Horisontaalse geeniülekande detekteerimise meetodite ülevaade

Aleksander Sudakov

Horizontal Gene Transfer

Horisontaalne geeniülekanne on geenide või geneetilise materjali ülekanne otse doonorilt retsipienti. See erineb normaalsest vertikaalsest ülekandest vanematelt järglastele.

Phylogenetic analysis would be easier if evolution occurred in a vertical fashion. However, horizontal or lateral transfer of genetic material (for instance through viruses) occurs, which makes it difficult to determine the phylogenetic origin of some evolutionary events.

Ajalugu

1977 - Carl Woese - SSU rRNA is ideal universal molecular chronometer

1982 - Serial Endosymbiosis Hypothesis - geenide ülekanne plastiididest ja mitokondritest tuuma genoomi

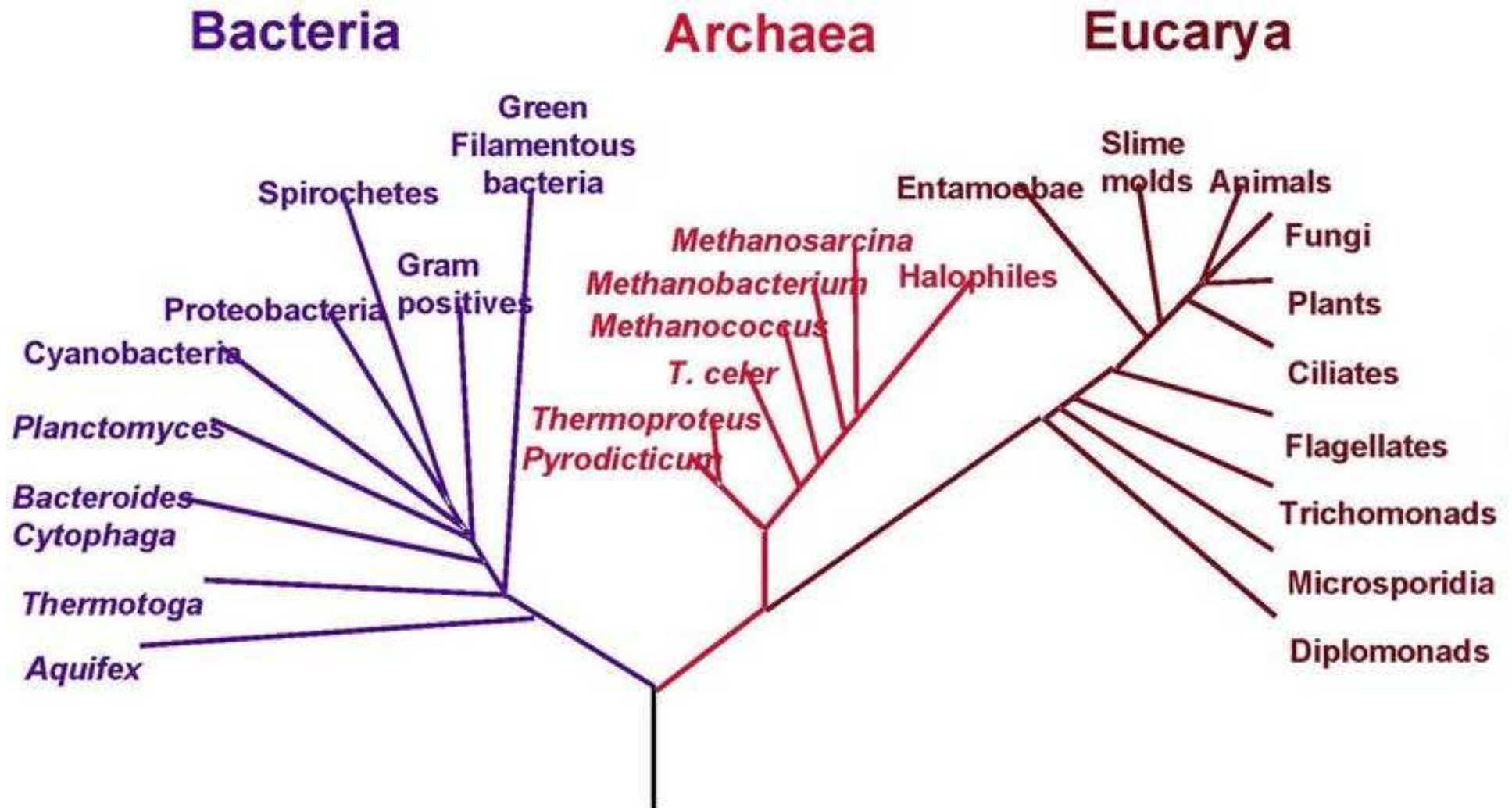1987 - Woese - HGT oli oluline varajases evolutsioonis

