

# *Amelioration*

ATTAATATTTT

# *Amelioration*

GGGCGCCGCC

ATTAATATTTT

# *Amelioration*

GGGAAATATCC

ATTCGCCGTTT

# *Amelioration*

ATTCGCCGTTT

# *Amelioration*

ATTCGCCGTTT

1Myr

# *Amelioration*

ATTCGCCGTTT

1Myr

ATTGCCGTTT

5Myr

# *Amelioration*

ATTCGCCGTTT

1Myr

ATTTGCCGTTT

5Myr

ATTTACCGTTT

20Myr

# *Amelioration*

ATTCGCCGTTT

1Myr

ATTTGCCGTTT

5Myr

ATTTACCGTTT

20Myr

ATTTATCGTTT

50Myr



# *Amelioration*

ATTCGCCGTTT

1Myr

ATTTGCCGTTT

5Myr

ATTTACC GTTT

20Myr

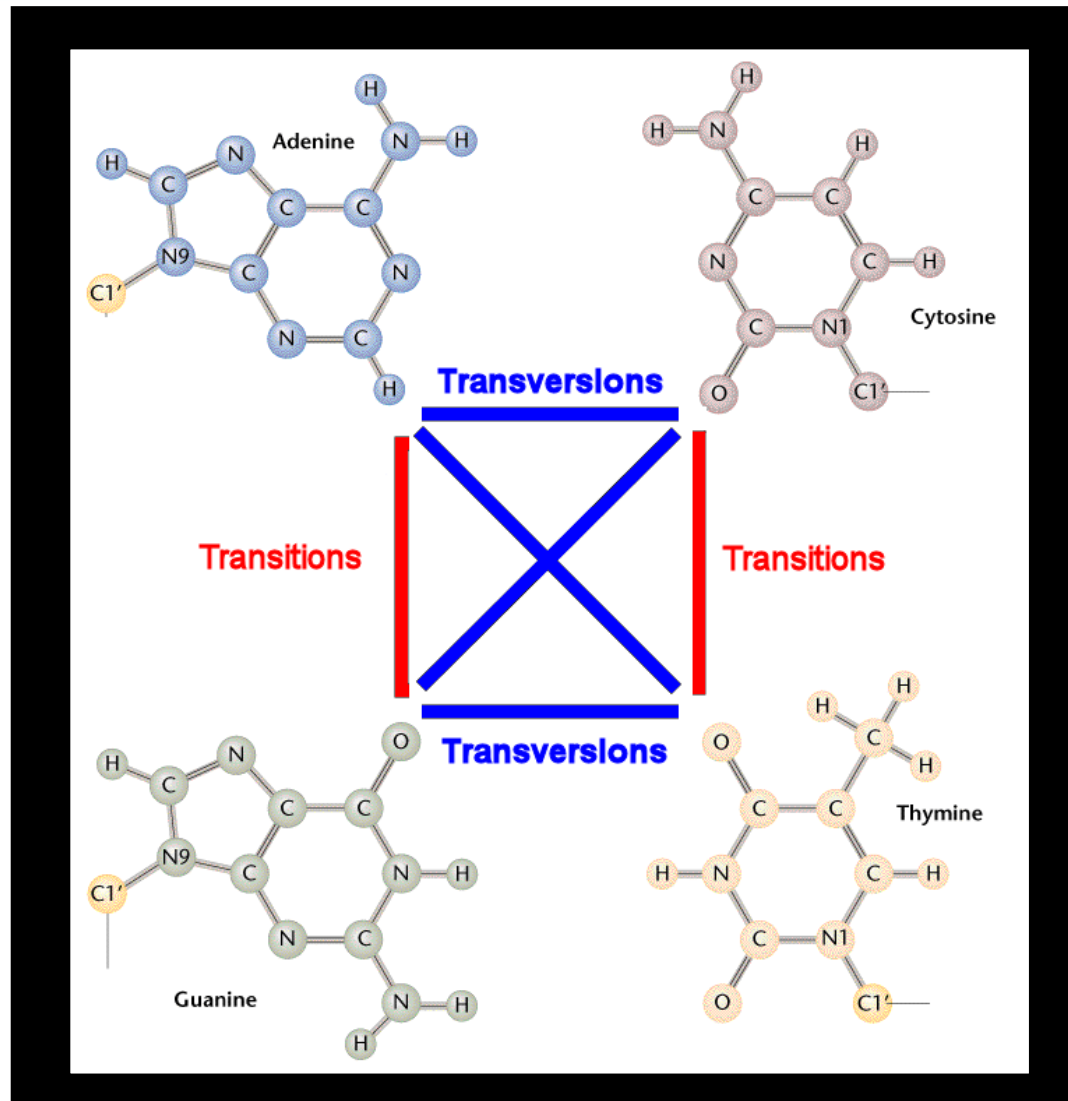
ATTTATCGTTT

50Myr

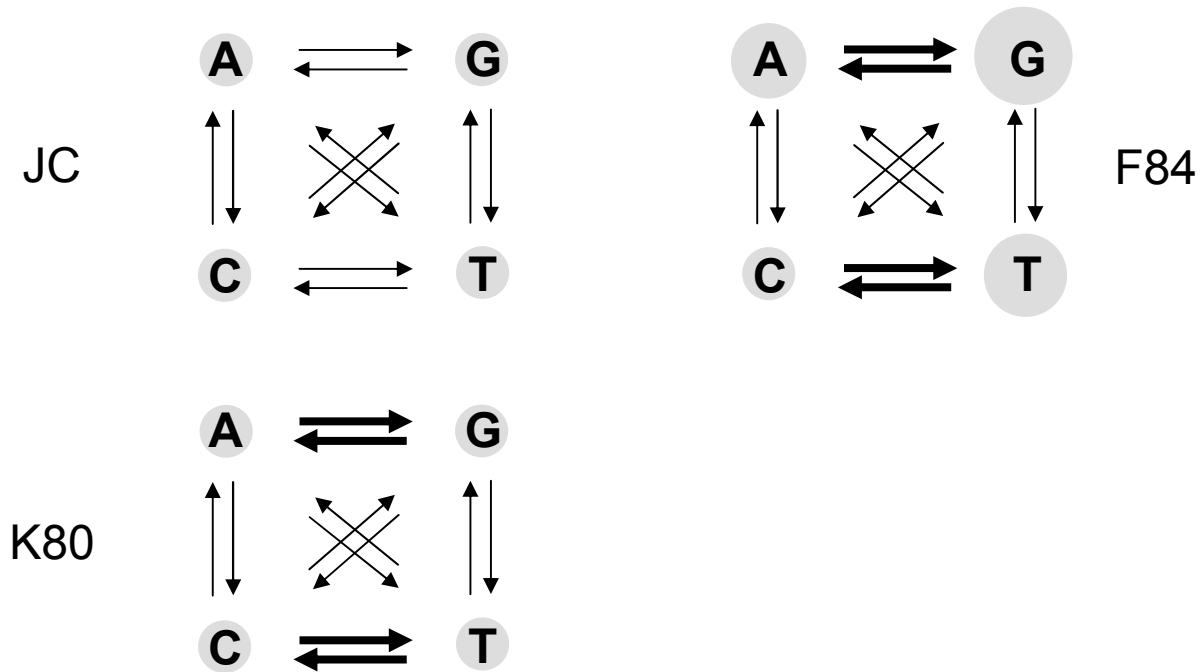
ATTTATAGTTT

100Myr

# Transition vs Transversion

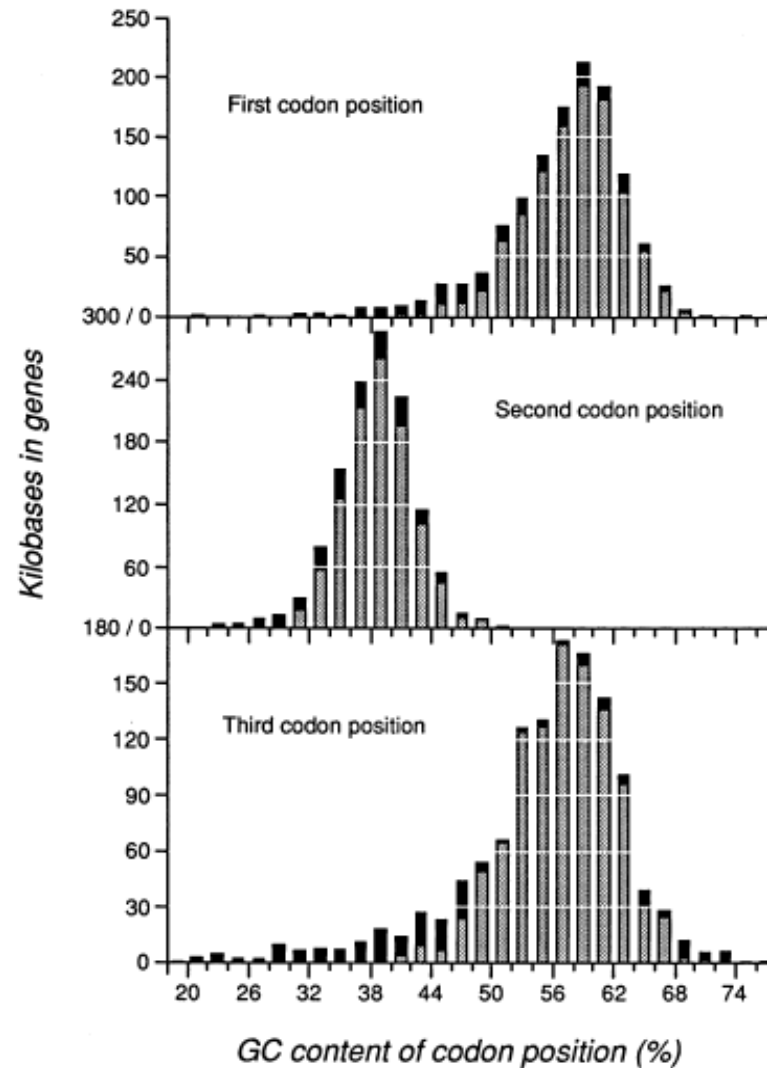


# Transition vs Transversion (models)

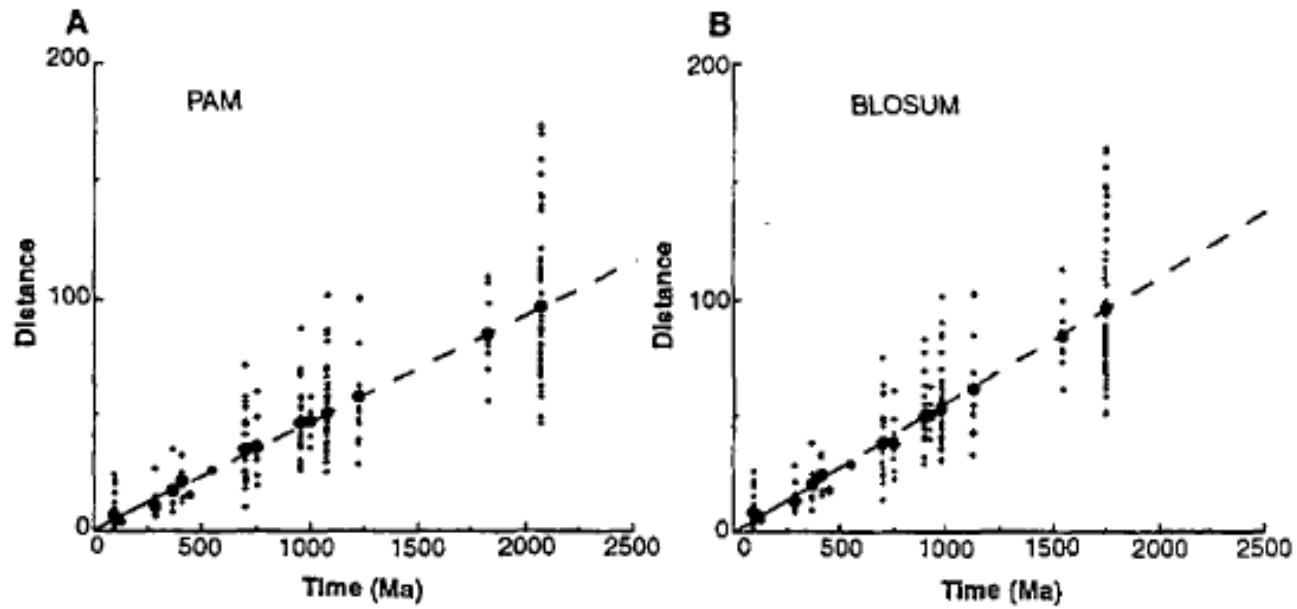


# $G+C_{1,2,3}$

GC contents of 1,294 *E. coli* genes. GC contents at each codon position are shown. *Gray bars* represent *native* (i.e., ancestral) genes; *black bars* represent genes inferred to have been acquired by *horizontal transfer*.



