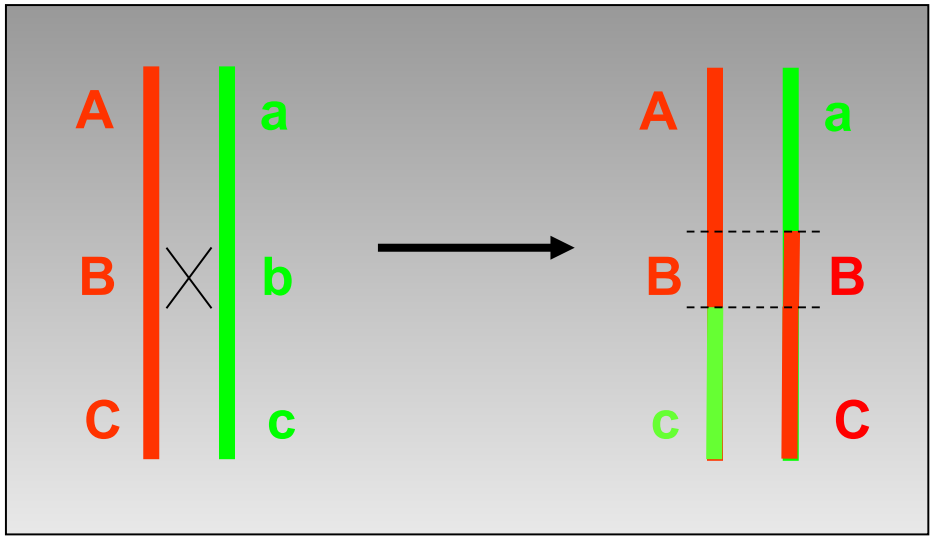
Mendel II. law! -



Gene conversion



Gene conversion + crossing over



Cell, Vol. 33, 25–35, May 1983, Copyright © 1983 by MIT

0092-8674/83/050025-11 $02.00/0

Review

The Double-Strand-Break Repair Model for Recombination

Jack W. Szostak,* Terry L. Orr-Weaver,*
Rodney J. Rothstein,† and Franklin W. Stahl‡
* Dana–Farber Cancer Institute
and Department of Biological Chemistry
Harvard Medical School
Boston, Massachusetts 02115
† Department of Microbiology
New Jersey Medical School
Newark, New Jersey 07103
‡ Institute of Molecular Biology
University of Oregon
Eugene, Oregon 97403

Summary

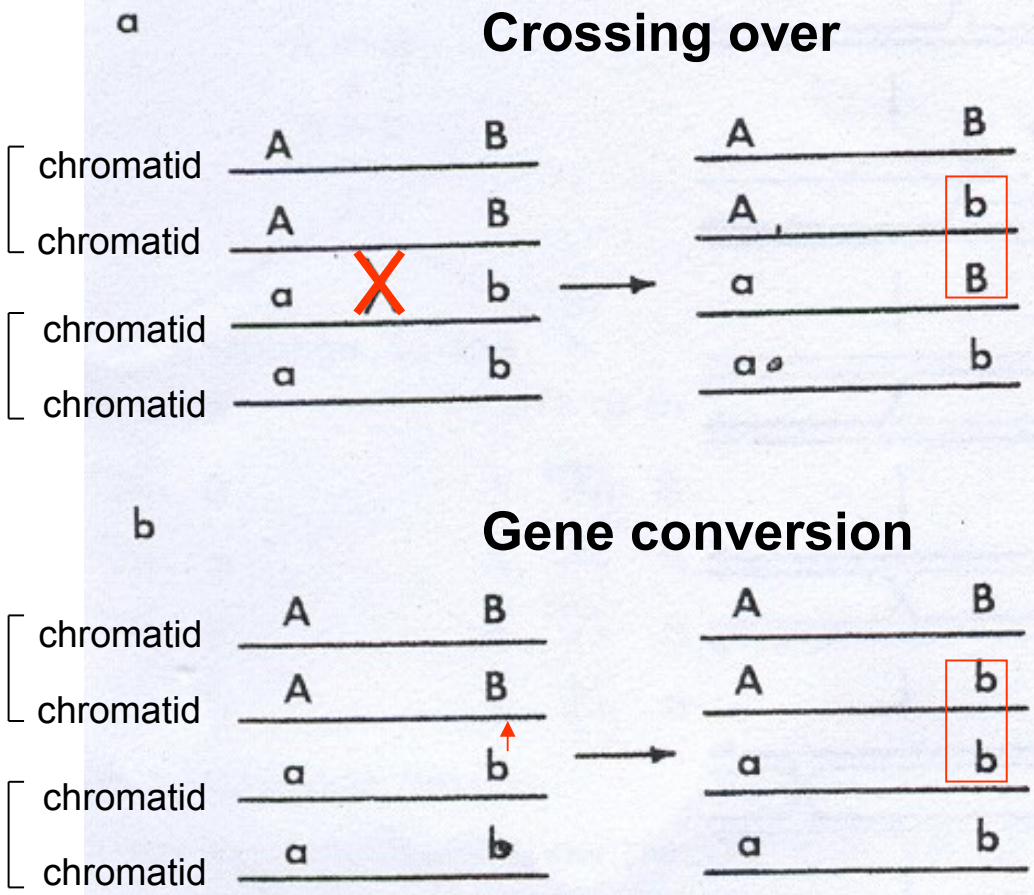
Gene conversion is the nonreciprocal transfer of information from one DNA duplex to another; in meiosis, it is frequently associated with crossing-over. We review the genetic properties of meiotic recombination and previous models of conversion and crossing-over. In these models, recombination is initiated by single-strand nicks, and heteroduplex DNA is generated. Gene conversion is explained by

