

Organisation et fluidité des génomés bactériens

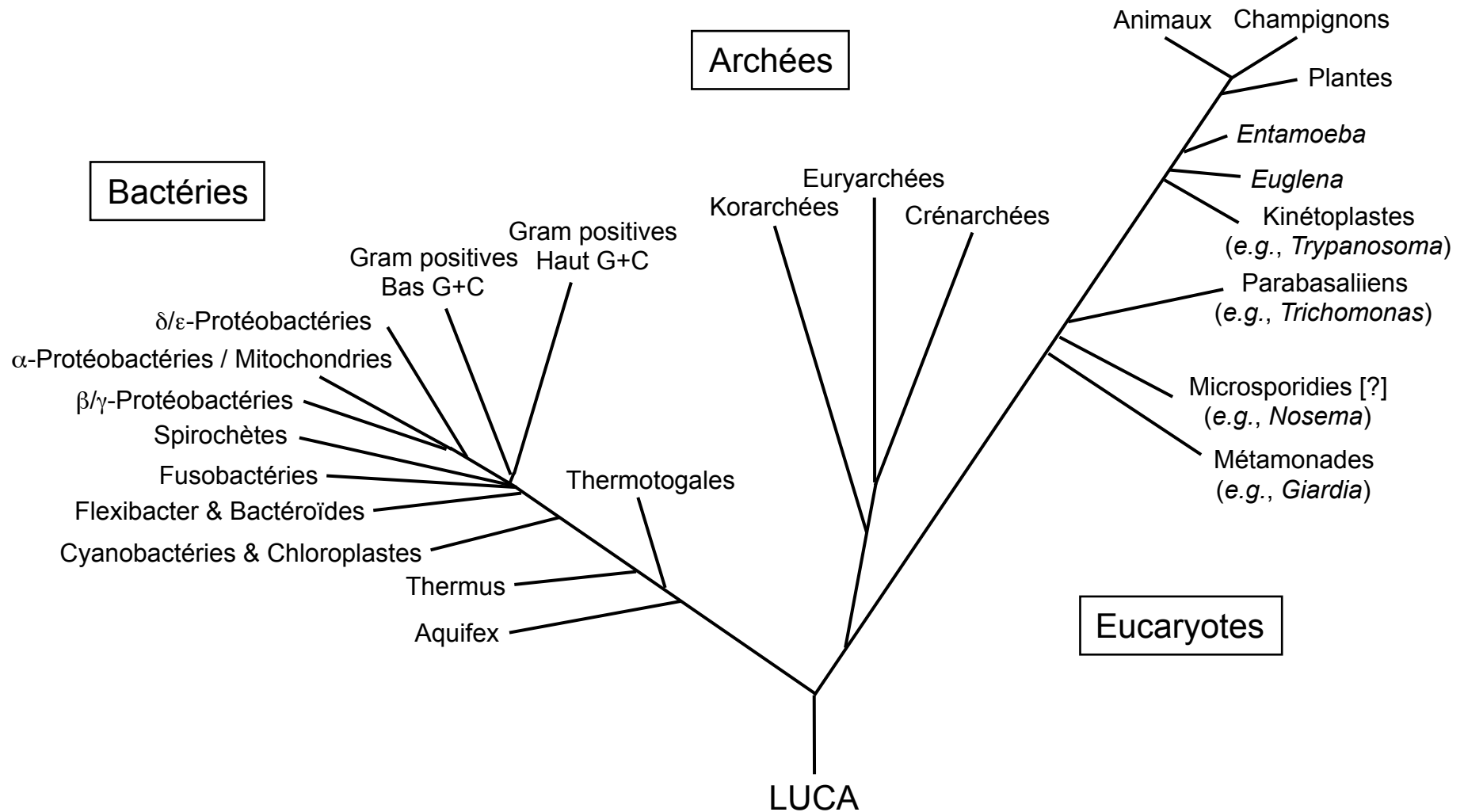
Guy Perrière

Pôle Rhône-Alpes de Bioinformatique
Laboratoire de Biométrie et Biologie Évolutive
UMR CNRS 5558

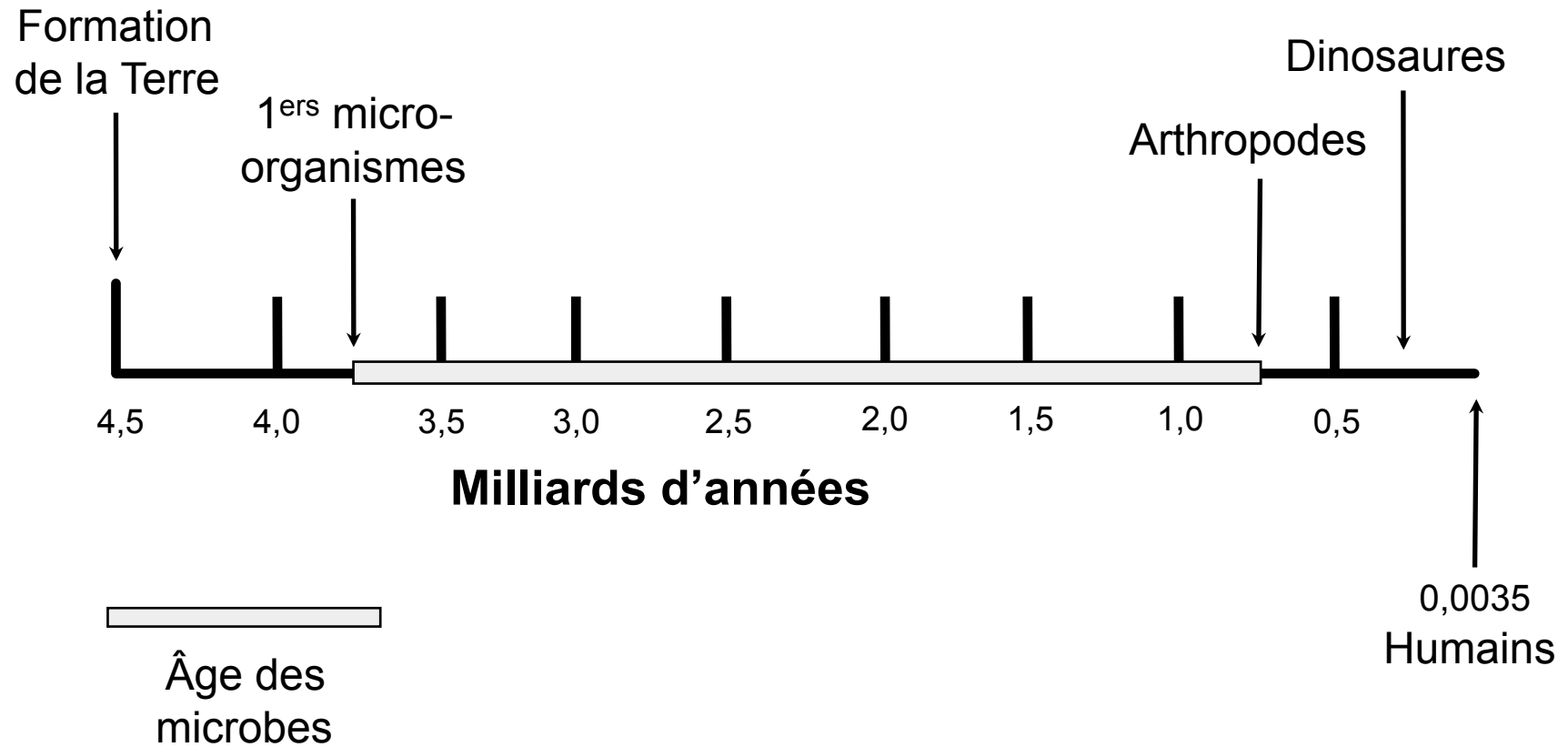
<http://pbil.univ-lyon1.fr/members/perriere/cours/MSBM>



Les trois domaines du vivant

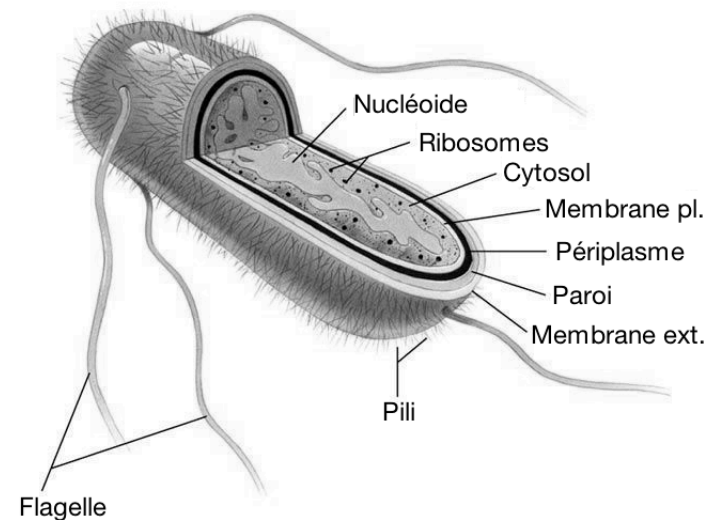


Histoire de la vie sur Terre



Les procaryotes

- Comprennent les bactéries et les archées :
 - Organisation générale comparable :
 - Organismes unicellulaires dépourvus de noyau.
 - Différences au niveau :
 - Des mécanismes de répllication et de traduction.
 - Des phospholipides de la membrane plasmique.



Structure générale d'une cellule procaryote

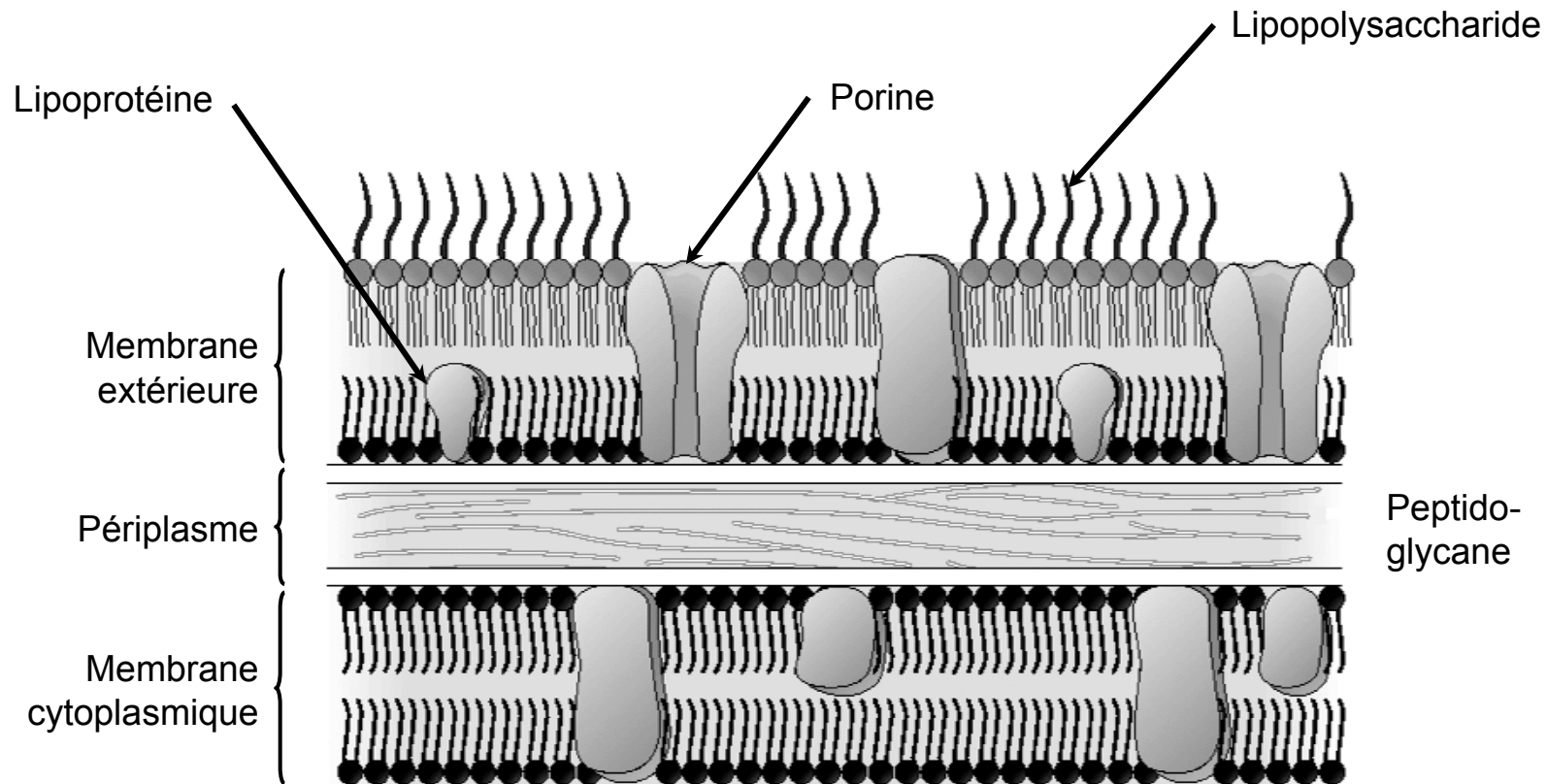
Comment les classer

- Critères morphologiques inopérants.
- Pendant longtemps, utilisation de critères biochimiques:
 - Coloration (Gram positives et Gram négatives).
 - Fonctions biochimiques (principe des galeries Api).
- Depuis 1977, utilisation des données de séquences.