# *Distance ~ Time*



# Distance ~ Time

**Table 3.** Average resemblances and divergence times by extrapolation.

	N*	ID† (%)	D‡	LCA§	LCA'	LCA''	LCA'''
Deuterostome-protostome	21	64 ± 10	36	750	656	675	675
Schizocoelome-pseudocoelome	9	64 ± 8	37	750	784	750	750
Fungi-animal	54	55 ± 8	52	1050	978	965	965
Plant-animal	33	57 ± 8	47	1000	1000	1000	1000
Protist-plant-animal-fungi	14	51 ± 10	59	1250	1236	1230	1230
Archaeobacteria-eukaryotes	9	39 ± 6	85	1800	1889	1700	1870
Eubacteria-eukaryotes	57	37 ± 9	96	2050	2080	1875	2156
<i>Bacilli-E. coli</i>	28	45 ± 9	75	(1500)	1610	1450	1523
<i>E. coli-Salmonella</i>	8	94 ± 6	6	(100)	(100)	(100)	(100)

**Table 2.** Average resemblances and divergence times from fossil record.

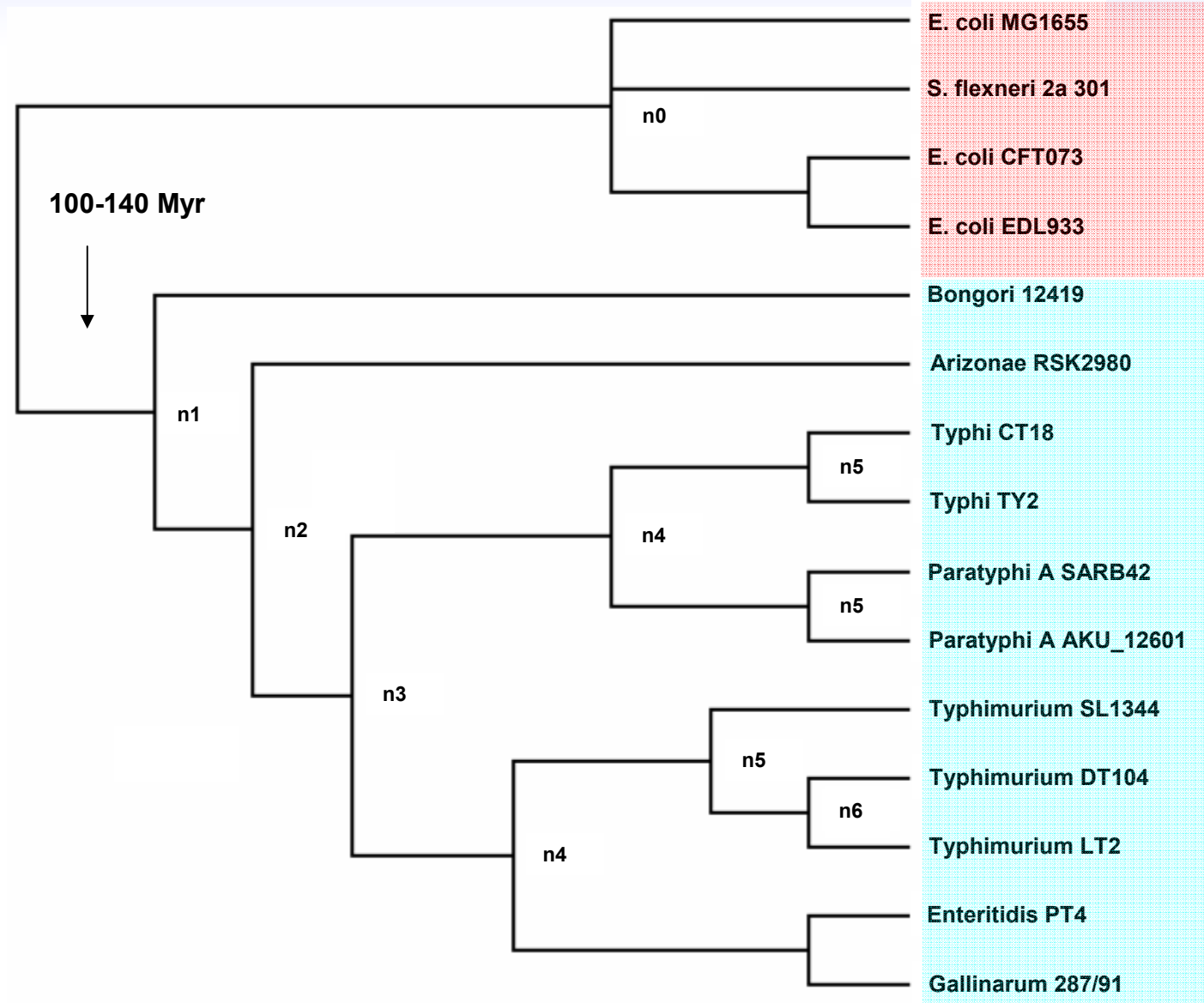
	N*	Identity† (% ± SD)	Dis- tance‡	LCA§ (Ma)
Mammal-mammal	43	91 ± 6	6	100
Eutheria-marsupial	2	92 ± 2	5	130
Mammal-bird-reptile	12	84 ± 6	11	300
Amniote-amphibian	5	78 ± 9	17	365
Tetrapod-fish	4	74 ± 8	22	400
Gnathostome-lamprey	1	78	16	450
Chordate-echinoderm	1	69	27	550

\*Number of enzyme sets compared. †Percent identity. ‡Distances taken from Fig. 2C. §Last common ancestor.