however, like to acknowledge the contributions of those investigators whose extensive analysis of segregation in the fungi has provided the facts upon which this discussion of our model is based.

Conversion, Postmeiotic Segregation, and Crossing-Over

Upon the completion of premeiotic DNA replication, a diploid cell contains four DNA duplexes. The examination of eight-spored asci (Ascobolus, Neurospora, Sordoria) or of sectored-spore clones (Saccharomyces, Schizosaccharomyces) allows one to determine the genetic content of each of the eight single strands present at the beginning of meiosis. We will discuss all recombination in terms of these eight meiotic products. Any heterozygous marker will normally segregate $4^+:4^-$. The examination by tetrad analysis of all of the products of individual meiotic recombination events shows that recombinants between distant markers are produced in pairs (Figure 1a), with all markers showing 4:4 segregation. Such recombination events are called reciprocal exchanges or crossovers.

Occasionally a heterozygous marker will not segregate 4:4, but will show some aberrant pattern of segregation. The most common type of aberrant segregation in yeast



reciprocal

Recombination

Non-reciprocal

Figure 1. Crossing-Over and Conversion

(a) Crossing-over results in the production of complementary pairs of recombinant chromatids, with both markers segregating 4:4.

(b) Gene conversion results in the 6:2 segregation of one marker, and is therefore a nonreciprocal transfer of information from one chromatid to another.