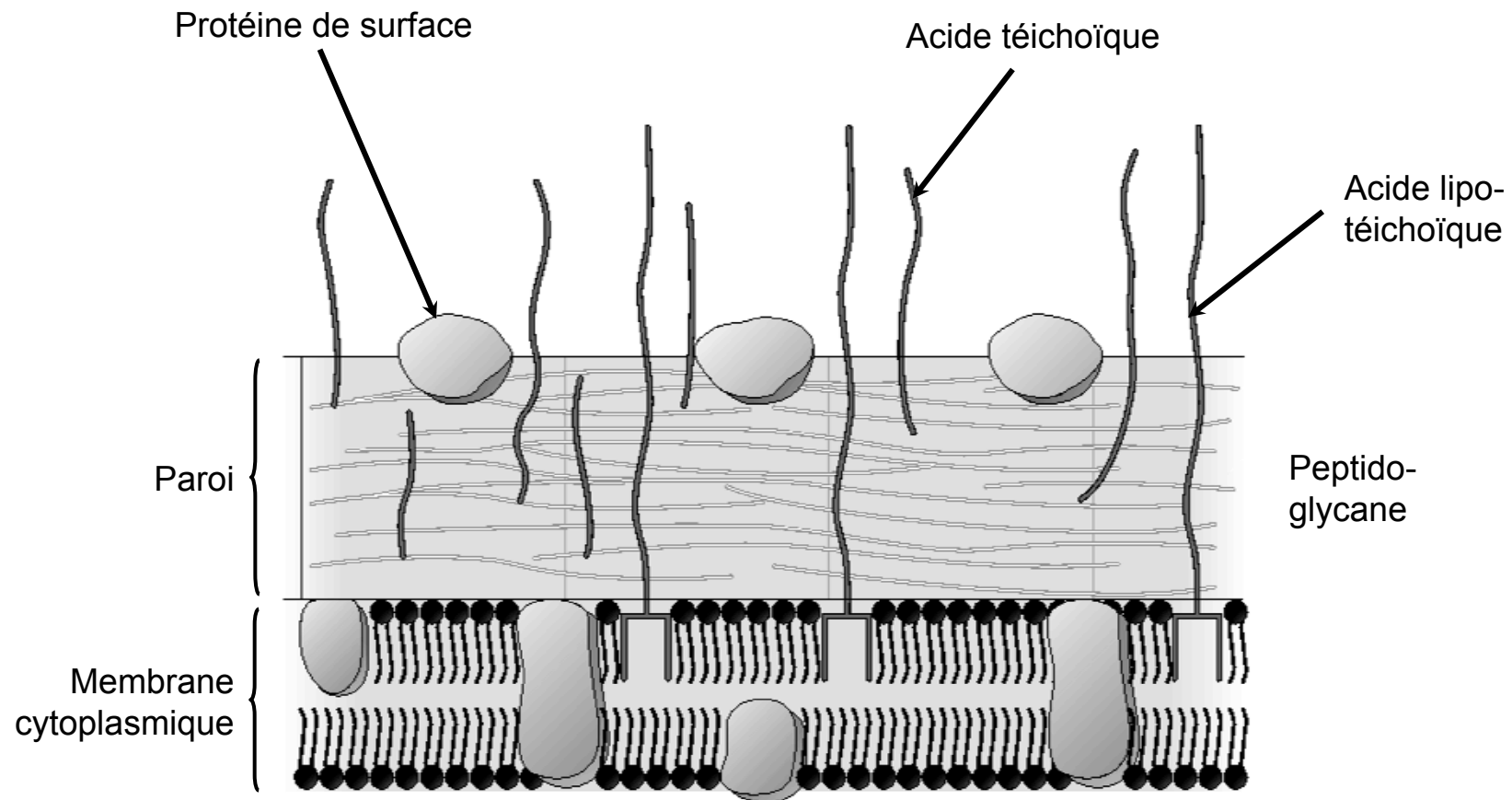
La paroi bactérienne

- La plupart des bactéries possèdent une paroi :
 - Structure de type peptidoglycane (polymère de glucides et d'acides aminés) :
 - Organisation différente chez les Gram positives et les Gram négatives.
 - Exceptions :
 - Certaines bactéries du genre *Mycoplasma* :
 - ✓ Membrane plasmique renforcée.
 - ✓ Vivent dans un milieu où la pression osmotique est faible (hôte).

Paroi des Gram négatives



Paroi des Gram positives



Diversité métabolique

- Animaux, plantes et champignons n'utilisent qu'une seule voie pour récupérer le carbone et produire de l'ATP.
- Bactéries et archées utilisent plusieurs voies métaboliques différentes :
 - Sources de carbone organiques ou inorganiques.
 - Production d'ATP par photosynthèse ou par oxydation de composés organiques ou inorganiques :
 - Augmentation de la quantité d'ATP produite en fonction de la différence d'énergie libre (ΔG) entre le donneur et l'accepteur d'électrons.

Donneurs et accepteurs

Donneur	Accepteur	Produit	Catégorie
H ₂ ou composés organiques	SO ₄ ²⁻	SH ₂	Réducteurs de sulfate
H ₂	CO ₂	CH ₄	Méthanogènes
CH ₄	O ₂	CO ₂	Méthanotrophes
S ₂ ou SH ₂	O ₂	SO ₄ ²⁻	Bactéries sulfureuses
Composés organiques	Fe ³⁺	Fe ²⁺	Réducteurs de fer
NH ₃	O ₂	NO ₂ ⁻	Nitrificateurs
Composés organiques	NO ₃ ⁻	N ₂ O, NO or N ₂	Dénitrificateurs (ou réducteurs de nitrate)
NO ₂ ⁻	O ₂	NO ₃ ⁻	Nitrosificateurs

Bactéries et archées peuvent utiliser différents types de donneurs et d'accepteurs d'électrons

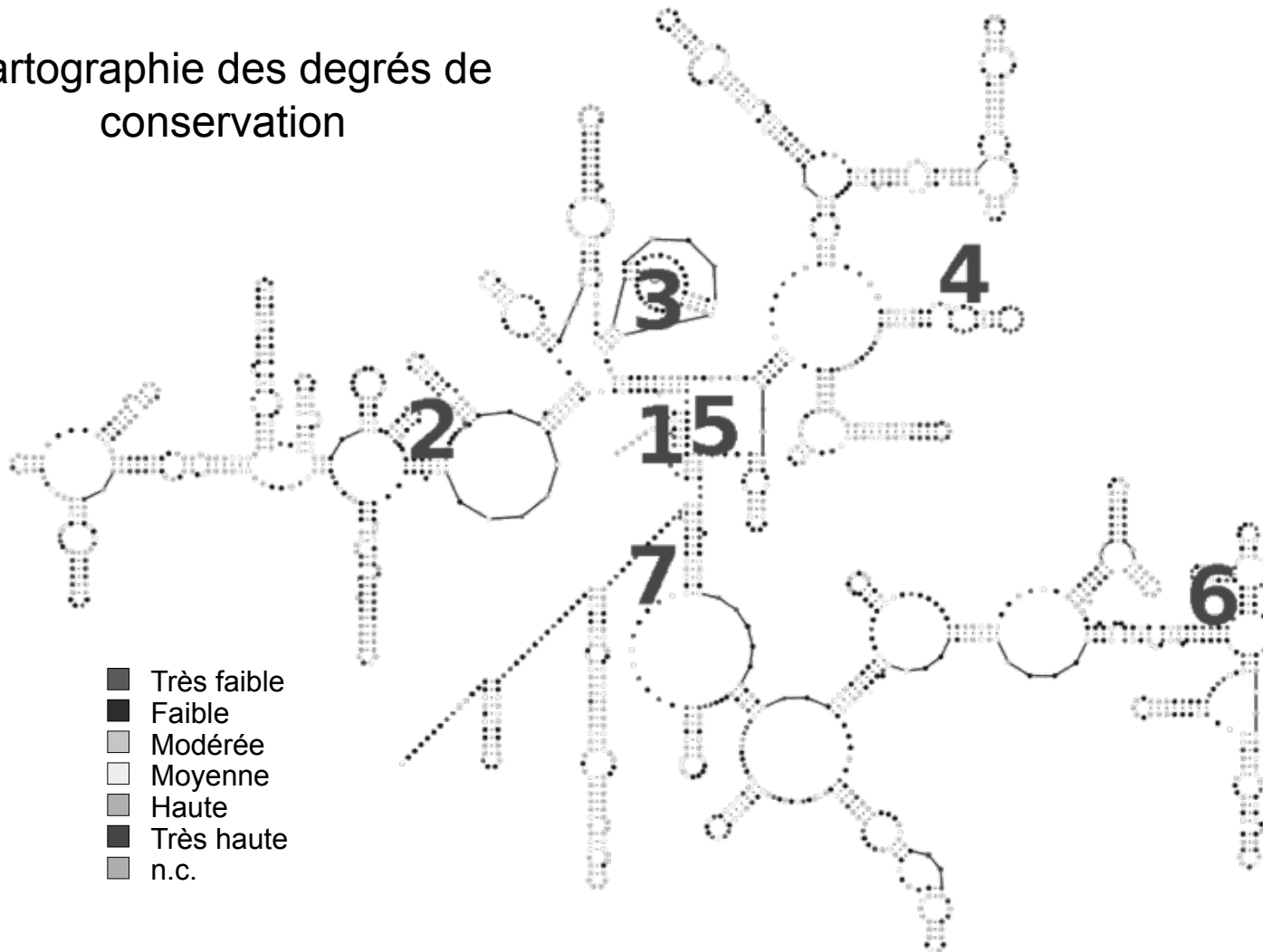
Fonctions et classification

	Utilisation du lactose	Utilisation du citrate	Production de SH ₂	Production d'indole	Lysine décarboxylase		
Escherichia coli	⊕	-	-	⊕	-	⊕	Commensal de mammifères
Shigella flexneri	-	-	-	⊕	-	-	Pathogène de primates
Salmonella enterica	-	⊕	⊕	-	-	⊕	Pathogène de mammifères
Klebsiella pneumoniae	⊕	⊕	-	-	⊕	⊕	Bactérie des sols
Serratia marcescens	-	⊕	-	-	⊕	⊕	Bactérie des sols
Yersinia pestis	-	-	-	-	-	-	Pathogène de mammifères

Ochman *et al.* (2000)

Interêt de l'ARNr 16S

Cartographie des degrés de conservation



Principaux taxons bactériens

Division	Subdivision	Genres représentatifs
Protéobactéries	α-Protéobactéries β-Protéobactéries δ-Protéobactéries ε-Protéobactéries γ-Protéobactéries	<i>Agrobacterium, Rickettsia</i> <i>Neisseria, Ralstonia</i> <i>Myxobacterium</i> <i>Helicobacter, Campylobacter</i> <i>Escherichia, Buchnera, Pseudomonas</i>
Gram positives	Haut G+C Bas G+C	<i>Actinomyces, Streptomyces, Mycobacterium</i> <i>Bacillus, Clostridium, Mycoplasma</i>
Cyanobactéries et apparentées		<i>Nostoc, Synechocystis</i>
Spirochète et apparentées	Spirochètes Leptospiras	<i>Treponema, Borrelia</i> <i>Leptonema, Leptospira</i>
Bactéries vertes sulfureuses		<i>Chlorobium, Chloroherpeton</i>
Bactéroïdes, Cytophagales et Flavobactéries	Bactéroïdes Flavobactéries	<i>Bacteroides, Fusobacterium</i> <i>Cytophaga, Flavobacterium</i>
Planctomycètes et apparentées	Groupe des Planctomycètes Groupe des Thermophiles	<i>Planctomyces, Pasteuria, Pirellula</i> <i>Isosphaera</i>
Chlamydiales		<i>Chlamydia</i>
Micrococcus radiorésistants et apparentées	Deinococcales Groupe des Thermophiles	<i>Deinococcus</i> <i>Thermus</i>
Bactéries vertes non sulfureuses	Groupe des Chloroflexus Groupe des Thermomicrobium	<i>Chloroflexus, Herpetosiphon</i> <i>Thermomicrobium</i>
Aquificales et Thermotogales	Aquificales Thermotogales	<i>Aquifex, Hydrogenobacter</i> <i>Thermotoga, Geotoga, Thermopallium</i>

Bactéries pathogènes

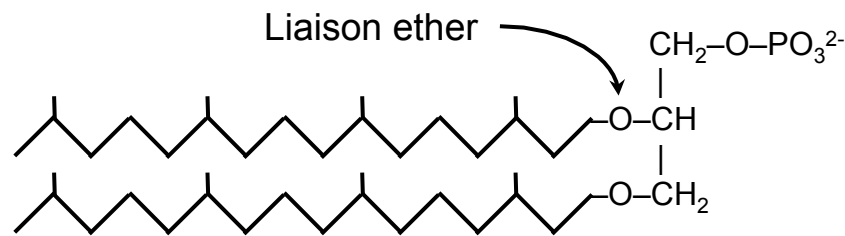
Espèce	Division	Tissus affectés	Maladie
<i>Chlamydia trachomatis</i>	Chlamydiales	Canal uro-génital, yeux	Infection du tractus génital
<i>Clostridium botulinum</i>	Gram positives	Système nerveux, tractus gastro-intestinal	Botulisme
<i>Clostridium tetani</i>	Gram positives	Système nerveux	Tétanos
<i>Haemophilus influenzae</i>	Protéobactéries	Conduit auditif, système nerveux	Infections de l'oreille, méningites
<i>Mycobacterium tuberculosis</i>	Gram positives	Tractus respiratoire	Tuberculose
<i>Neisseria gonorrhoeae</i>	Protéobactéries	Canal uro-génital	Gonorrhée
<i>Propionibacterium acnes</i>	Gram positives	Peau	Acné
<i>Pseudomonas aeruginosa</i>	Protéobactéries	Canal uro-génital, yeux, conduit auditif	Infections du tractus urinaire, des yeux et de l'oreille
<i>Salmonella enteritidis</i>	Protéobactéries	Tractus gastro-intestinal	Intoxication alimentaire
<i>Streptococcus pneumoniae</i>	Gram positives	Tractus respiratoire	Pneumonie
<i>Streptococcus pyogenes</i>	Gram positives	Tractus respiratoire	Scarlatine
<i>Treponema pallidum</i>	Spirochètes	Canal uro-génital	Syphilis
<i>Vibrio parahaemolyticus</i>	Protéobactéries	Tractus gastro-intestinal	Intoxication alimentaire
<i>Yersinia pastis</i>	Protéobactéries	Lymphes et sang	Peste

La découverte des archées

- Avant 1977, considérées comme des bactéries.
- Proposition d'un nouveau domaine du vivant par Woese et Fox (1977).
- Les données moléculaires (ARNr 16S/18S) justifient la séparation.
- Confirmation par de nombreuses études au niveau génomique et/ou biochimique.
- Longtemps supposées ne vivre que dans des milieux extrêmes (température, salinité, pH...)

Paroi des archées

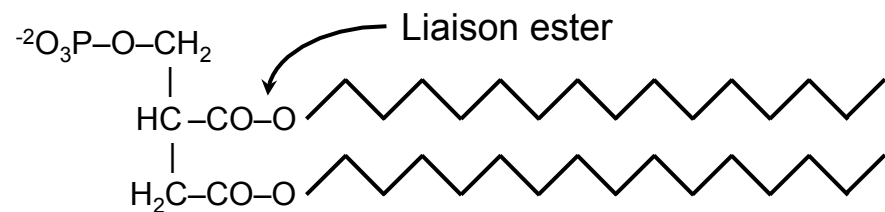
- Structure particulière de la paroi :
 - Pas de peptidoglycane.
 - Diglycérides-phosphates membranaires spécifiques :
 - L-Glycérol-3-phosphate au lieu de D-Glycérol.
 - Isoprènes au lieu d'acides gras.
 - Liaisons avec le glycérol de type ether au lieu d'ester.