**1.  $D = -\ln S \times 100$**

**2. Extrapolate on  
vertebrate fossil  
records**

# Divergence of *Salmonella* and *E. coli*



# Substitution Rate (S)

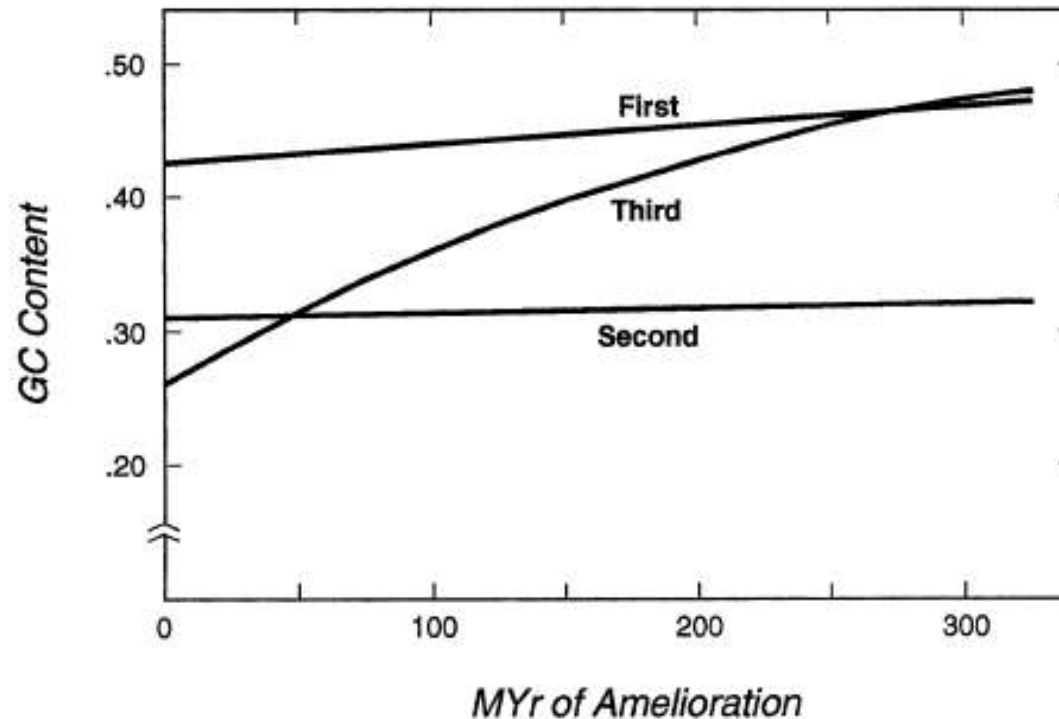
Table 1. Amelioration parameters used for *E. coli* and *S. enterica*

Class	Proportion of sites		Substitution rate <sup>a</sup> (% Divergence per Myr)
	Synonymous	Nonsynonymous	
Synonymous	1	0	0.910
Nonsynonymous	0	1	0.045
First codon position	0.09	0.91	0.123
Second codon position	0	1	0.045
Third codon position	0.72	0.28	0.668

<sup>a</sup> Based on a divergence time between *E. coli* and *S. enterica* of 100 million years (Doolittle et al. 1996; Ochman and Wilson 1987, 1988)



# Amelioration: *Shigella flexneri spa* genes

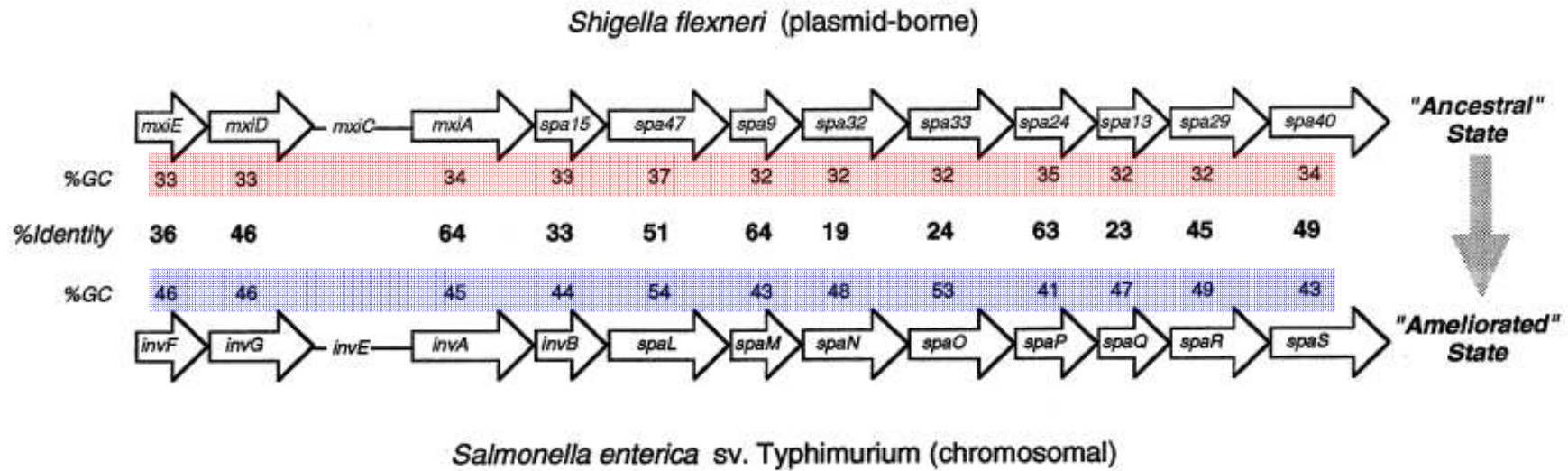


**Fig. 4.** The GC contents of the three codon positions of the *Shigella flexneri spa* genes are plotted for 325 iterations of 1 Myr of amelioration toward the base composition of the *Salmonella enterica* genome.

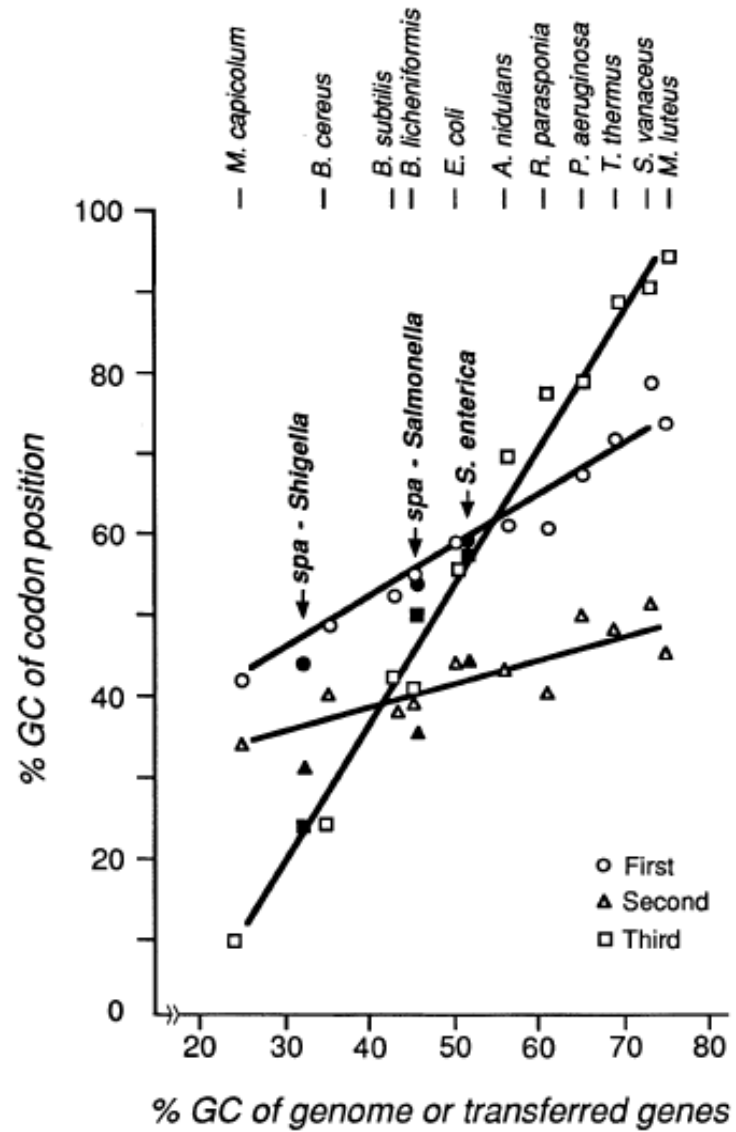
*inv/spa* genes of *Salmonella enterica* are required for bacterial entry into mammalian intestinal cells