The double-strand break (gap) repair model for recombination

Nem fizikai átkereszteződés

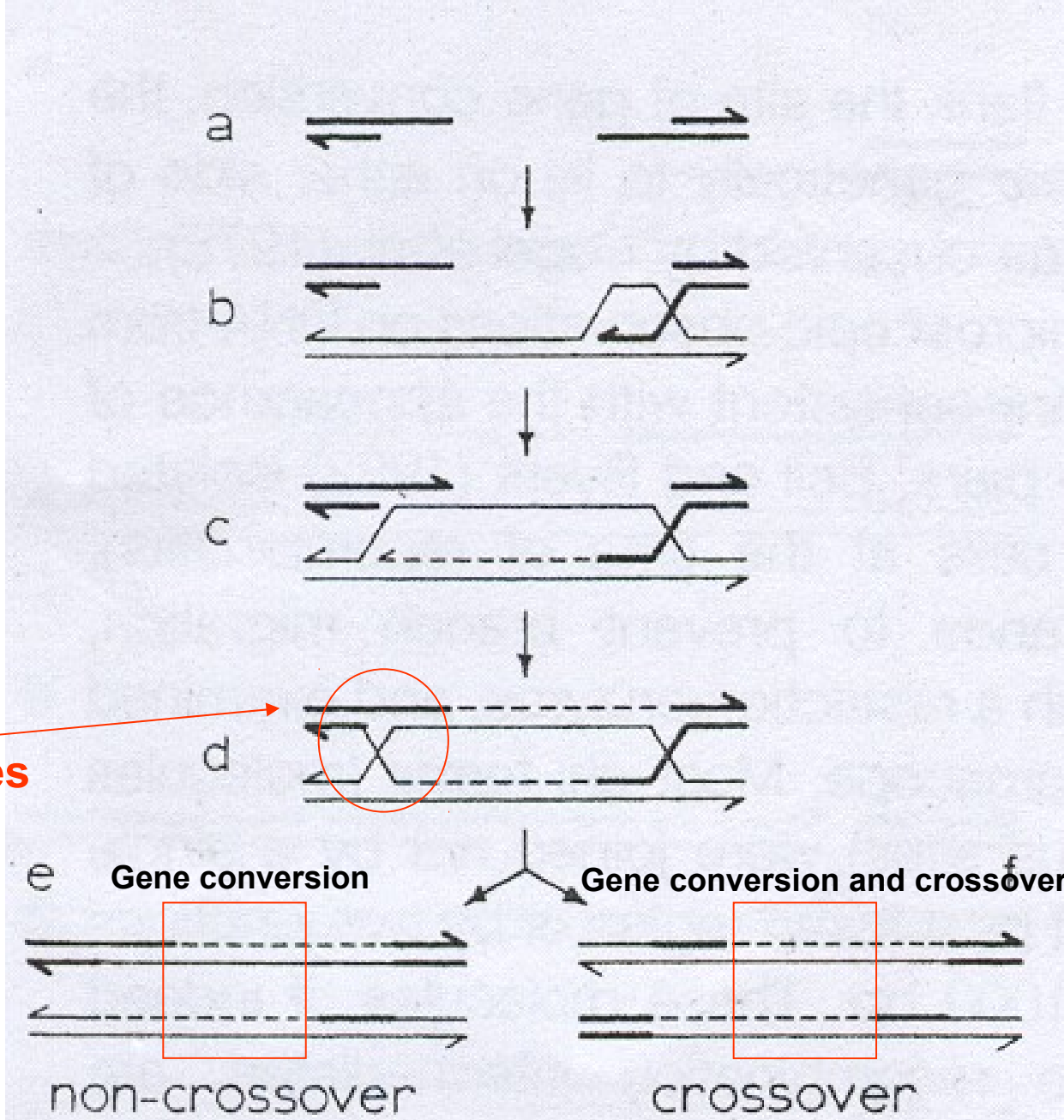


Figure 8. A Double-Strand-Break Repair Model

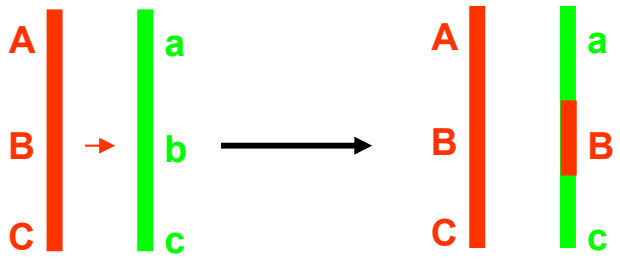
Recombination events:

100% Gene conversion

50% crossing over

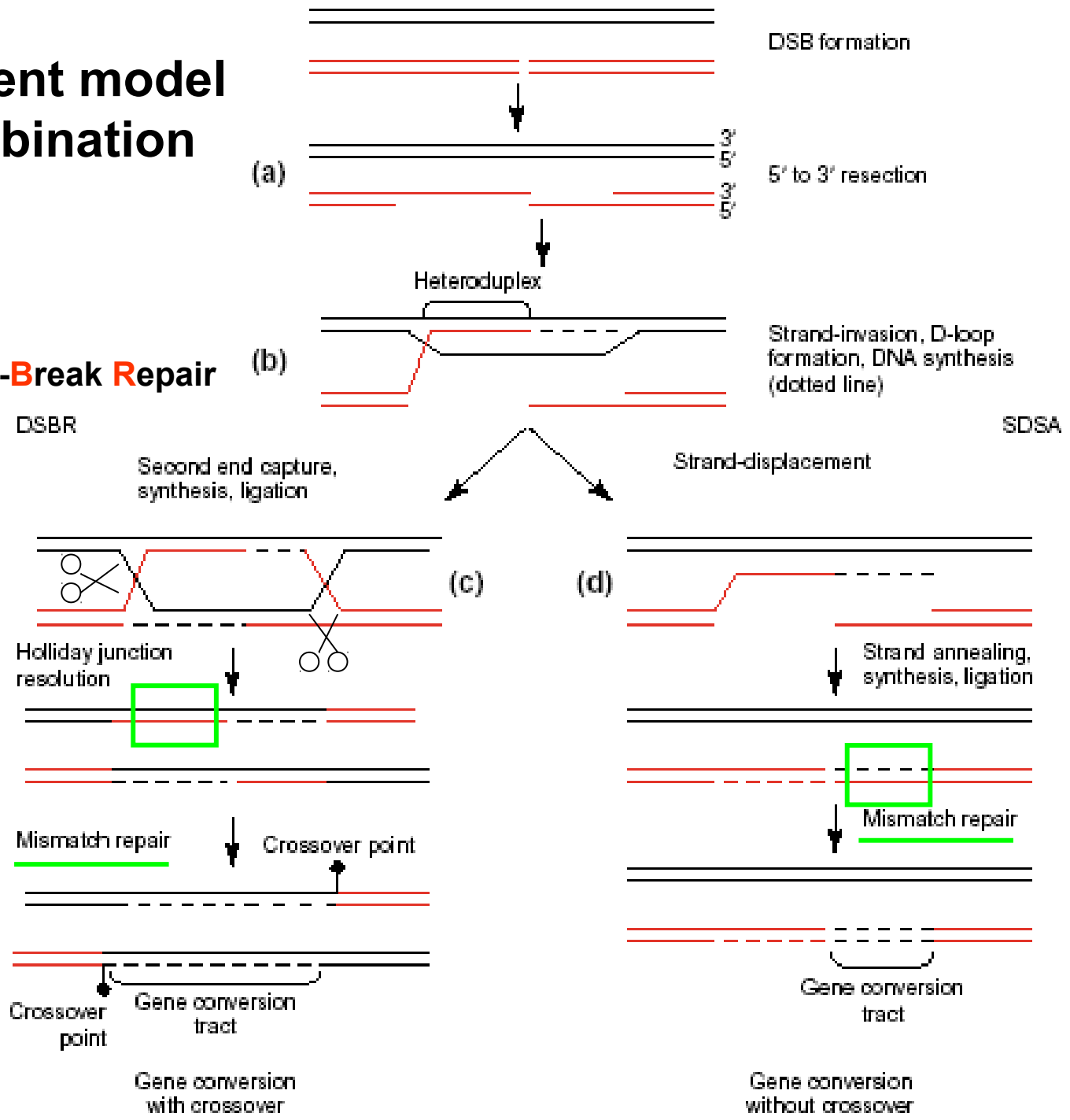
Crossover is often but NOT ALWAYS associated with gene conversion – these two processes are regulated independently

Remember: gene conversion always generates recombinant DNA



The current model of recombination

Double-Strand-Break Repair



Synthesis Dependent Strand Annealing

What we know about the mechanism…

- Meiotic recombination is induced by a DSB
- Homologs inteact before DSB occurs
- Chromatin structure highly affects DSB formation
- DSB formation is influenced by sequences at allelic position
- DSB is generated by Spo11
  - a topoisomearse and NOT an endonuclease
  - can interact simultaneously with two duplexes
- Certain sequences act as hotspots only because of a more accessible conformation for proteins involved in making the break

Recombination may be induced by a self-promoting element

Most evolutionary models have assumed that recombination is the evolutionary value of sex

Population genetics model for the evolution of recombination (sex) have usually ignored gene conversion

Gene conversion at the site of initiation is strong enough to promote the transmission of an allele – even if it gives no advantage to the individual or to the population

Disparity in gene conversion!!!!