- ◈ inadequate phylogenetic information
- ◈ mutational saturation
- ◈ differential gene loss

SSU rRNA

The most common gene to be used for constructing phylogenetic relationships in prokaryotes is the 16s rRNA gene, since its sequences tend to be conserved among members with close phylogenetic distances, but variable enough that differences can be measured. However, in recent years it has also been argued that 16s rRNA genes can also be horizontally transferred. Although this may be infrequent, validity of 16s rRNA-constructed phylogenetic trees must be reevaluated.

Phylogenetic Tree of Life

1999: Thermotoga heats up HGT

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C., Fraser, C.M. 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima. Nature 399:323-329.

Table 4 Top eubacterial or archaeal match in *T. maritima* by role ID

Role category	Eubacteria	Archaea	Eukaryotes	None
Amino acid biosynthesis	49	20	2	1
Purines, pyrimidines, etc.	32	11	2	0
Fatty acid and phospholipid metab.	12	3	0	0
Biosynthesis of cofactors etc.	22	10	0	0
Central intermediary metabolism	27	12	1	3
Autotrophic metabolism	0	0	0	0
Energy metabolism	117	56	1	18
Transport	89	92	0	7
DNA metabolism	45	6	0	2
Transcription	17	0	0	0
Translation	118	6	0	6
Regulatory functions	61	9	0	0
Cell envelope	57	9	1	7
Cellular processes	55	10	0	0
Other	9	8	0	1
Hypotheticals	213	173	2	19
Unknown	52	26	0	5
TOTAL	975	451	9	69

The complete sequence of the bacterium Thermotoga maritima genome has revealed a large fraction of genes most closely related to those of archaeal species

2000: HGT Bakterites ja arhedes

Garcia-Vallvé, S., Romeu, A., Palau, J. 2000. Horizontal Gene transfer in Bacterial and Archaeal Complete Genomes. Genome Research 10:1719–1725.

Statistiline meetod, põhineb GC tasakaalul, koodonkasutusel, aminohappe kasutusel ja geenide grupeerumisel.

17 bakteri genoomi

7 arhe genoomi

Tulemused:

HGT 1.5%-14.5%

Arhedes % suurem

Info töötlemises vähem HGT

2000

Species	Disease caused	Genome size (bp)	Number of open reading frames	HGT	Percentage HGT	GC+	GC−
Proteobacteria							
Escherichia coli	. . .	4,639,221	4289	381	9.62	68	258
Haemophilus influenzae	Pneumonia	1,830,138	1709	96	6.19	31	33
Helicobacter pylori 26695	Ulcer	1,667,867	1553	89	6.41	0	68
Helicobacter pylori J99	Ulcer	1,643,831	1491	80	5.81	5	59
Rickettsia prowazekii	Typhus	1,111,523	834	28	3.62	10	8
Gram-positive bacteria							
Bacillus subtilis	. . .	4,214,814	4100	537	14.47	85	402
Mycoplasma genitalium	Urethritis	580,074	480	67	14.47	46	19
Mycoplasma pneumoniae	Pneumonia	816,394	677	39	5.93	0	32
Mycobacterium tuberculosis	Tuberculosis	4,411,529	3918	187	5.01	55	53
Spirochaete							
Borrelia burgdorferi	Lyme disease	910,724	850	12	1.56	0	5
Treponema pallidum	Syphilis	1,138,011	1031	77	8.32	30	38
Chlamydiae							
Chlamydia trachomatis	Trachoma, epidymitis	1,042,519	894	36	4.32	8	13
Chlamydia pneumoniae	Pneumonia, bronchitis	1,230,230	1052	55	5.70	28	16
Aquifex aeolicus	. . .	1,551,335	1522	72	4.84	6	43
Deinococcus radiodurans	. . .	2,648,638	2580	95	3.92	6	38
Synechocystis PCC6803	. . .	3,573,470	3169	219	7.50	12	151
Thermotoga maritima	. . .	1,860,725	1846	198	11.63	69	100
Ureaplasma urealyticum	Adverse pregnancy outcome, neonatal disease, and superative arthritis	751,719	610	32	5.70	8	9
Archaea							
Aeropyrum pernix	. . .	1,669,695	2694	370	14.01	108	203
Archaeoglobus fulgidus	. . .	2,178,400	2407	179	8.44	16	116
Methanobacterium thermoautotrophicum	. . .	1,751,377	1869	179	10.73	30	111
Methanococcus jannaschii	. . .	1,664,970	1715	77	5.00	18	38
Pyrococcus abyssi	. . .	1,765,118	1765	124	7.35	22	63
Pyrococcus horikoshii	. . .	1,738,505	2064	154	7.68	68	42

2001

Koonin, E.V., Makarova, K.S., Aravind L. 2001. Horisontal gene transfer in prokaryotes: quantification and classification. Annu. Rev. Microbiol. 55:709–42.

31 prokarüootse genoomi valgud
Taxonomic classification of database hits
Kui kõige sarnasem valk asub teises taksonis, on see
paradoxical best hit ja tõenäoline HGT kandidaat

Vaadeldi ka võimalikku ülekannet eukarüootidest
prokarüootidesse. Tüüpiliselt oli selliste geenide osakaal
1% ringis, ainult rakusisestel parasiitidel tunduvalt
suurem.

2001: HGT Arhede ja bakterite vahel

Species		Taxon	Paralog acquisition (%)		New gene acquisition (%)	
Aeropyrum	*pernix*	Archaea	34	(1.8)	47	(2.5)
Archaeoglobus	**fulgidus**	**Archaea**	**103**	**(4.3)**	**100**	**(4.2)**
Methanobac.	**thermoaut.**	**Archaea**	**100**	**(5.3)**	**61**	**(3.3)**
Methanococcus	*jannaschii*	Archaea	43	(2.5)	39	(2.3)
Pyrococcus	*horikoshii*	Archaea	55	(2.7)	39	(1.9)
Pyrococcus	*abyssi*	Archaea	72	(4.1)	39	(2.2)
Thermoplasma	**acidophilum**	**Archaea**	**112**	**(7.8)**	**54**	**(3.7)**
Halobacterium	**sp.**	**Archaea**	**204**	**(8.4)**	**174**	**(7.2)**
Aquifex	**aeolicus**	**Bacteria**	**87**	**(5.7)**	**45**	**(3.0)**
Thermotoga	**maritima**	**Bacteria**	**207**	**(11.1)**	**53**	**(2.9)**
Deinococcus	*radiodurans*	Bacteria	47	(1.5)	45	(1.5)
Bacillus	*subtilis*	Bacteria	71	(1.7)	28	(0.7)
Bacillus	*halodurans*	Bacteria	79	(1.9)	40	(1.0)
Mycobacterium	*tuberculosis*	Bacteria	50	(1.3)	62	(1.7)
Escherichia	*coli*	Bacteria	26	(0.6)	13	(0.3)
Haemophilus	*influenzae*	Bacteria	3	(0.2)	3	(0.2)
Rickettsia	*prowazekii*	Bacteria	23	(2.8)	7	(0.8)
Pseudomonas	*aeruginosa*	Bacteria	66	(1.2)	39	(0.7)
Neisseria	*meningitidis*	Bacteria	6	(0.3)	5	(0.2)
Vibrio	*cholerae*	Bacteria	12	(0.3)	16	(0.4)
Xylella	*fastidiosa*	Bacteria	22	(0.8)	8	(0.3)
Buchnera	*sp.*	Bacteria	0	(0.0)	0	(0.0)
Treponema	*pallidum*	Bacteria	10	(1.0)	4	(0.4)
Borrelia	*burgdorferi*	Bacteria	3	(0.4)	6	(0.7)
Synechocystis	**PCC6803**	**Bacteria**	**219**	**(6.9)**	**115**	**(3.6)**
Chlamydophyla	*pneumoniae*	Bacteria	23	(2.2)	9	(0.9)
Mycoplasma	*pneumoniae*	Bacteria	0	(0.0)	1	(0.1)
Ureaplasma	*urealyticum*	Bacteria	1	(0.2)	1	(0.2)
Helicobacter	*pylori*	Bacteria	5	(0.3)	3	(0.2)
Campylobacter	*jejuni*	Bacteria	5	(0.3)	4	(0.2)