Phytanol

L-Glycérol

Archées



D-Glycérol

Acide stéarique

Bactéries

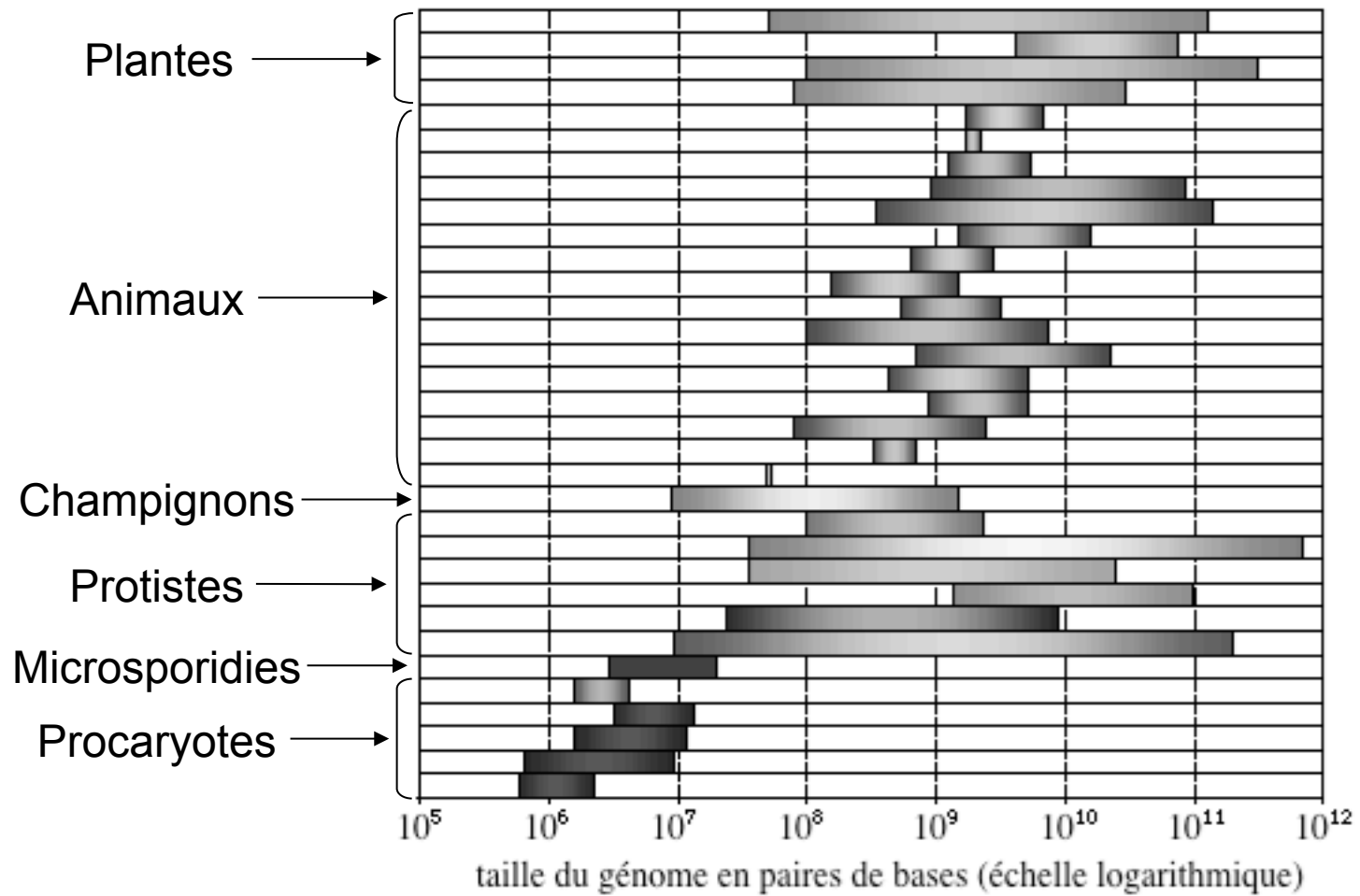
Principaux taxons archéens

Division	Subdivision	Genres représentatifs
Crénarchées	Thermoprotéales	<i>Thermoproteus, Pyrobaculum, Thermofilum</i>
	Sulfolobales	<i>Sulfolobus, Acidianus</i>
	Désulfurococcales	<i>Aeropyrum, Desulfurococcus</i>
	Cénarchéales	<i>Cenarchaeum</i>
	Caldisphérales	<i>Caldisphaera</i>
Euryarchées	Méthanobactériales	<i>Methanobacterium, Methanothermobacter</i>
	Méthanococcales	<i>Methanococcus, Methanothermococcus</i>
	Halobactériales	<i>Halobacterium, Halococcus</i>
	Thermoplasmatales	<i>Thermoplasma, Ferroplasma</i>
	Thermococcales	<i>Pyrococcus, Thermococcus</i>
	Archaeoglobales	<i>Archaeoglobus</i>
	Méthanopyrales	<i>Methanopyrus</i>
	Méthanomicrobiales	<i>Methanogenium</i>
	Méthanosarcinales	<i>Methanosarcina, Methanococcoides</i>
Thaumarchées [?]		<i>Cenarchaeum</i>
Nanoarchées [?]		<i>Nanoarchaeum</i>
Korarchées [?]		<i>Candidatus</i>

Archées pathogènes ?

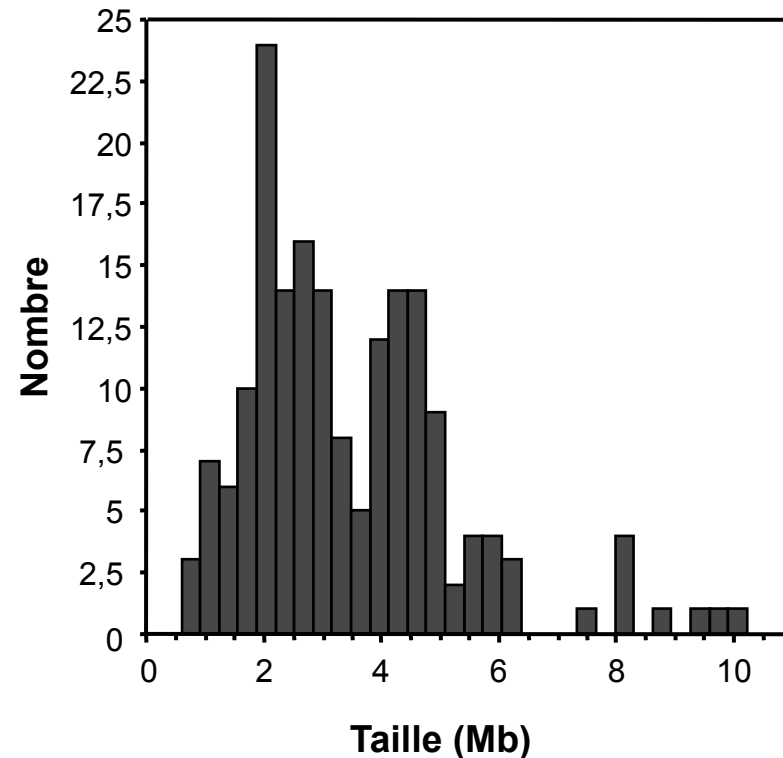
- Aucune archée pathogène (humain ou animal) décrite à ce jour.
- De nombreuses archées méthanogènes sont des commensaux des mammifères.
- Certaines d'entre elles possèdent des homologues de gènes impliqués dans la pathogénicité chez les bactéries :
 - Systèmes de sécrétion.
 - Synthèse de toxines.

Taille des génomes



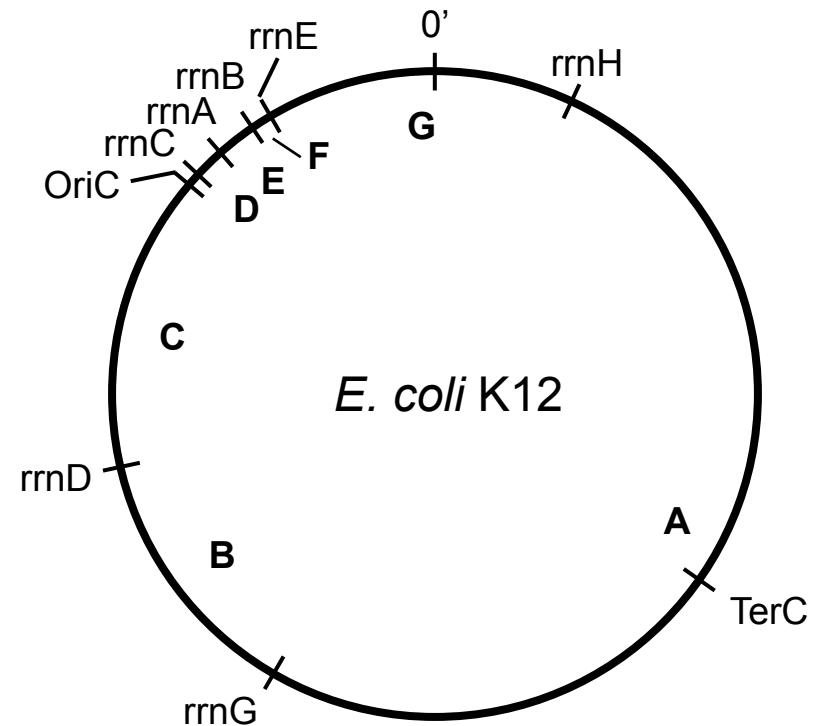
Génomes bactériens

- La variabilité de la taille est grande :
 - Le plus petit : *Carsonella ruddii* (0,16 Mb).
 - Le plus grand : *Sorangium cellulosum* (13,0 Mb).
- Parasites et symbiotes ont un génome plus petit que les bactéries cultivables :
 - Perte de nombreux gènes liés à des voies de biosynthèse.



Polymorphisme de la taille

- Étudié chez *Escherichia coli* :
 - Variation entre 4,5 Mb et 5,5 Mb pour les différentes souches.
 - Étude à l'aide de l'enzyme I-CeuI spécifique des gènes *rrn*.
 - Équidistance entre TerC et OriC de chaque côté du chromosome.

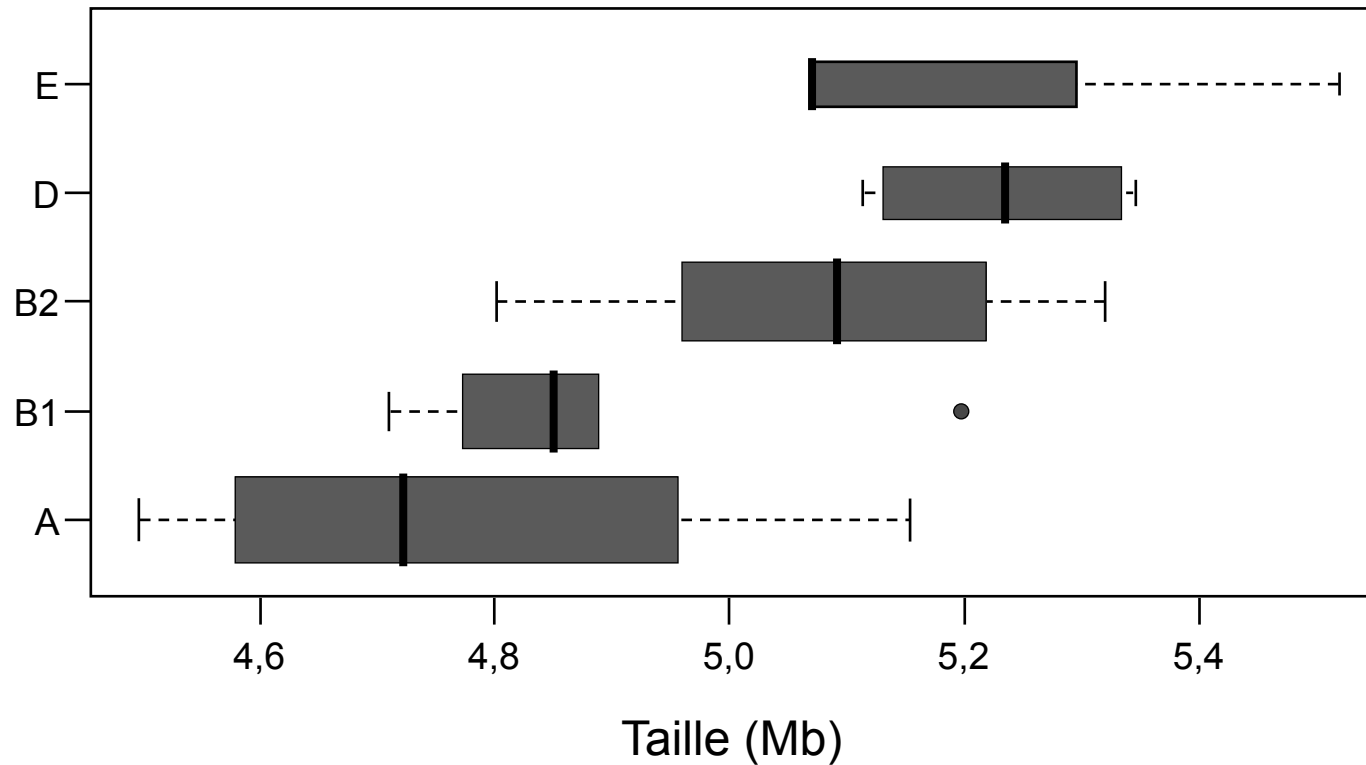


Résultats

Groupe	Souche	Hôte (sexe)	Localisation	A	B	C	D	E	F	G
A	ECOR4	Humain (F)	Iowa	2585	707	527	90	166	38	608
A	ECOR5	Humain (F)	Iowa	2940	743	515	90	128	38	699
A	ECOR11	Humain (F)	Suède	2750	824	556	90	128	38	735
A	ECOR13	Humain (F)	Suède	2485	680	515	90	128	38	639
A	ECOR14	Humain (F)	Suède	2645	735	608	90	128	38	707
A	ECOR15	Humain (F)	Suède	2690	735	575	90	138	38	639
A	ECOR18	Macaque	Washington	2510	699	515	90	122	38	608
A	ECOR19	Macaque	Washington	2480	699	527	90	122	38	639
A	ECOR20	Bœuf	Bali	2505	654	480	90	122	38	608
A	ECOR21	Bœuf	Bali	2505	654	480	90	122	38	608
A	ECOR23	Eléphant	Washington	2675	807	532	90	138	38	680
B1	ECOR27	Girafe	Washington	2600	707	515	90	143	38	616
B1	ECOR28	Humain (F)	Iowa	2620	743	527	94	128	38	639
B1	ECOR29	Rat kangourou	Nevada	2610	787	527	94	138	38	639
B1	ECOR34	Chien	Massachusetts	2500	790	515	94	138	38	680
B1	ECOR58	Lion	Washington	2700	743	515	94	136	38	639
B1	ECOR68	Girafe	Washington	2745	843	532	94	138	38	807
B1	ECOR71	Humain (F)	Suède	2650	771	547	90	138	38	654
B1	ECOR72	Humain (F)	Suède	2635	771	532	94	138	38	680
B2	ECOR51	Humain (enfant)	Massachusetts	2750	810	550	112	138	38	810
...
D	ECOR39	Humain (F)	Suède	2780	787	581	104	143	38	713
D	ECOR40	Humain (F)	Suède	2845	807	616	104	143	43	787
E	ECOR31	Léopard	Washington	2775	743	547	94	138	38	735
E	ECOR37	Marmouset	Washington	3100	787	581	94	175	38	743
E	ECOR42	Humain (M)	Massachusetts	2735	743	616	94	143	38	699