Lawrence 1997

# Shigella & Salmonella spa genes



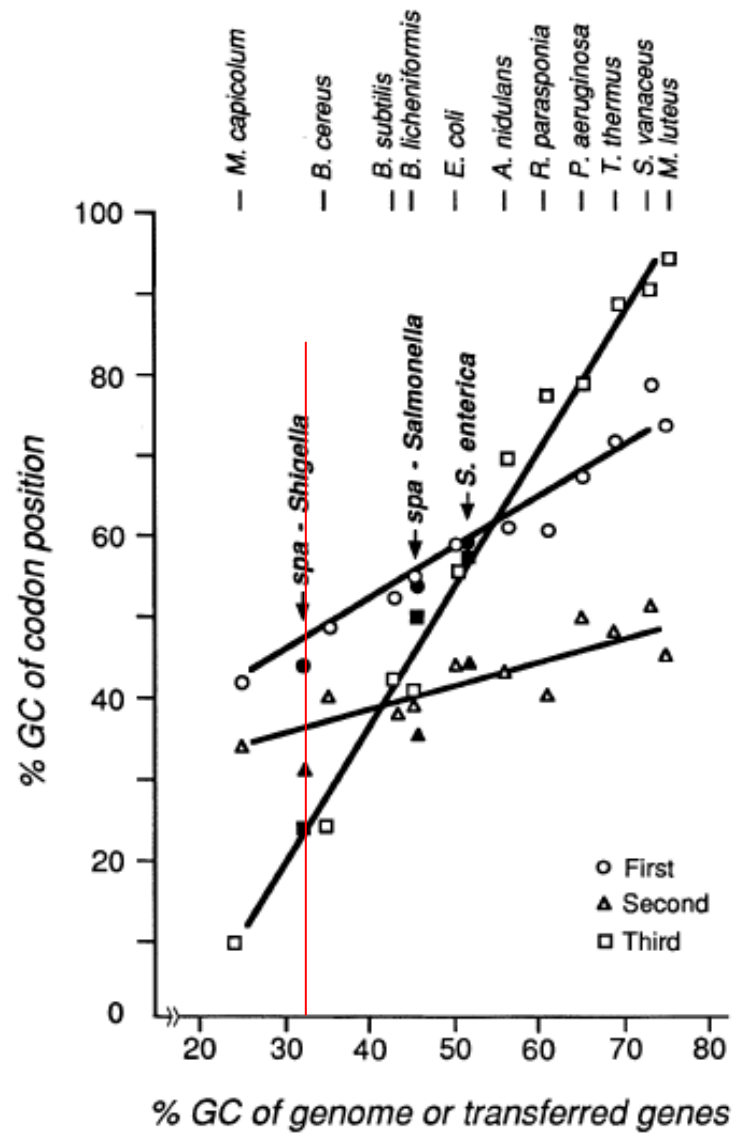
# G+C<sub>1,2,3</sub> VS G+C



$GC_1 = 0.615 \times GC + 26.9$   
 $GC_2 = 0.270 \times GC + 26.7$   
 $GC_3 = 1.692 \times GC - 32.3$