Gene conversion events involving heterozygous deletions or insertions often show disparity in favour of the conversion events that duplicate the insertion or results in loss of the deletion

In contrast, gene conversion events involving heterozygous point mutations usually show no disparity

Meiotic Recombination Involving Heterozygous Large Insertions in *Saccharomyces cerevisiae*: Formation and Repair of Large, Unpaired DNA Loops

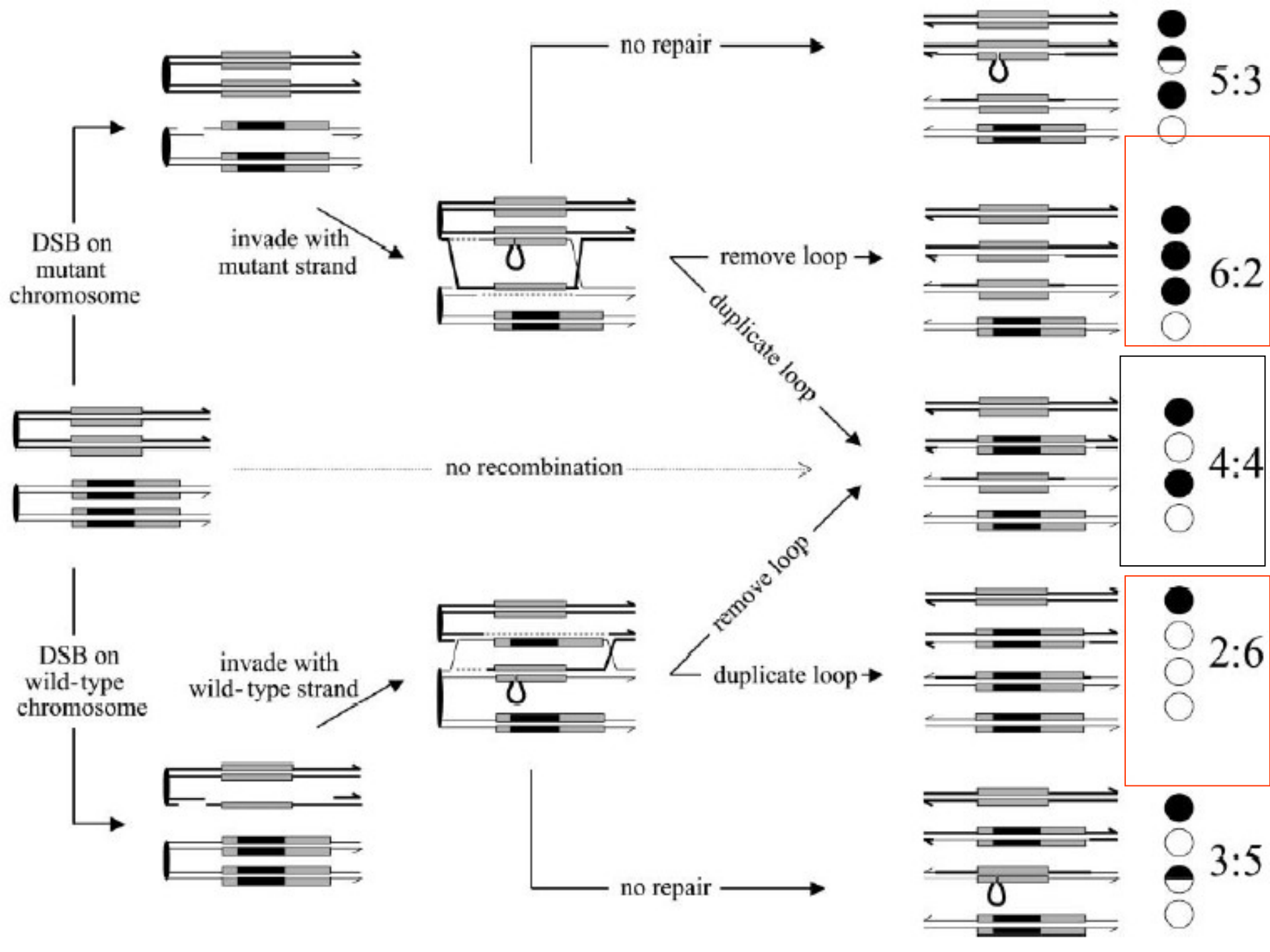
Hutton M. Kearney, David T. Kirkpatrick,[1] Jennifer L. Gerton[2] and Thomas D. Petes