Taille des fragments (A-G) en kb

Variations par sérotype



Distributions de tailles au sein des cinq sérogroupes d'*E. coli*

Le code génétique standard

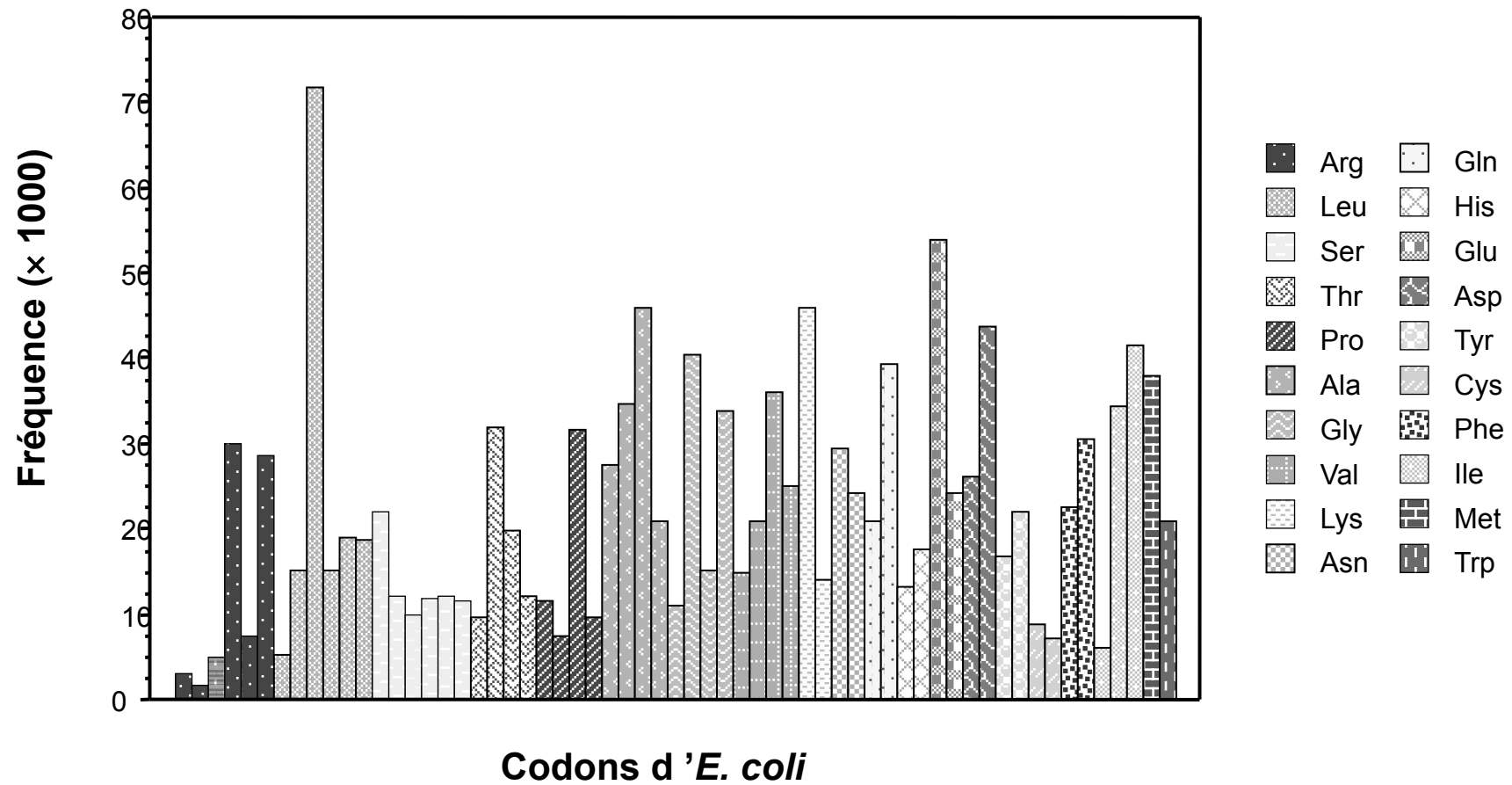
I \ II	U	C	A	G	III
U	UUU Phe F	UCU Ser S	UAU Tyr Y	UGU Cys C	U
	UUC Phe F	UCC Ser S	UAC Tyr Y	UGC Cys C	C
	UUA Leu L	UCA Ser S	UAA Stop	UGA Stop	A
	UUG Leu L	UCG Ser S	UAG Stop	UGG Trp W	G
C	CUU Leu L	CCU Pro P	CAU His H	CGU Arg R	U
	CUC Leu L	CCC Pro P	CAC His H	CGC Arg R	C
	CUA Leu L	CCA Pro P	CAA Gln Q	CGA Arg R	A
	CUG Leu L	CCG Pro P	CAG Gln Q	CGG Arg R	G
A	AUU Ile I	ACU Thr T	AAU Asn N	AGU Ser S	U
	AUC Ile I	ACC Thr T	AAC Asn N	AGC Ser S	C
	AUA Ile I	ACA Thr T	AAA Lys K	AGA Arg R	A
	AUG Met M	ACG Thr T	AAG Lys K	AGG Arg R	G
G	GUU Val V	GCU Ala A	GAU Asp D	GGU Gly G	U
	GUC Val V	GCC Ala A	GAC Asp D	GGC Gly G	C
	GUA Val V	GCA Ala A	GAA Glu E	GGA Gly G	A
	GUG Val V	GCG Ala A	GAG Glu E	GGG Gly G	G

Biais au niveau d'un gène

UUU	Phe	6	UCU	Ser	5	UAU	Tyr	4	UGU	Cys	0
UUC	Phe	10	UCC	Ser	6	UAC	Tyr	12	UGC	Cys	3
UUA	Leu	8	UCA	Ser	8	UAA	Ter	*	UGA	Ter	*
UUG	Leu	6	UCG	Ser	10	UAG	Ter	*	UGG	Trp	12
CUU	Leu	6	CCU	Pro	5	CAU	His	2	CGU	Arg	7
CUC	Leu	9	CCC	Pro	5	CAC	His	3	CGC	Arg	6
CUA	Leu	5	CCA	Pro	4	CAA	Gln	9	CGA	Arg	6
CUG	Leu	2	CCG	Pro	3	CAG	Gln	9	CGG	Arg	3
AUU	Ile	1	ACU	Thr	11	AAU	Asn	2	AGU	Ser	4
AUC	Ile	8	ACC	Thr	5	AAC	Asn	15	AGC	Ser	3
AUA	Ile	7	ACA	Thr	5	AAA	Lys	5	AGA	Arg	3
AUG	Met	7	ACG	Thr	6	AAG	Lys	9	AGG	Arg	4
GUU	Val	8	GCU	Ala	6	GAU	Asp	8	GGU	Gly	15
GUC	Val	7	GCC	Ala	12	GAC	Asp	5	GGC	Gly	6
GUA	Val	7	GCA	Ala	7	GAA	Glu	5	GGA	Gly	2
GUG	Val	9	GCG	Ala	10	GAG	Glu	12	GGG	Gly	5

Protéine A du phage MS2 (Fiers *et al.*, 1975)

Biais au niveau d'un génome

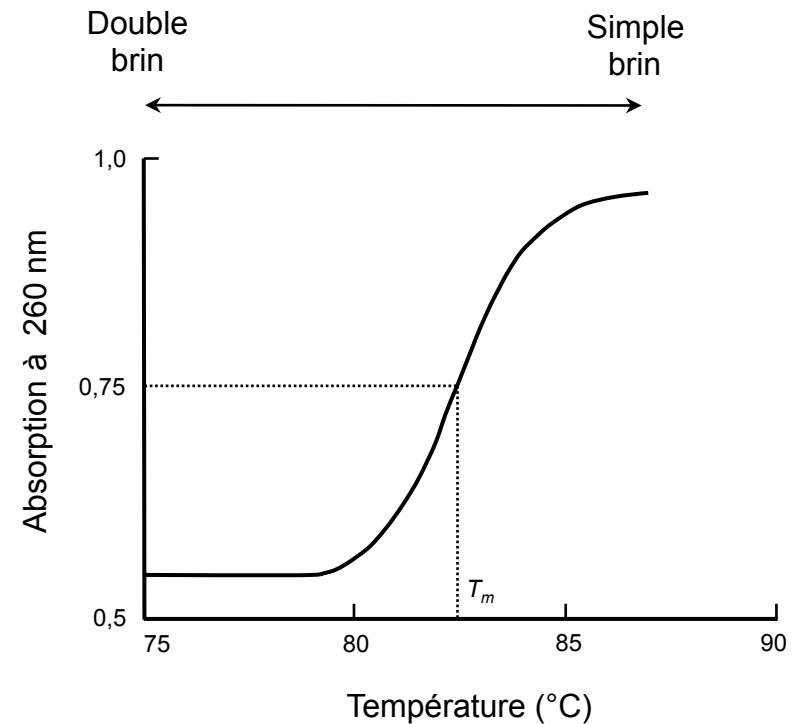


Facteurs connus

- Les facteurs influençant la composition en codons chez les bactéries sont multiples :
 - Contenu global en G+C du génome.
 - Sélection traductionnelle.
 - Localisation sur l'un ou l'autre des deux brins du chromosome.
 - Distance à l'origine de réplication.
 - Transferts horizontaux.

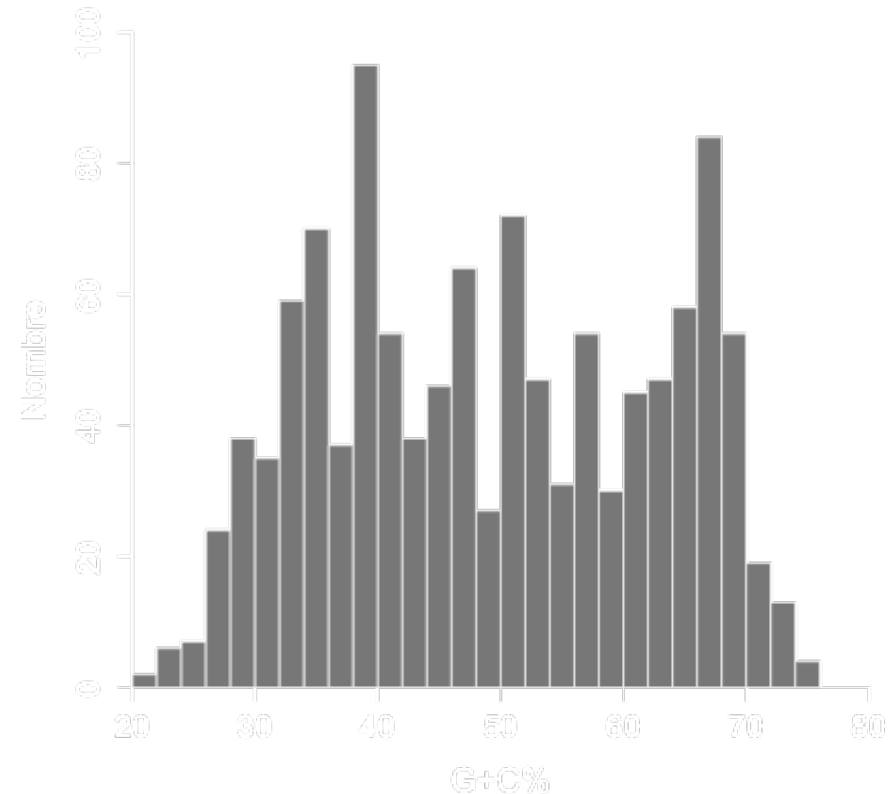
Contenu global en G+C

- Calculé en pourcentage de bases G+C :
 - Constitue une des premières mesures moléculaires appliquée à la systématique.
- Méthodes de dénaturation par la chaleur :
 - Mesure de la variation de l'absorption UV en fonction de la température.
 - La valeur du T_m est linéairement proportionnelle au G+C% de l'organisme.



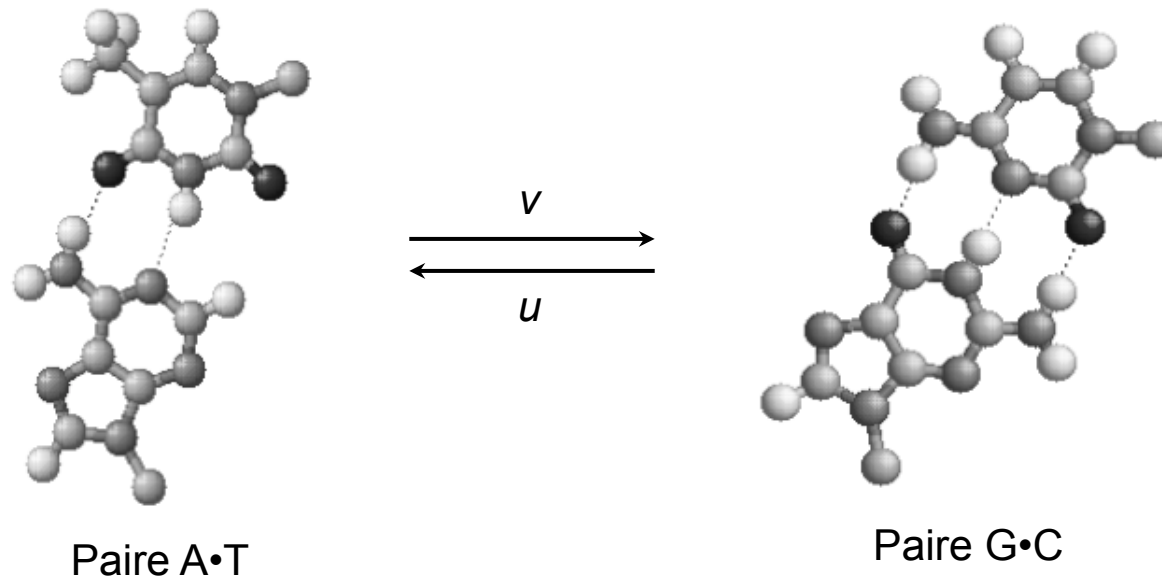
Distribution

- Mesure faite sur 1160 bactéries.
- Minimum : *Buchnera aphidicola* (20 %).
- Maximum : *Micrococcus luteus* (77 %).
- Variations de 5 % et 10 % au sein d'une espèce et d'un genre.