Lawrence 1997

# G+C<sub>1,2,3</sub> VS G+C



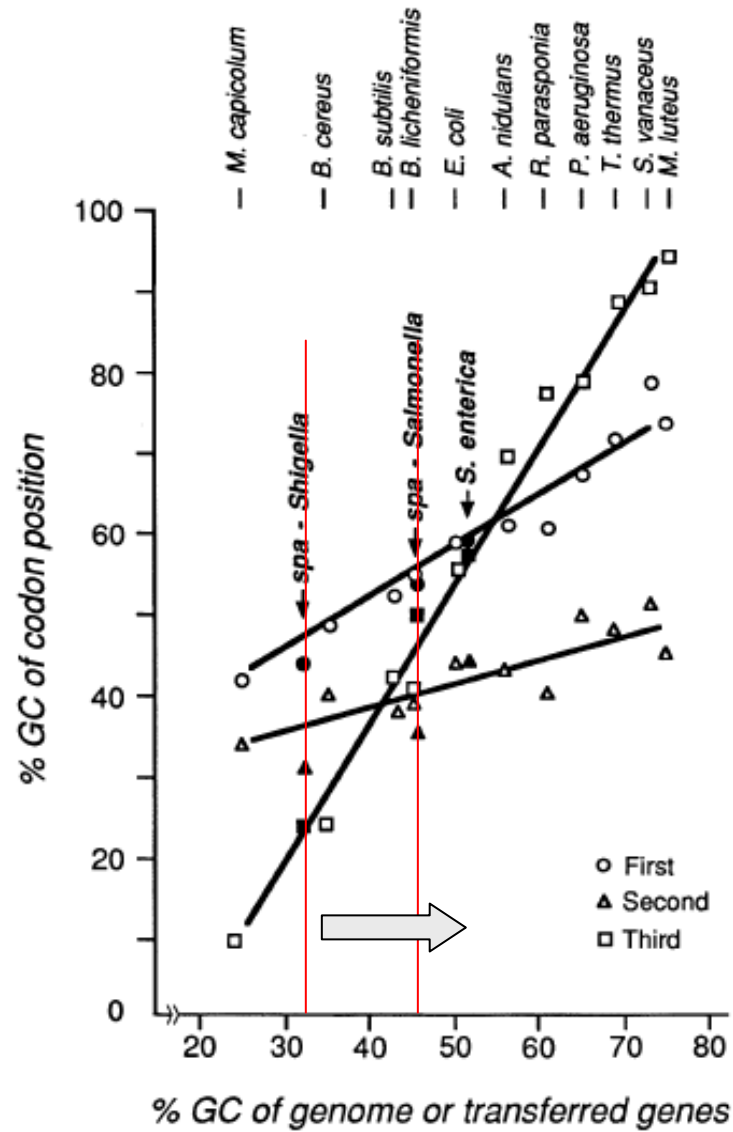
$$GC_1 = 0.615 \times GC + 26.9$$

$$GC_2 = 0.270 \times GC + 26.7$$

$$GC_3 = 1.692 \times GC - 32.3$$

Lawrence 1997

# G+C<sub>1,2,3</sub> VS G+C



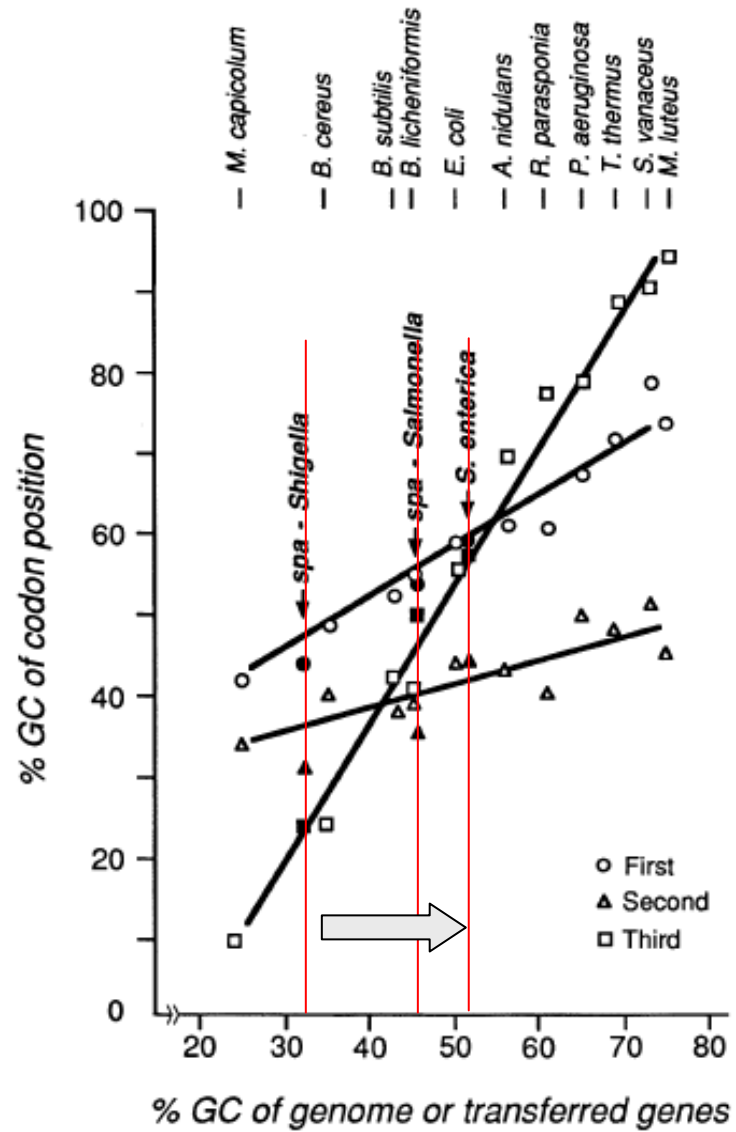
$$GC_1 = 0.615 \times GC + 26.9$$

$$GC_2 = 0.270 \times GC + 26.7$$

$$GC_3 = 1.692 \times GC - 32.3$$

Lawrence 1997

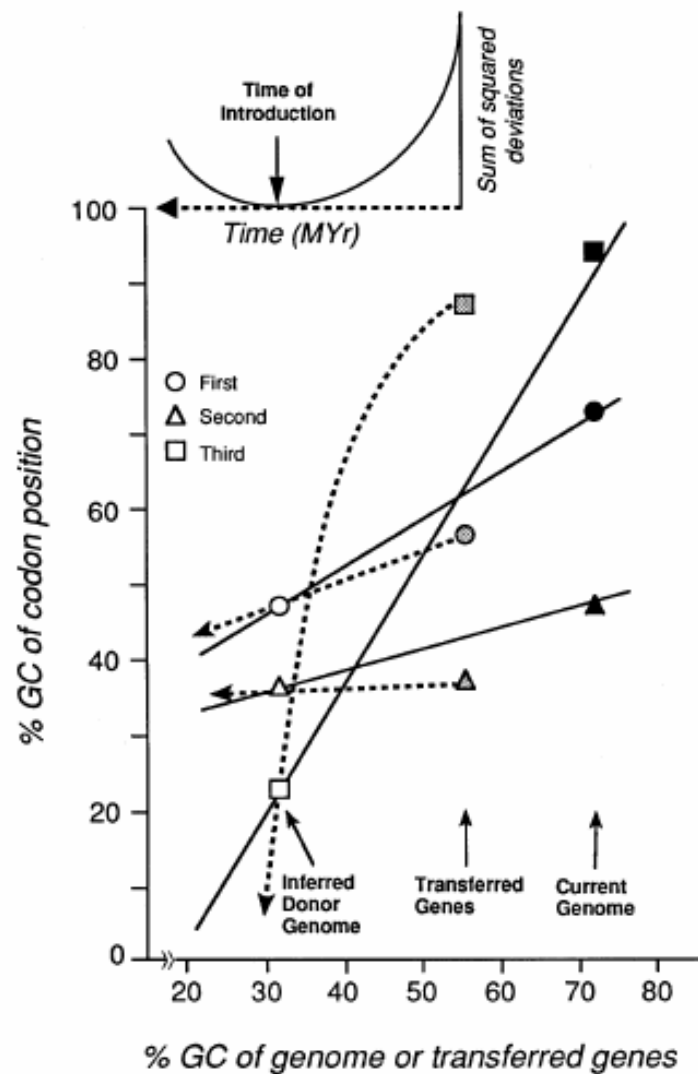
# G+C<sub>1,2,3</sub> VS G+C



$GC_1 = 0.615 \times GC + 26.9$   
 $GC_2 = 0.270 \times GC + 26.7$   
 $GC_3 = 1.692 \times GC - 32.3$