2001: HGT bakteri liinide vahel

Species	Taxon	Paralog acquisition	Acquisition of new gene
Mycoplasma genitalium	Firmicutes	6 (1.2)	2 (0.4)
Mycoplasma pneumoniae	Firmicutes	9 (0.9)	8 (1.2)
Bacillus subtilis	**Firmicutes**	**685 (16.7)**	**383 (9.3)**
Bacillus halodurans	**Firmicutes**	**772 (19.0)**	**400 (9.8)**
Treponema pallidum	**Spirochaetales**	**132 (12.8)**	**204 (19.8)**
Borrelia burgdorferi	**Spirochaetales**	**109 (12.8)**	**141 (16.6)**
Haemophilus influenzae	Proteobacteria	32 (1.9)	21 (1.2)
Rickettsia prowazekii	Proteobacteria	49 (5.9)	32 (3.8)
Escherichia coli	Proteobacteria	223 (5.2)	102 (2.4)
Pseudomonas aeruginosa	Proteobacteria	448 (8.1)	275 (5.0)
Neisseria meningitidis	Proteobacteria	55 (2.7)	34 (1.7)
Vibrio cholerae	Proteobacteria	130 (3.4)	85 (2.2)
Xylella fastidiosa	Proteobacteria	88 (3.2)	83 (3.0)
Buchnera sp.	Proteobacteria	0 (0.0)	0 (0.0)
Mycoplasma genitalium	Mycoplasma	0 (0.0)	0 (0.0)
Chlamydia pneumoniae	Chlamydiales	4 (0.4)	25 (2.4)

2001: puude analüüs

Et saada parem ülevaade, oli tehtud kandidaatide detailne fülogeneetiline analüüs nelja genoomi jaoks. HGT oli tõestatud umbes kolmandikul neist, ülejäänutel on puu topoloogia liiga keeruline ja nad märgiti kui "uncertain"

Species	Reference taxon	Xenologous gene displacement	Acquisition of paralog	Uncertain
Haemophilus influenzae	Proteobacteria	6	6	20
Vibrio cholerae	Proteobacteria	5	40	85
Aquifex aeolicus	Bacteria	31	8	48
Methanobac. Thermoautotr.	Archaea	19	17	64

2001: Surrogate methods

Leiavad sellised geenid, mis on erinevad ülejäänud genoomist.

Plussid: ei vaja fülogeneesi puude ehitamist - vähem arvutust. Saab vaadelda geene, millel ei ole leitud ortolooge.