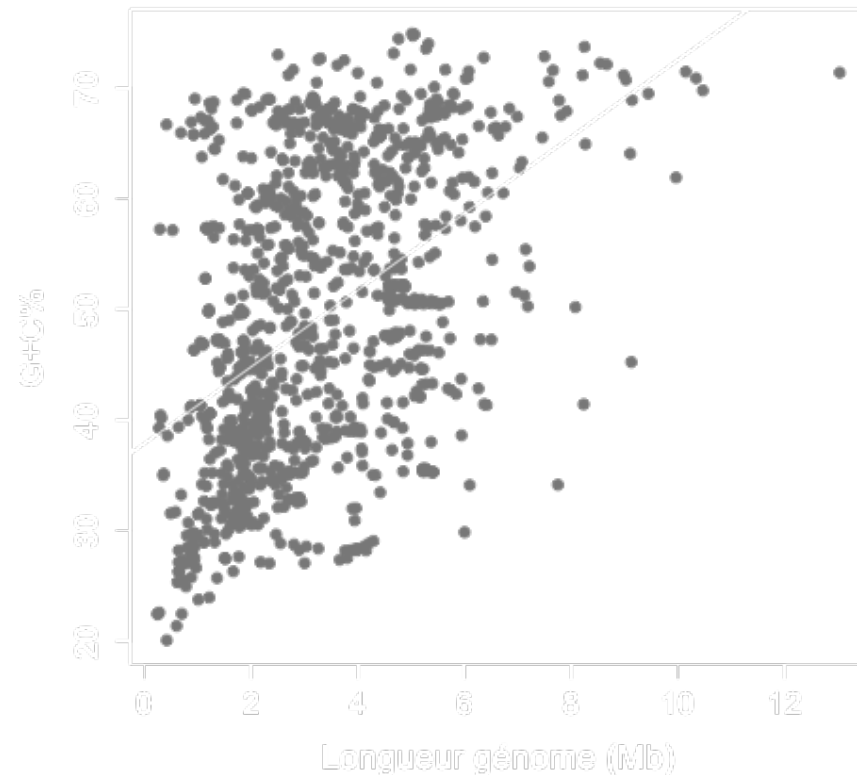
Source de la variation

- Différences de pression de mutation entre les paires A•T et G•C.
- Le ratio u/v est différent d'une espèce à l'autre :
 - Distribution observée chez les bactéries.

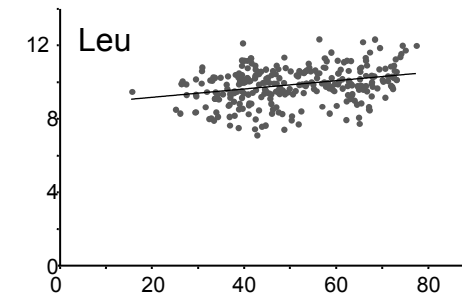
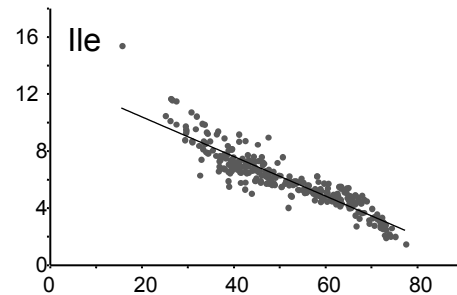
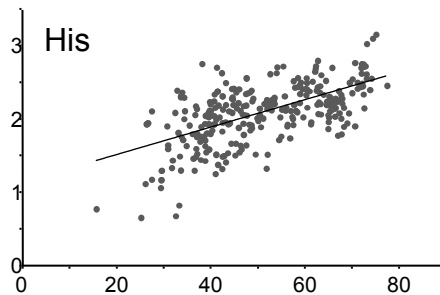
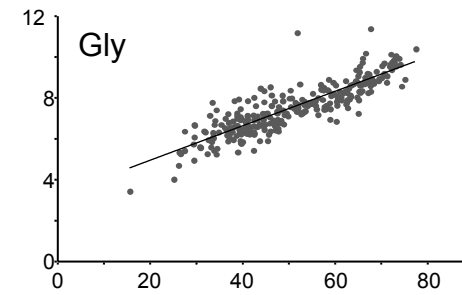
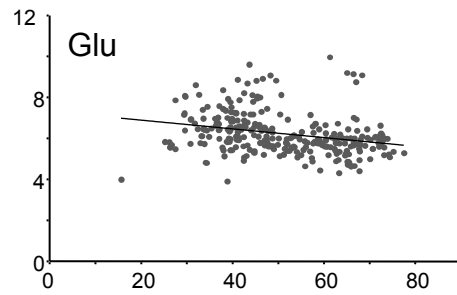
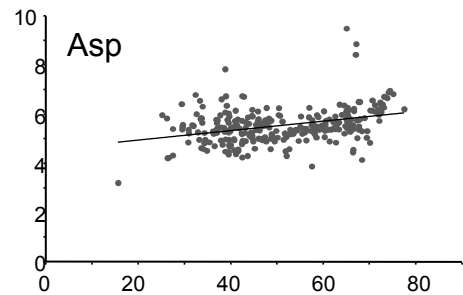
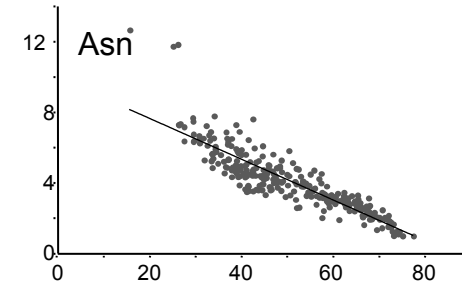
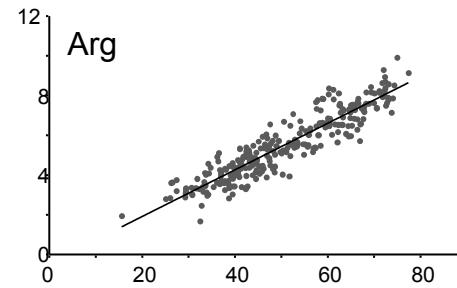
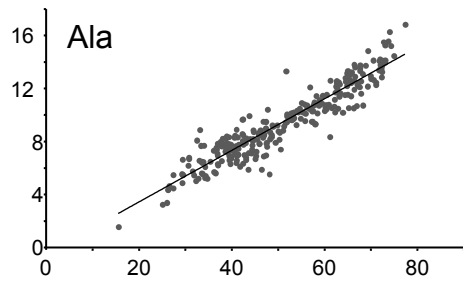


G+C et taille des génomes

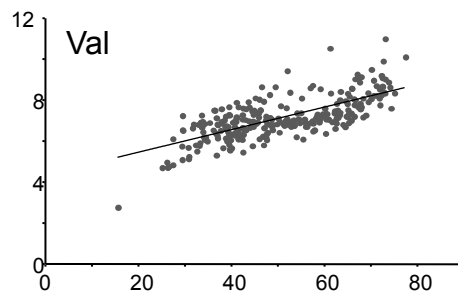
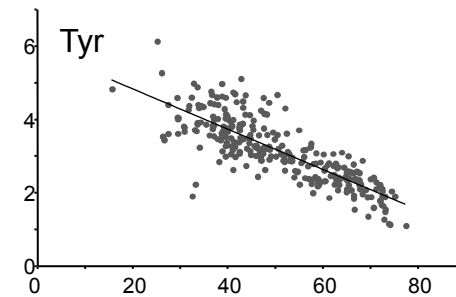
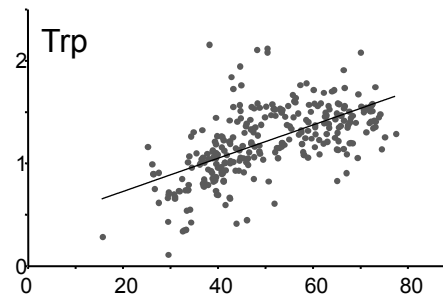
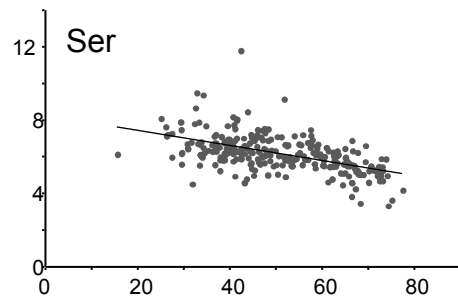
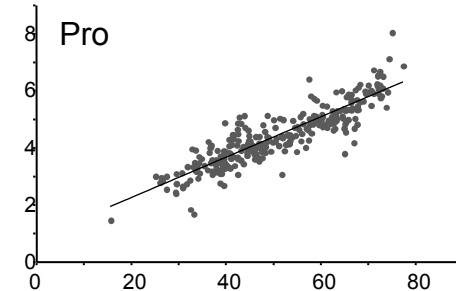
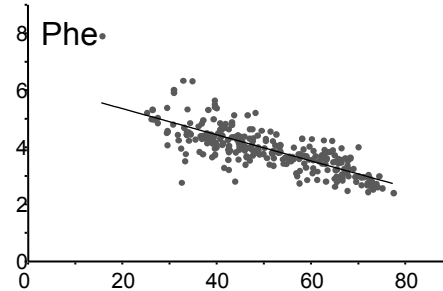
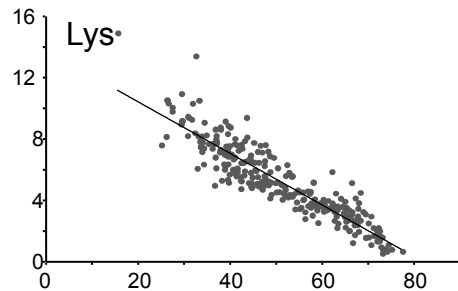
- Corrélation taille génome *vs.* G+C%.
- Génomes de parasites et de symbiotes :
 - Perte d'au moins un mécanisme de réparation :
 - Dérive mutationnelle vers A+T.
 - Accessibilité plus réduite aux dNTP « coûteux ».