Lawrence 1997

# Amelioration ... in reverse



Lawrence 1997

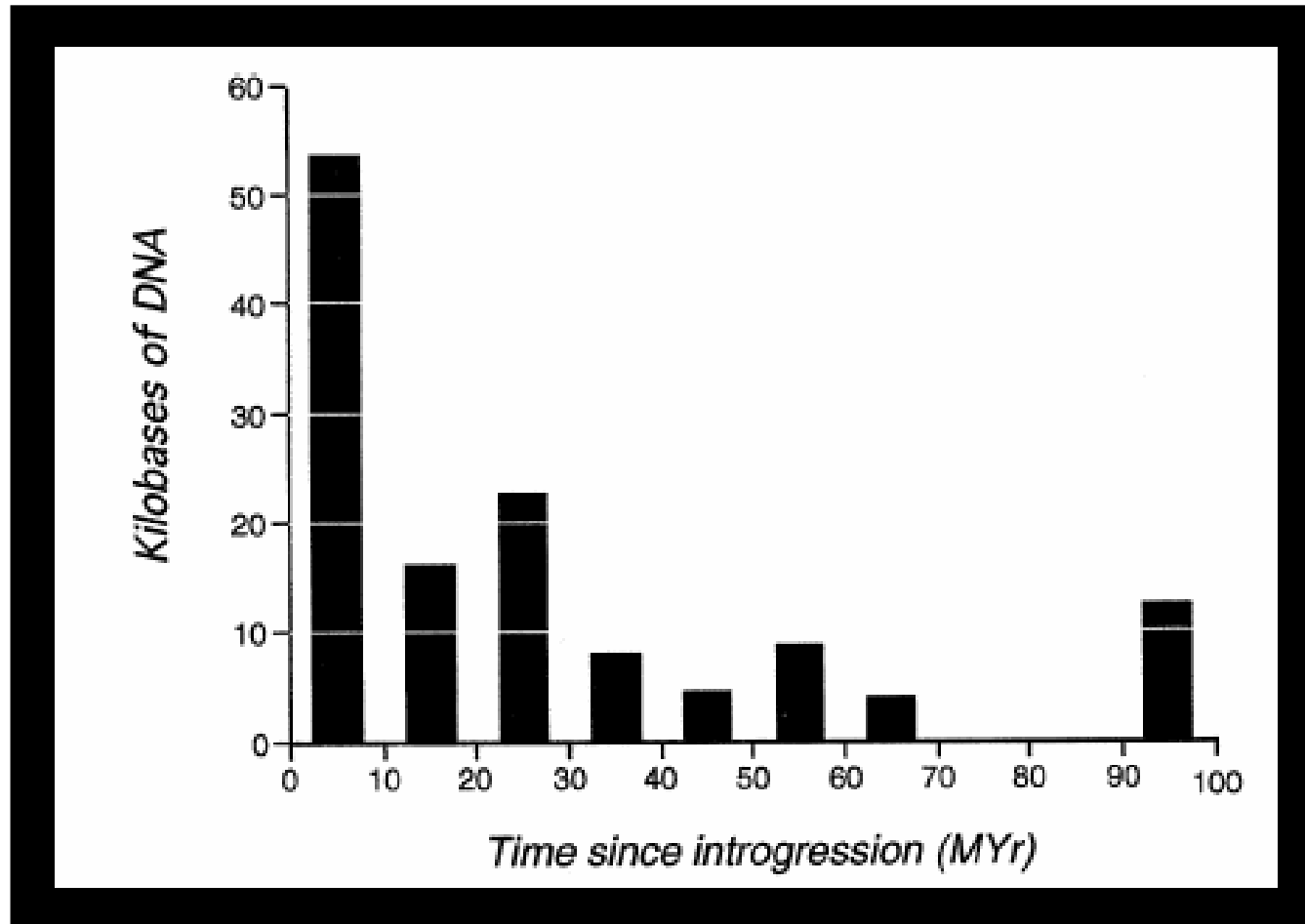
# Amelioration ... in reverse HGTs in *E.coli*

Pool	Genes <sup>a</sup>	Bases <sup>b</sup>	Average GC content			Time ameliorating <sup>d</sup> (Myr)	Range <sup>e</sup> (Myr)	
			ORF	1 <sup>c</sup>	2 <sup>c</sup>			3 <sup>c</sup>
1	4	2682	29.1	37.1	27.3	23.0	17	12–24
2	9	7479	33.8	41.9	33.3	26.3	26	24–29
3	2	879	32.8	34.8	31.4	32.1	22	14–34
4	6	6135	33.7	39.5	30.2	31.3	na <sup>f</sup>	
5	5	6768	36.7	44.8	32.4	33.1	93	85–109
6	4	4470	39.7	49.6	37.7	31.7	0	0–4
7	4	2586	35.2	35.0	26.8	39.4	na	
8	7	4446	38.2	44.1	34.2	36.4	na	
9	8	6414	41.6	48.9	37.6	38.2	17	12–22
10	7	5595	44.2	54.3	40.3	38.0	0	0–4
11	5	3480	45.6	59.7	39.6	37.4	na	
12	11	6243	41.9	45.0	37.9	42.9	na	
13	8	7620	43.3	48.7	38.7	42.5	37	29–48
14	7	6492	44.5	51.6	40.4	41.6	22	17–26
15	9	6660	45.0	55.5	36.8	42.8	15	9–24
16	11	5118	46.6	59.8	37.0	42.8	na	
17	7	4056	49.6	65.1	39.8	43.9	na	
18	9	8367	44.4	45.9	39.0	48.2	54	42–69
19	6	4893	45.9	49.4	40.4	47.9	na	
20	10	8943	46.5	52.5	39.0	47.9	3	0–12
21	12	9712	47.7	56.0	39.8	47.4	2	0–7
22	10	6471	49.4	60.2	39.9	48.0	0	0–5
23	6	5097	50.4	64.3	38.6	48.3	na	
24	7	3894	48.3	50.4	42.2	52.5	63	58–72
25	8	5304	53.1	62.6	40.8	55.9	98	91–112
26	4	4281	51.0	53.4	37.5	62.2	48	42–57
27	6	7533	54.1	60.9	38.6	62.8	0	0–3
28	4	7350	57.3	61.3	43.8	66.9	26	24–29
29	8	6777	58.2	65.5	42.5	66.7	4	0–13
30	6	6687	57.1	59.0	41.3	71.2	na	
31	9	5802	60.2	65.4	43.7	71.4	0	0–4
32	4	2706	58.4	56.6	42.0	76.6	na	
33	5	4521	61.6	67.9	41.6	75.4	1	0–5
AVE <sup>g</sup>	229	185461					25.2	