Miinused: tulemused ei ole usaldusväärsed

Mõned meetodid rakendatavad ainult väikestele järjestuse regioonidele (n. 50kb)

2001: Surrogate methods

Criterion	#ORFs exp.	#ORFs obs.	% +/-
GC only	535.1	462**	-14
MM only	449.6	332**	-26
PD only	270.4	322**	+19
DP only	187.3	200	+7
GC∩MM	95.5	201**	+110
GC∩PD	57.4	18**	-69
GC∩DP	39.8	24*	-40
MM∩PD	48.3	39	-19
MM∩DP	33.4	32	-4
PD∩DP	20.1	12	-40
GC∩MM∩PD	10.2	20*	+96
GC∩MM∩DP	7.1	24**	+238
GC∩PD∩DP	4.3	3	-
MM∩PD∩DP	3.6	2	-
GC∩MM∩PD∩DP	0.8	0	-

Numbers of E. coli K12 MG1655 ORFs identifed by four surrogate methods, individually and in combinations

- GC - base composition
- MM - Markov model
- PD - phylogenetic discordance - ORFs that show patterns of BLAST matches signifcantly different from the median pattern shown by ORFs in that same genome
- DP - distributional profle - unusual patterns of gene distribution among organisms (phyletic pattern)

2002: Eliminate phylogenetically discordant sequences

Clarke, G.D.P., Beiko, R.G., Ragan, M.A., Charlebois R.L. 2002. Inferring genome trees by using a filter to eliminate phylogenetically discordant sequences and a distance matrix based on mean normalized BLASTP scores.J. of Bacteriol. 184:2072-2080.

Hüpotees:

Kui eemaldada analüüsist "incongruent" geenid, siis ülejäänute põhjal ehitatud puud peaksid olema sarnasemad rRNA puuga.

Phylogenetically discordant sisaldab nii HGT geene kui teisi mittesobivaid geene.

Nende eemaldamine muutis puu topoloogiat minimaalselt, aga bootstrap confidence suurenes. Puu tuli enamjaolt rRNA puuga sarnane.

2003: HGT mõju metabolismisüsteemide levikule

Boucher, Y., Douady, C.J., Papke, R.T., Walsh, D.A., Boudreau, M.E.R., Nesbø, C.L., Case, R.J., Doolittle, W.F. 2003. Lateral gene transfer and the origins of prokaryotic groups.

Algselt tugines prokarüootide klassifikatsioon peamiste füsioloogiliste tunnuste olemasolul (fotosüntees, aeroobne hingamine, lämmastiku fikseerimine, sulfaadi redutseerimine jne.). Nende levik evolutsioonipuul ei ole korrapärane.

Antud uuring leiab mitmed tõenäolised ülekanded eri metabolismisüsteemides.

"Selfish operon" – kui vajalik funktsioon kasutab mitut geeni, siis on ainult nende kõikide (operoni) ülekanne kasulik. Kõik vaadeldud metabolismisüsteemid on kogunenud ühte või mitmesse operoni – kaudne tõendus nende levikule HGT kaudu.