G+C et contenu en AA

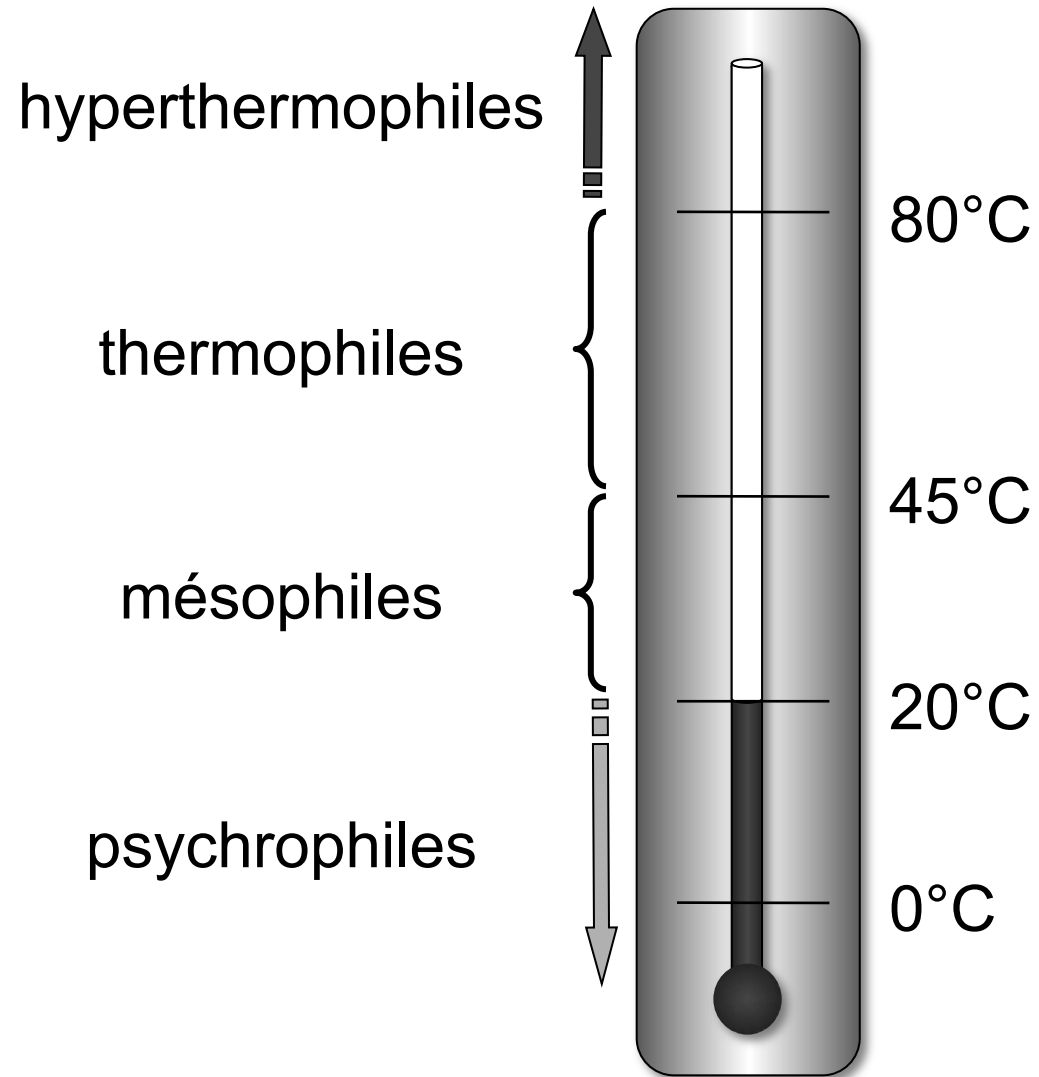


G+C global et contenu en AA

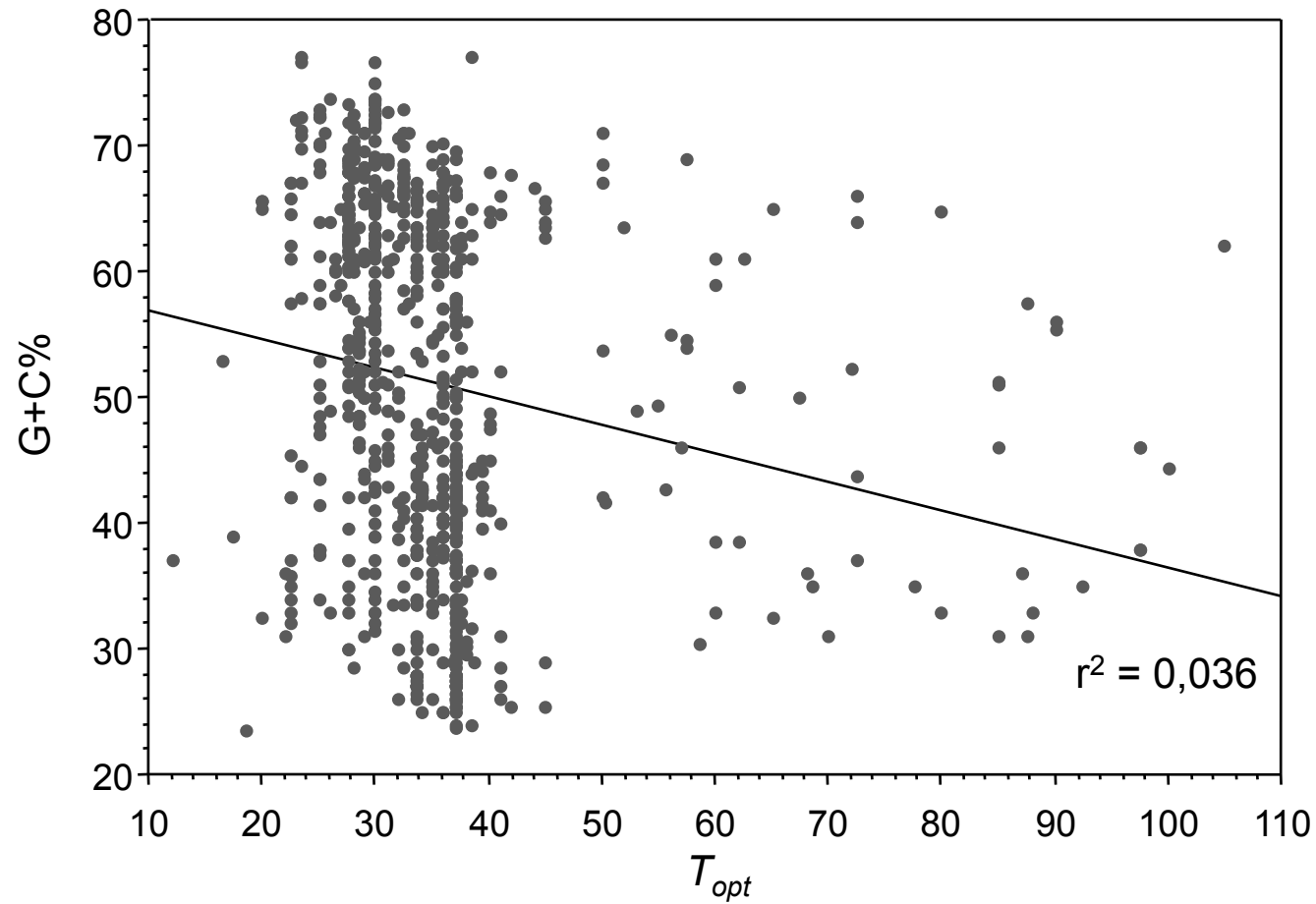


Quatre AA non corrélés au G+C% : Cys, Gln, Met, Thr

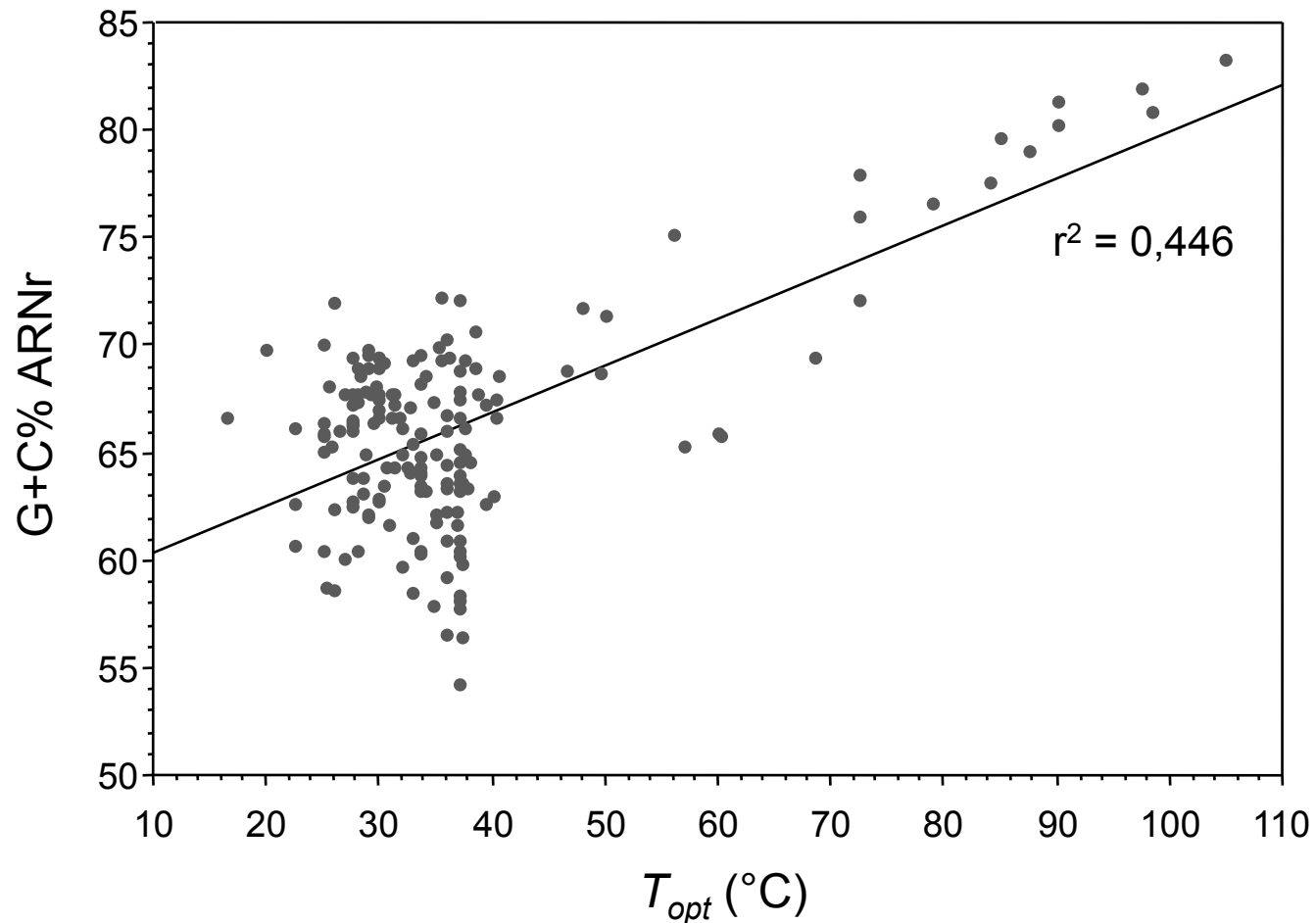
Température de croissance



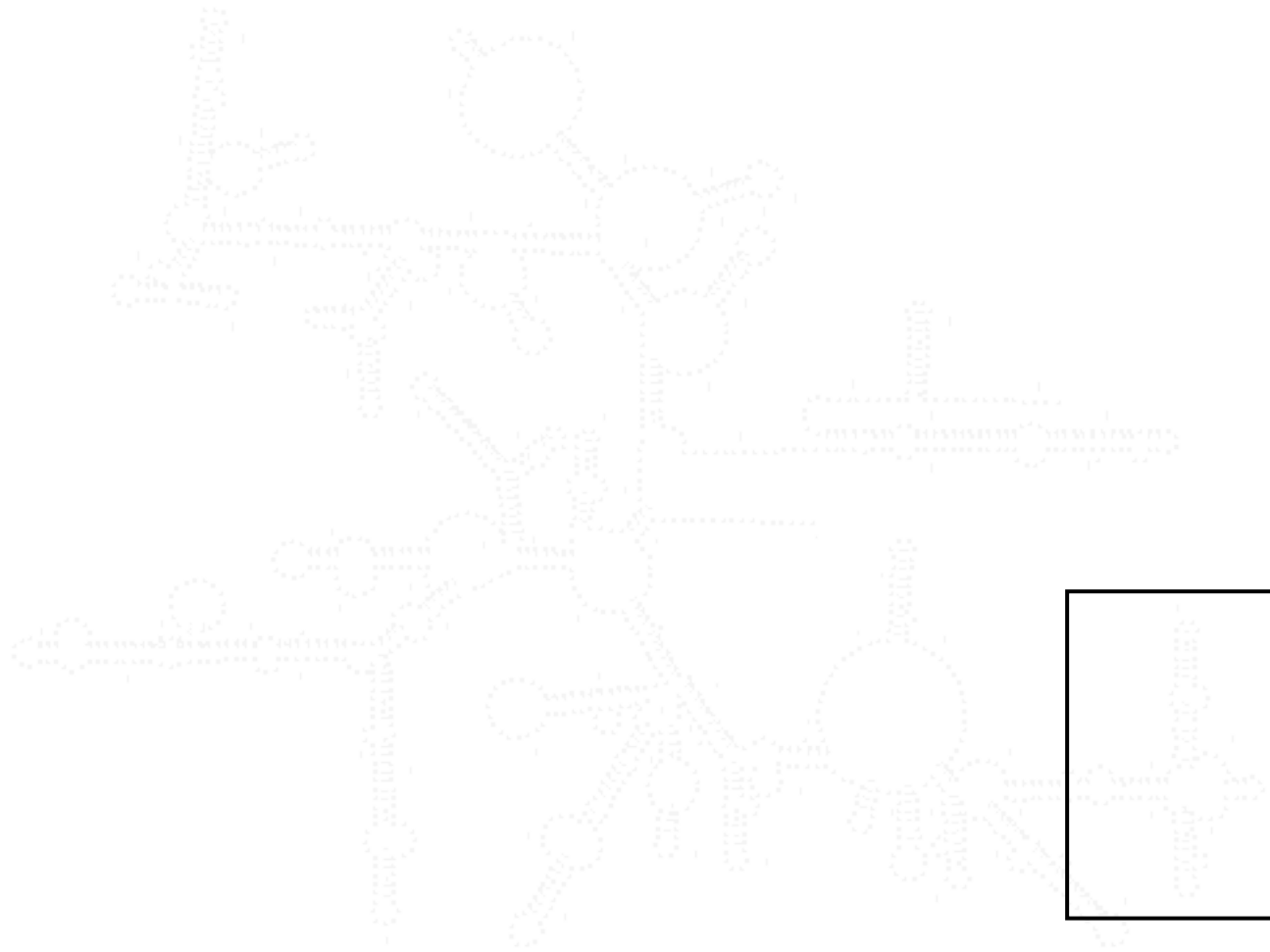
Pas de lien G+C% / T_{opt}



Relation G+C% ARNr / T_{opt}



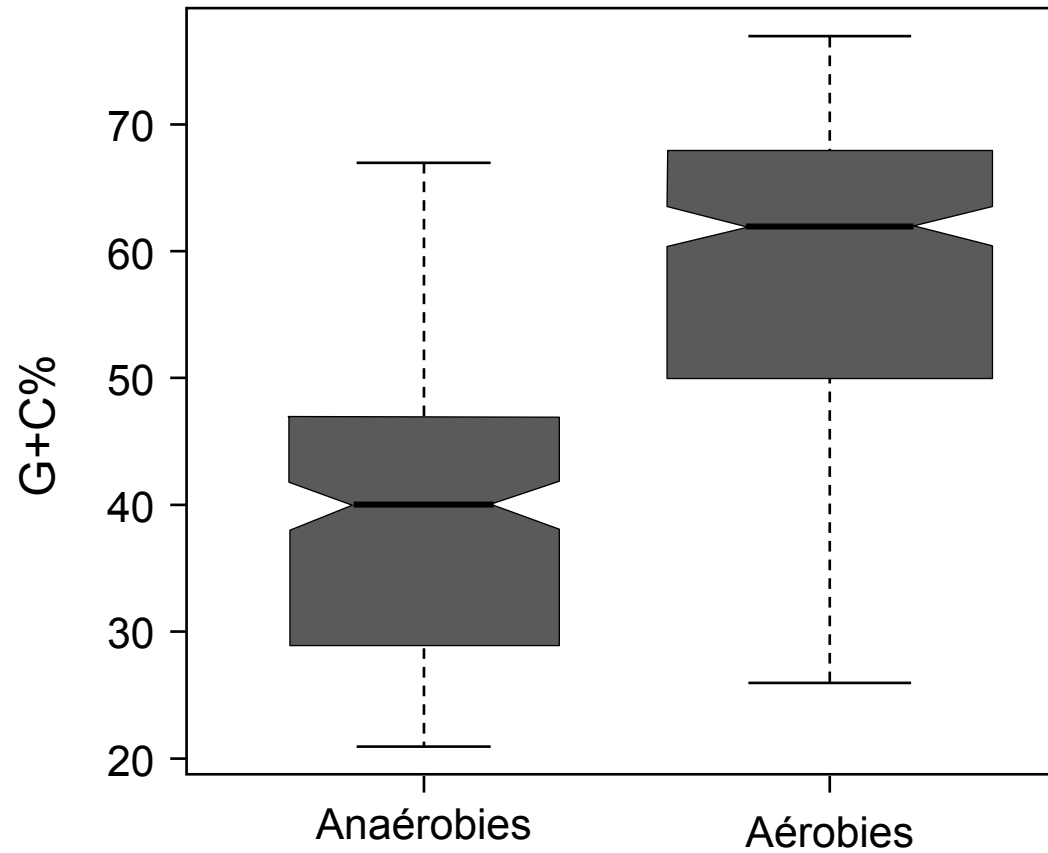
Structure de l'ARNr 16S



Le biais hyperthermophile

- La structure secondaire des ARNr est indispensable à leur fonction.
- Les régions appariées sont d'autant plus stables qu'elles sont riches en G+C :
 - Trois liaisons H au lieu de deux.
 - Enrichissement en G+C chez tous les organismes hyperthermophiles :
 - Existence d'un biais de composition en bases uniquement au niveau des structures de l'ARNr.