Lawrence 1997



# *Amelioration ... in reverse HGTs in E.coli*



# ***Amelioration equations (1)***

$$S = R_{GC} + R_{AT} \quad (1)$$

$$R_{GC} = R_{GC \rightarrow AT} + R_{G,C \rightarrow C,G} \quad (2a)$$

$$R_{AT} = R_{AT \rightarrow GC} + R_{A,T \rightarrow T,A} \quad (2b)$$

# Amelioration equations (1)

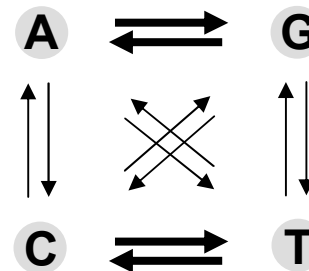
$$S = R_{GC} + R_{AT} \quad (1)$$

$$R_{GC} = R_{GC \rightarrow AT} + R_{G,C \rightarrow C,G} \quad (2a)$$

$$R_{AT} = R_{AT \rightarrow GC} + R_{A,T \rightarrow T,A} \quad (2b)$$

$$R_{GC} = R_{GC \rightarrow AT} + \left[ \frac{1}{2 \times \text{IV ratio} + 1} \times R_{GC \rightarrow AT} \right]$$

$$= \frac{\text{IV ratio} + 1}{\text{IV ratio} + 1/2} \times R_{GC \rightarrow AT} \quad (3)$$



All transition mutations, and half of the transversions, change the GC content of DNA

IV ratio: transition/transversion rate

# Amelioration equations (1)

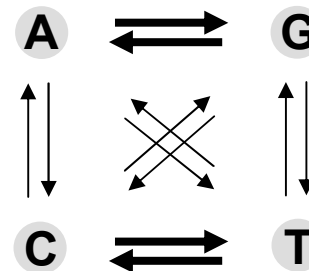
$$S = R_{GC} + R_{AT} \quad (1)$$

$$R_{GC} = R_{GC \rightarrow AT} + R_{G,C \rightarrow C,G} \quad (2a)$$

$$R_{AT} = R_{AT \rightarrow GC} + R_{A,T \rightarrow T,A} \quad (2b)$$

$$R_{GC} = R_{GC \rightarrow AT} + \left[ \frac{1}{2 \times \text{IV ratio} + 1} \times R_{GC \rightarrow AT} \right]$$

$$= \frac{\text{IV ratio} + 1}{\text{IV ratio} + 1/2} \times R_{GC \rightarrow AT} \quad (3)$$



All transition mutations, and half of the transversions, change the GC content of DNA

**IV ratio:** transition/transversion rate

$$S = \left[ \frac{\text{IV ratio} + 1}{\text{IV ratio} + 1/2} \times R_{GC \rightarrow AT} \right]$$

$$+ \left[ \frac{\text{IV ratio} + 1}{\text{IV ratio} + 1/2} \times R_{AT \rightarrow GC} \right] = \frac{\text{IV ratio} + 1}{\text{IV ratio} + 1/2}$$

$$\times [R_{GC \rightarrow AT} + R_{AT \rightarrow GC}] \quad (4)$$

## Amelioration equations (2)

$$GC^{EQ} = \frac{R_{AT \rightarrow GC}}{R_{AT \rightarrow GC} + R_{GC \rightarrow AT}} \quad (5a)$$

$$AT^{EQ} = \frac{R_{GC \rightarrow AT}}{R_{AT \rightarrow GC} + R_{GC \rightarrow AT}} \quad (5b)$$

$$R_{AT \rightarrow GC} + R_{GC \rightarrow AT} = \frac{R_{AT \rightarrow GC}}{GC^{EQ}} = \frac{R_{GC \rightarrow AT}}{AT^{EQ}} \quad (6)$$

$$S = \frac{\text{IV ratio} + 1}{\text{IV ratio} + 1/2} \times \frac{R_{AT \rightarrow GC}}{GC^{EQ}} \quad (7a)$$

$$S = \frac{\text{IV ratio} + 1}{\text{IV ratio} + 1/2} \times \frac{R_{GC \rightarrow AT}}{AT^{EQ}} \quad (7b)$$

## Amelioration equations (3)

$$R_{AT \rightarrow GC} = S \times \frac{\text{IV ratio} + 1/2}{\text{IV ratio} + 1} \times GC^{EQ} \quad (8a)$$

$$R_{GC \rightarrow AT} = S \times \frac{\text{IV ratio} + 1/2}{\text{IV ratio} + 1} \times AT^{EQ} \quad (8b)$$

$$\Delta GC^{HT} = [AT^{HT} \times R_{AT \rightarrow GC}] - [GC^{HT} \times R_{GC \rightarrow AT}] \quad (9)$$

$$\Delta GC^{HT} = S \times \frac{\text{IV ratio} + 1/2}{\text{IV ratio} + 1} \times [AT^{HT} \times GC^{EQ} - GC^{HT} \times AT^{EQ}] \quad (10)$$

the change in the GC content over time can be expressed as the gain in GC nucleotides minus the loss of GC nucleotides

## Amelioration equations (4)

$$\Delta GC^{HT} = S \times \frac{IV \text{ ratio} + 1/2}{IV \text{ ratio} + 1} \times [(1 - GC^{HT}) \times GC^{EQ} - GC^{HT} \times (1 - GC^{EQ})] \quad (11)$$

$$\Delta GC^{HT} = S \times \frac{IV \text{ ratio} + 1/2}{IV \text{ ratio} + 1} \times [GC^{EQ} - GC^{HT}] \quad (12)$$

$$GC_{1st} = 0.615 \times GC_{\text{Genome}} + 26.9 \quad (13)$$

$$GC_{2nd} = 0.270 \times GC_{\text{Genome}} + 26.7 \quad (14)$$

$$GC_{3rd} = 1.692 \times GC_{\text{Genome}} - 32.3 \quad (15)$$

( $\Delta GC$ ) the rate of change in GC content of horizontally transferred DNA is **proportional** both to the substitution rate ( $S$ ) and to the magnitude of the **difference** in nucleotide composition between **horizontally** transferred DNA and **equilibrium** values.

# *Summary of findings*

- 600 kb
- 31 kb / Myr
- 3Mb
- Average time of introgression 25.2 Myr



# *Amelioration in reverse (demo)*