Department of Biology, Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, North Carolina 27599-3280

Manuscript received March 30, 2001
Accepted for publication May 24, 2001

ABSTRACT

Meiotic recombination in *Saccharomyces cerevisiae* involves the formation of heteroduplexes, duplexes containing DNA strands derived from two different homologues. If the two strands of DNA differ by an insertion or deletion, the heteroduplex will contain an unpaired DNA loop. We found that unpaired loops as large as 5.6 kb can be accommodated within a heteroduplex. Repair of these loops involved the nucleotide excision repair (NER) enzymes Rad1p and Rad10p and the mismatch repair (MMR) proteins Msh2p and Msh3p, but not several other NER (Rad2p and Rad14p) and MMR (Msh4p, Msh6p, Mlh1p, Pms1p, Mlh2p, Mlh3p) proteins. Heteroduplexes were also formed with DNA strands derived from alleles containing two different large insertions, creating a large "bubble"; repair of this substrate was dependent on Rad1p. Although meiotic recombination events in yeast are initiated by double-strand DNA breaks (DSBs), we showed that DSBs occurring within heterozygous insertions do not stimulate interhomologue recombination.


DSB on mutant chromosome
invade with mutant strand
no repair
remove loop
duplicate loop
5:3
6:2
no recombination
4:4
DSB on wild-type chromosome
invade with wild-type strand
remove loop
duplicate loop
2:6
no repair
3:5

Insertions like to be spred

Meiotic segregation patterns of strains with larger insertions and bubbles

HIS4 alleles	Other homozygous alleles	Total tetrads	Ab. seg. (%)	PMS (%)	PMS/ Ab. (%)	% of total tetrads					
						4:4	6:2	2:6	5:3	3:5	Other
HIS4/his4::k1.5	Wild type	169	12	0	0	88	1	11	0	0	1
HIS4/his4::k1.5	*rad1-Δ*	175	27	17	63	73	5	5	7	10	1
HIS4/his4::U5.6	Wild type	213	13	0	0	87	2	11	0	0	0
HIS4/his4::U5.6	*rad1-Δ*	472	14	1	6	86	8	5	0.4	0.4	0
HIS4/his4::U5.6	*msh3-Δ*	237	16	0.4	3	84	7	8	0	0	0.4
HIS4/his4::U5.6	*pol4-Δ*	273	10	0	0	90	1	10	0	0	0
his4::U1.1a/his4::k1.5	Wild type	213	6	0	0	94	2	3	0	0	0.5
his4::U1.1a/his4::k1.5	*rad1-Δ*	204	26	19	73	75	5	2	11	7	0.5
his4::U1.1a/his4::k1.5	*msh3-Δ*	138	25	1	3	75	17	8	1	0	0
his4::U1.1a/his4::k1.5	*mlh1-Δ*	137	8	0	0	92	5	3	0	0	0
his4::U1.1a/his4::k1.5	*rad2-Δ*	232	7	0	0	94	4	2	0	0	0.4

Deletions like to be lost

Meiotic segregation patterns of strains homozygous for the *his4-51* allele and isogenic controls

Strain	*HIS4* alleles	Other relevant alleles	Total tetrads	Ab. seg. (%)	PMS (%)	PMS/Ab. (%)	% of total tetrads					*HIS4-LEU2* distance[a] (cM)
							4:4	6:2	2:6	5:3	3:5	
HMY100	*HIS4/his4::U1.1a*		349	22	0	0	78	9	13	0	0	33
HMY157	*HIS4/his4::U1.1a* *his4-51/his4-51*		156	3	0	0	97	1	2	0	0	17*
HMY190	*HIS4/his4::k1.5*		169	12	0	0	88	1	11	0	0	30
HMY234	*HIS4/his4::k1.5* *his4-51/his4-51*		186	4	0	0	96	2	3	0	0	20
HMY219	*HIS4/his4::k1.5*	*rad1-Δ/rad1-Δ*	175	27	17	63	73	5	5	7	10	37
HMY239	*HIS4/his4::k1.5* *his4-51/his4-51*	*rad1-Δ/rad1-Δ*	142	9	6	69	91	1	2	3	4	16*