Relation G+C% / aérobiose

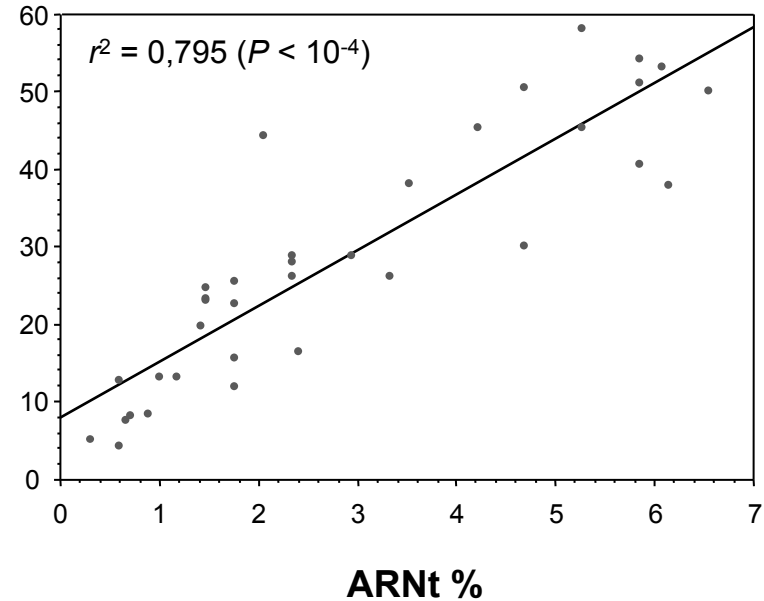
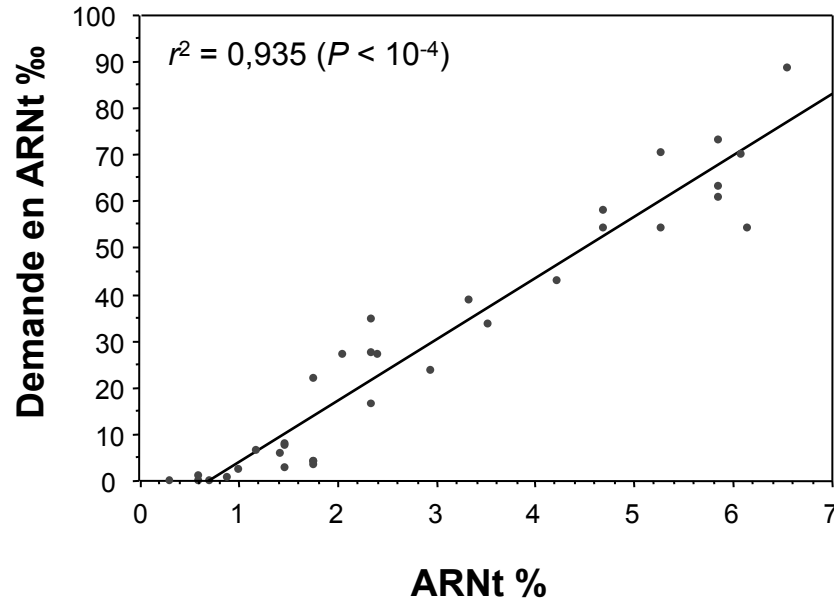


Fréquences des ARNt

AA	Codon	Anti-codon	ARNt %	AA	Codon	Anti-codon	ARNt %
Arg	CGH	ICG	5.25	Val	GUY	GAC	2.33
	CGG	idem	0.64		GUD	UAC	6.12
	AGR	idem	0.70	Lys	AAR	UUU	5.83
Leu	CUY	GAG	1.75	Asn	AAY	QUU	3.50
	CUA	idem	0.58	Gln	CAA	UUG	1.75
	CUG	CAG	5.83		CAG	CUG	2.33
	UUR	AAA	1.46	His	CAY	QUG	2.33
Ser	UCY	GGA	2.39	Glu	GAR	UUC	5.25
	UCD	UGA	1.46	Asp	GAY	QUC	4.67
	AGY	GCU	1.46	Tyr	UAY	QUA	2.92
Thr	ACY	GGU	4.67	Cys	UGY	GCA	1.17
	ACR	UGU	1.40	Phe	UUY	GAA	2.04
Pro	CCY	GGG	0.99	Ile	AUY	GAU	5.83
	CCR	UGG	3.32		AUA	NAU	0.29
Ala	GCY	GGC	4.20	Met	AUG	CAU	1.75
	GCD	UGC	6.06		Trp	UGG	CCA
Gly	GGY	GCC	6.53				
	GGA	UCC	0.88				
	GGG	CCC	0.58				

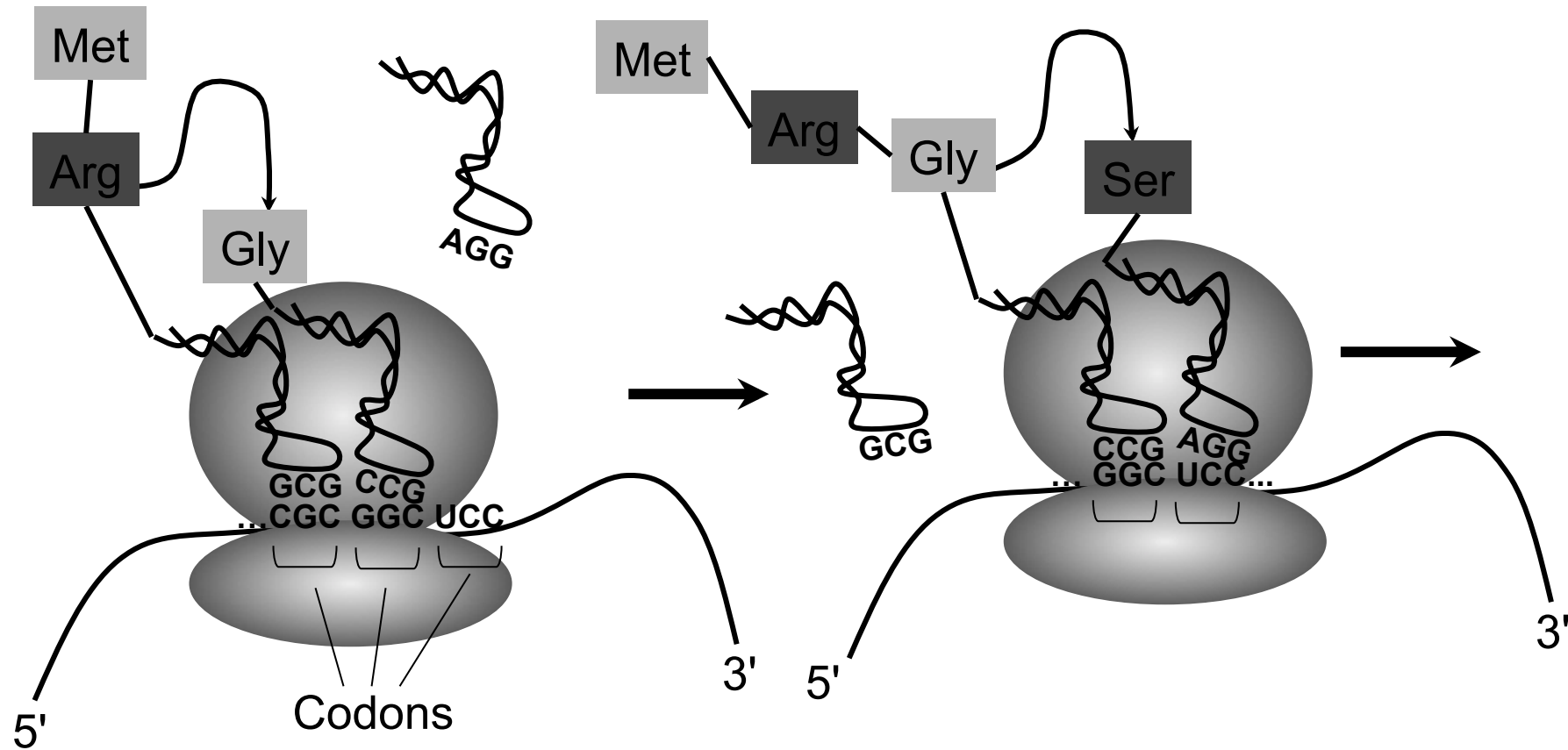
Q = Quosine, I = Inosine, R = {A, G}, Y = {C, U}, D = {A, G, U}, H = {A, C, U}

Demande en ARNt

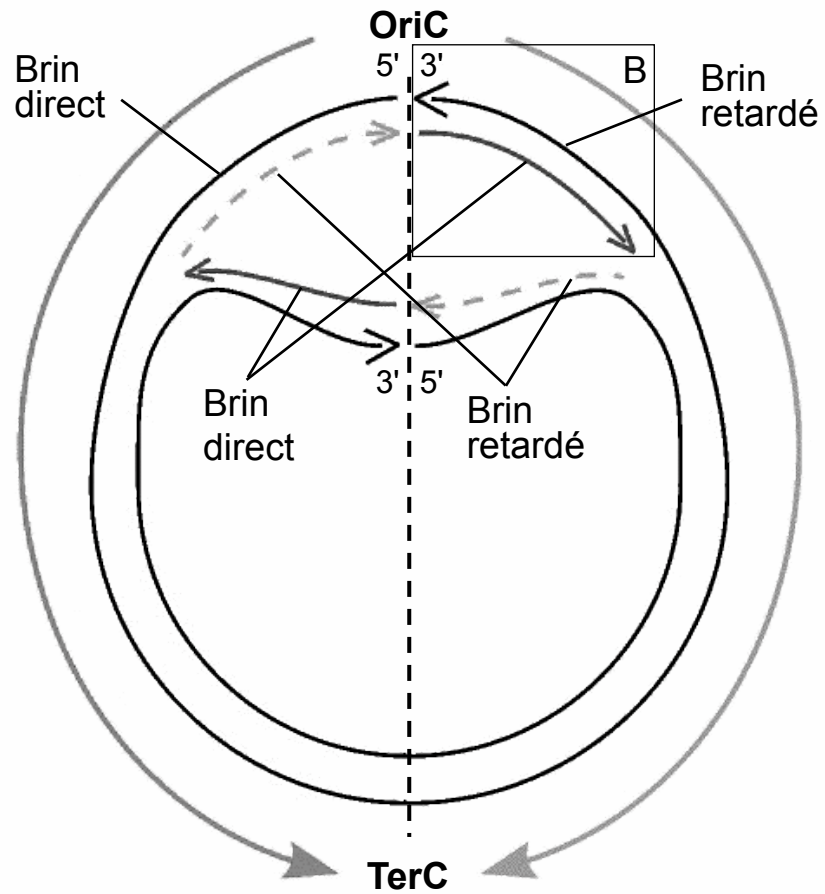


- La demande au niveau des gènes et la fréquence relative en ARNt sont corrélées positivement :
 - Cette corrélation est d'autant plus forte que les gènes sont hautement exprimés.