2003

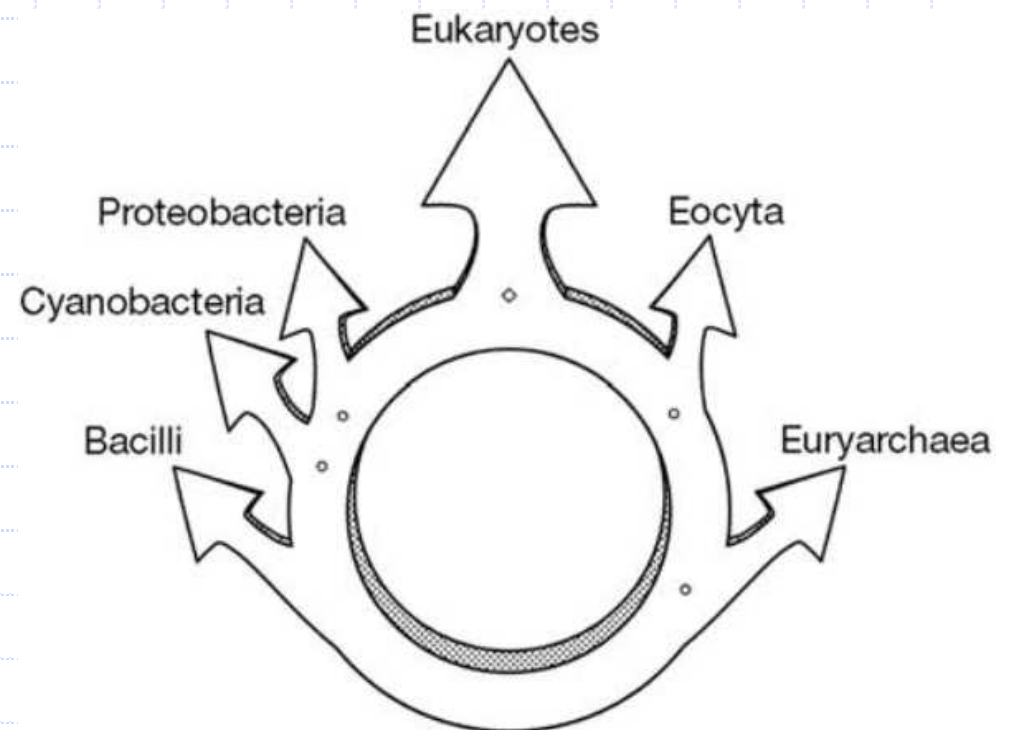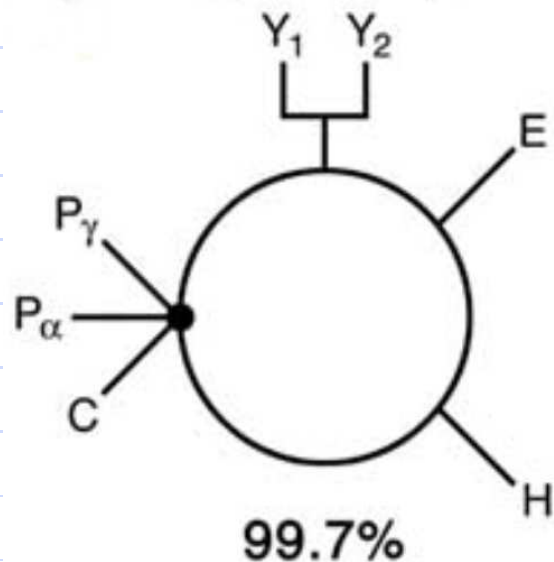HGT toimub kõikidel fülogeneesi tasemetel!

Osa süsteeme ei levinud arhede ja bakterite vahel.

HGT takistused

- ◆ Väga keerulised, vajavad mitut spetsiifilist geeni
- ◆ Ei ole operonidena
- ◆ Lisatingimused – arhedel erinev membraan; lämmastiku fikseerimine peab olema eraldatud hingamisest.

Ring of life

The eukaryotic genome resulted from a fusion it is expected that in some gene trees eukaryotes will be related to Bacteria, whereas in other gene trees eukaryotes will be related to Archaea. Informational genes of eukaryotes are primarily derived from Archaea and the operational genes are primarily derived from Bacteria what is also consistent with the ring of life.

2002: Anti GMO

Genetic engineering bypasses reproduction altogether by exploiting horizontal gene transfer, so genes can be transferred between distant species that would never interbreed in nature. For example, human genes are transferred into pig, sheep, fish and bacteria. Toad genes are transferred into potatoes. Completely new, exotic genes, can therefore be introduced into food crops.

Research results released earlier this year by the Food Standards Agency, indicating that transgenic DNA from GM soya flour, eaten in a single hamburger and milk shake meal, was found transferred to the bacteria in the gut contents from the colostomy bags of human volunteers.

Genome of a Marsupial:
The Gray, Short-tailed Opossum

January 18 2005

The opossum (Monodelphis domestica) is the latest vertebrate

Sequenced and assembled by The Broad Institute, Cambridge, MA, USA.