The evolution of microsatellites through gene conversion

The evolution of genome size (DNA content) in eukaryotes

Junk DNA – selfish gene

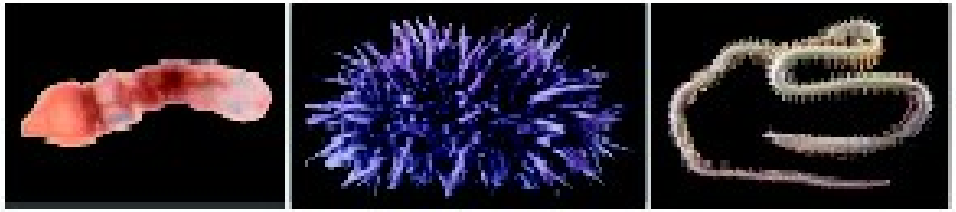
Repetitive elements

Large fraction of non-functional sequences

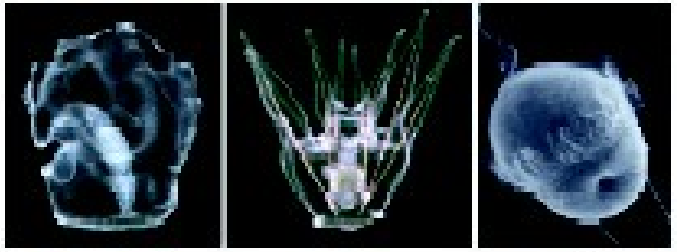
C-value paradox

More complex… more larger: Generally, genome size and biological complexity are correlated

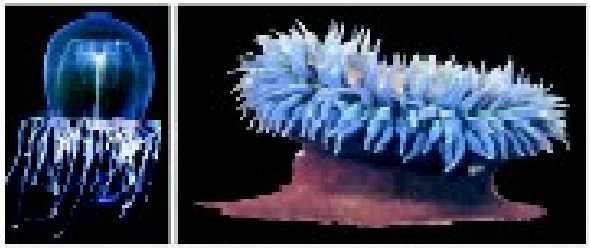
Grade of Organization



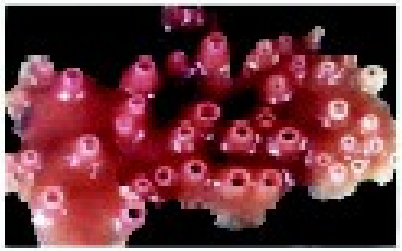
Adult Bilateria



Modern Bilaterian Primary Larvae

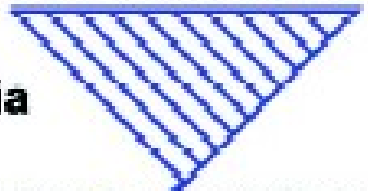


Cnidaria

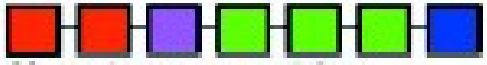


Porifera

Cladogram



All signaling pathways and TXF families present



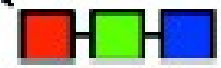
Hox cluster used for A/P vectorial patterning

Set-aside cells, Regional specification mechanisms

Ctenophora

Pax-6

Bilateral symmetry

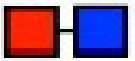


Central class *Hox* gene

Endomesoderm

Type 1 embryogenesis

nuclear hormone receptors



Two *Hox* Genes

Apical/blastoporal axis with *Brachyury* expressed in blastoporal gut

Gastrulation

Primordial *Hox* Gene

ots gene family, *paired box* genes

Multicellularity with ECM

A general negative relationship between selection efficiency and genome complexity: complex genomic structures have originated via non-adaptive, stochastic processes