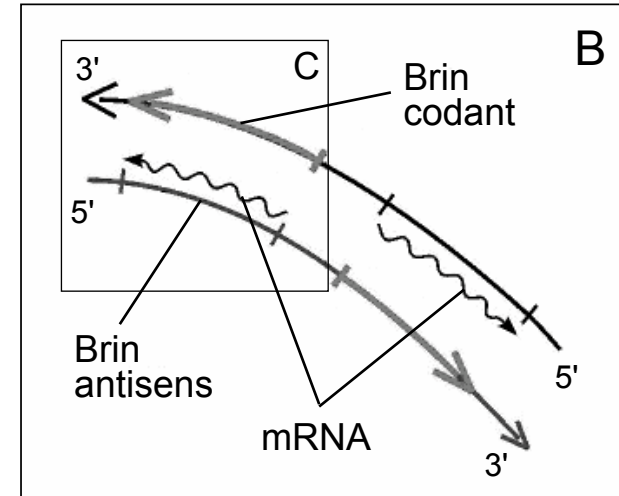
Sélection traductionnelle



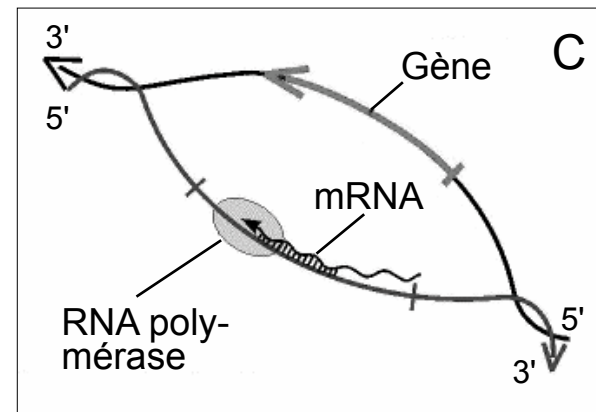
Réplication et transcription



A

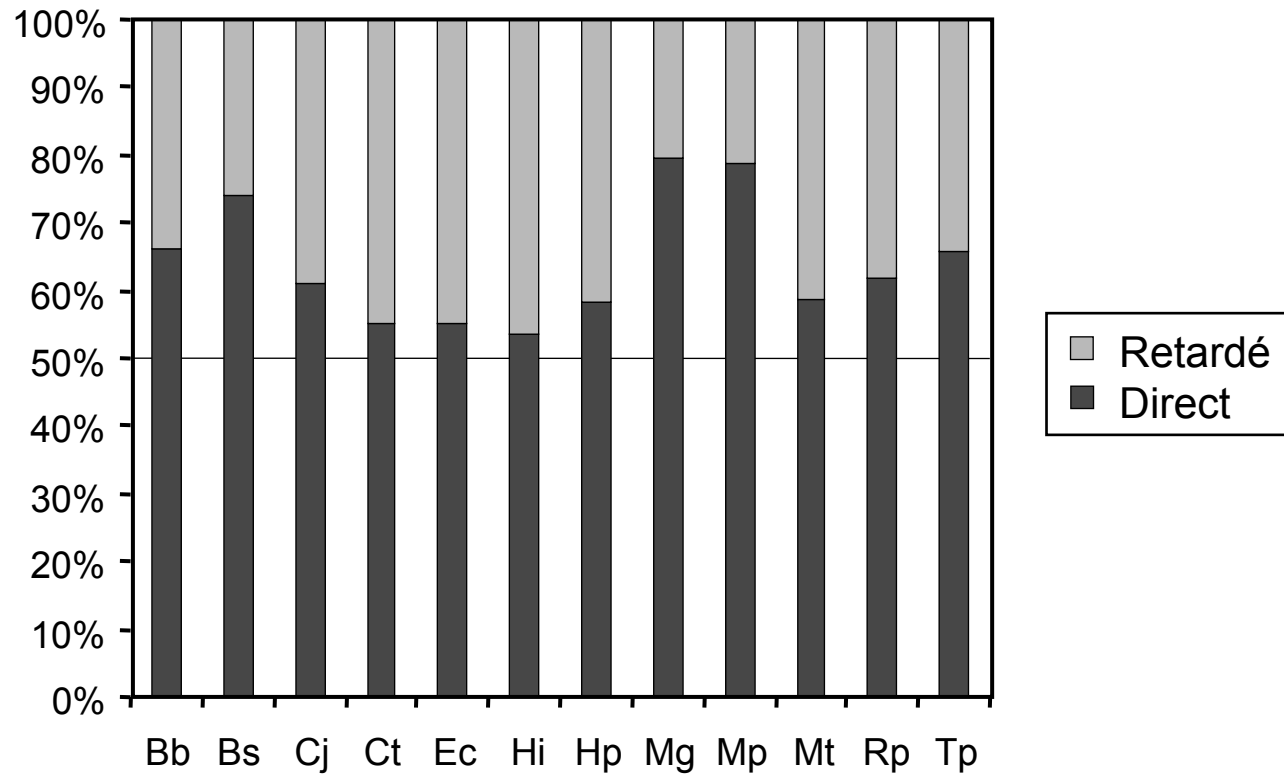


B



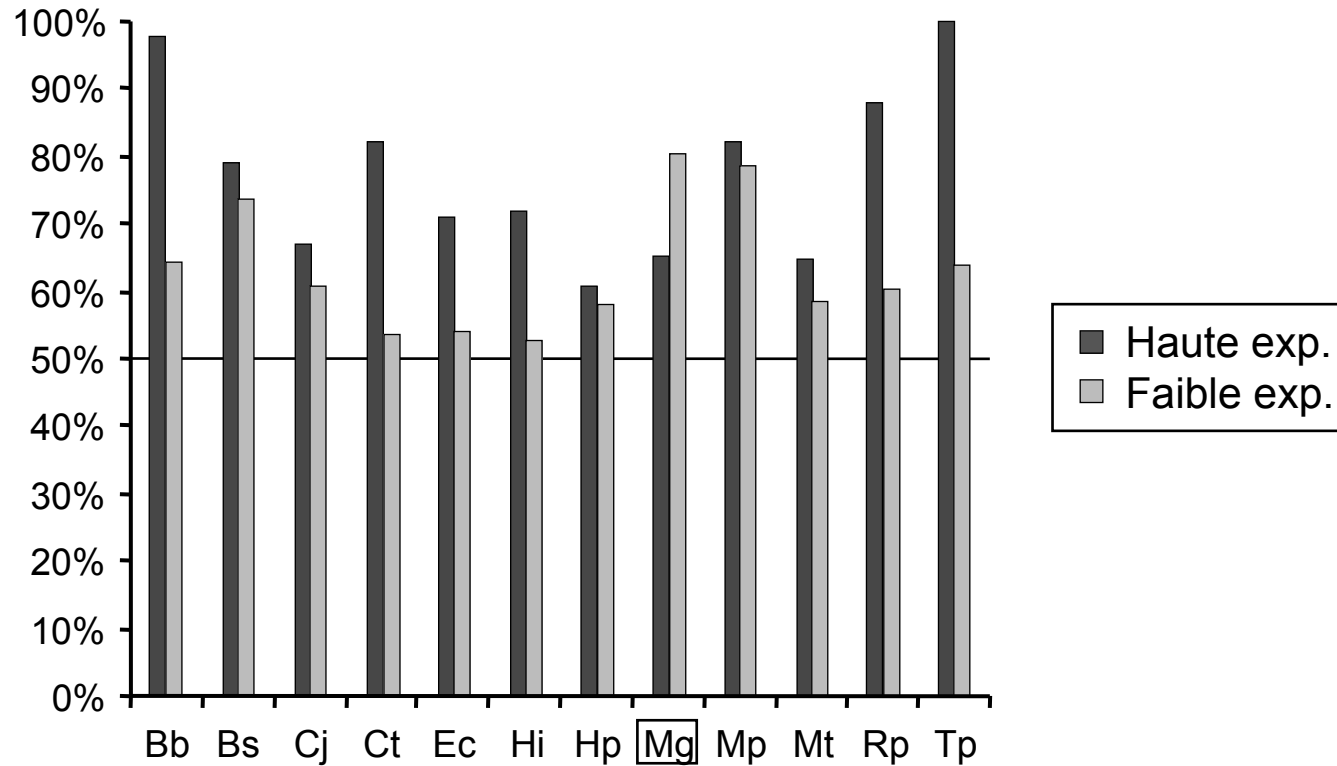
C

Distribution des gènes



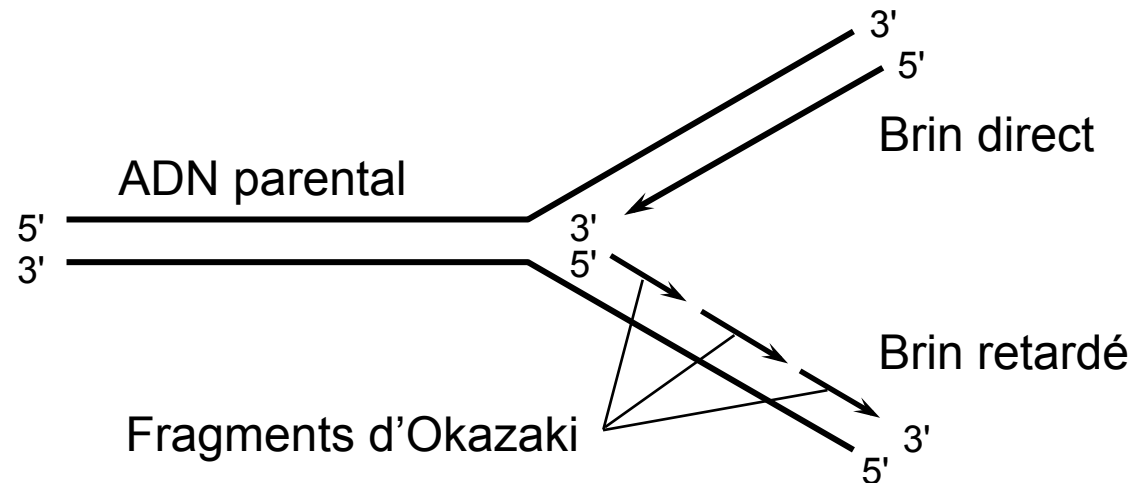
Les gènes sont plus souvent
localisés sur le brin direct

Localisation et expression



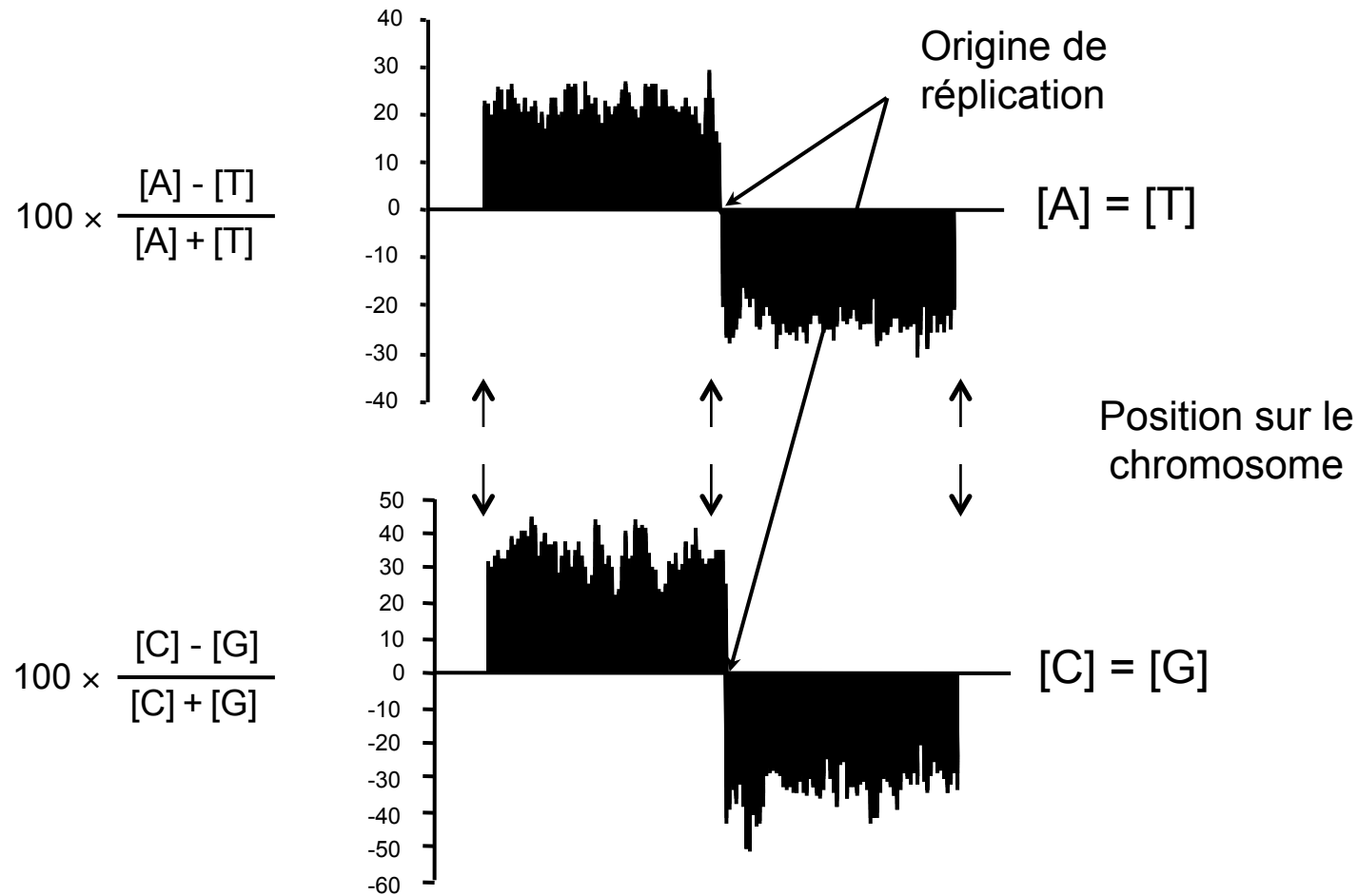
Le biais est plus fort pour
les gènes hautement exprimés

Différence entre les deux brins



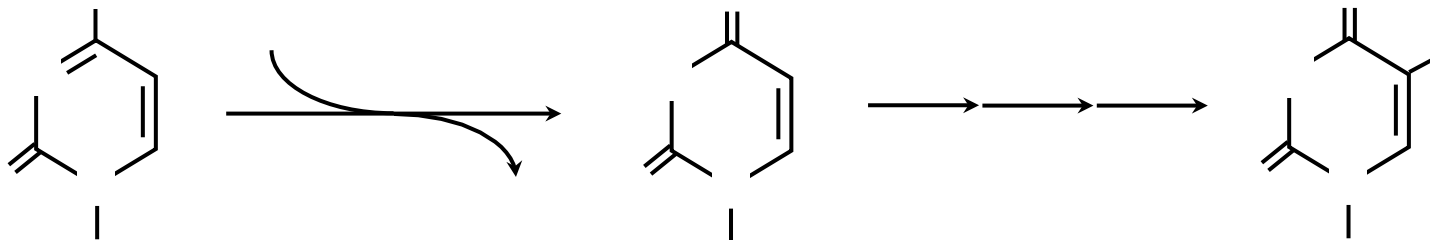
- Le brin direct et le brin retardé sont répliqués de façon différente :
 - Il existe une asymétrie de composition en base entre les deux brins chez la plupart des bactéries.
 - Cette asymétrie est très marquée chez certaines espèces.

Asymétrie chez *B. burgdorferi*



Origine du biais

- Mutations spontanées C → U par déamination des cytosines.
- Phénomène plus fréquent dans l'ADN simple brin (accessibilité au solvant) :
 - Le brin direct se retrouve fréquemment non apparié du fait du mécanisme de réplication :
 - Enrichissement concomittant en Thymines après réparation.



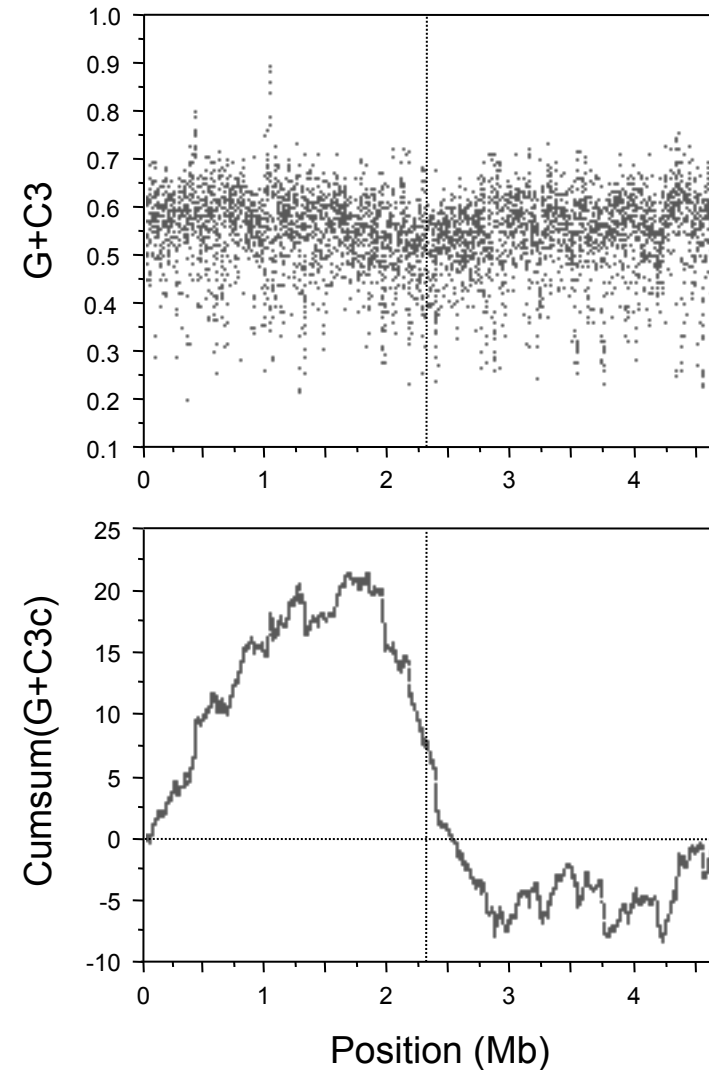
Contenu en G+C3

- Contenu en G+C3 d'un gène :
 - Pourcentage en bases G+C en position III des codons :
 - Valeur brute (G+C3).
 - Centrage par la moyenne (G+C3c).
 - La position III est le plus souvent conservatrice en cas de mutation ponctuelle :
 - Pas de changement de l'acide aminé codé dans 62,5 % des cas (composition uniforme).
 - La plasticité joue essentiellement sur cette position.

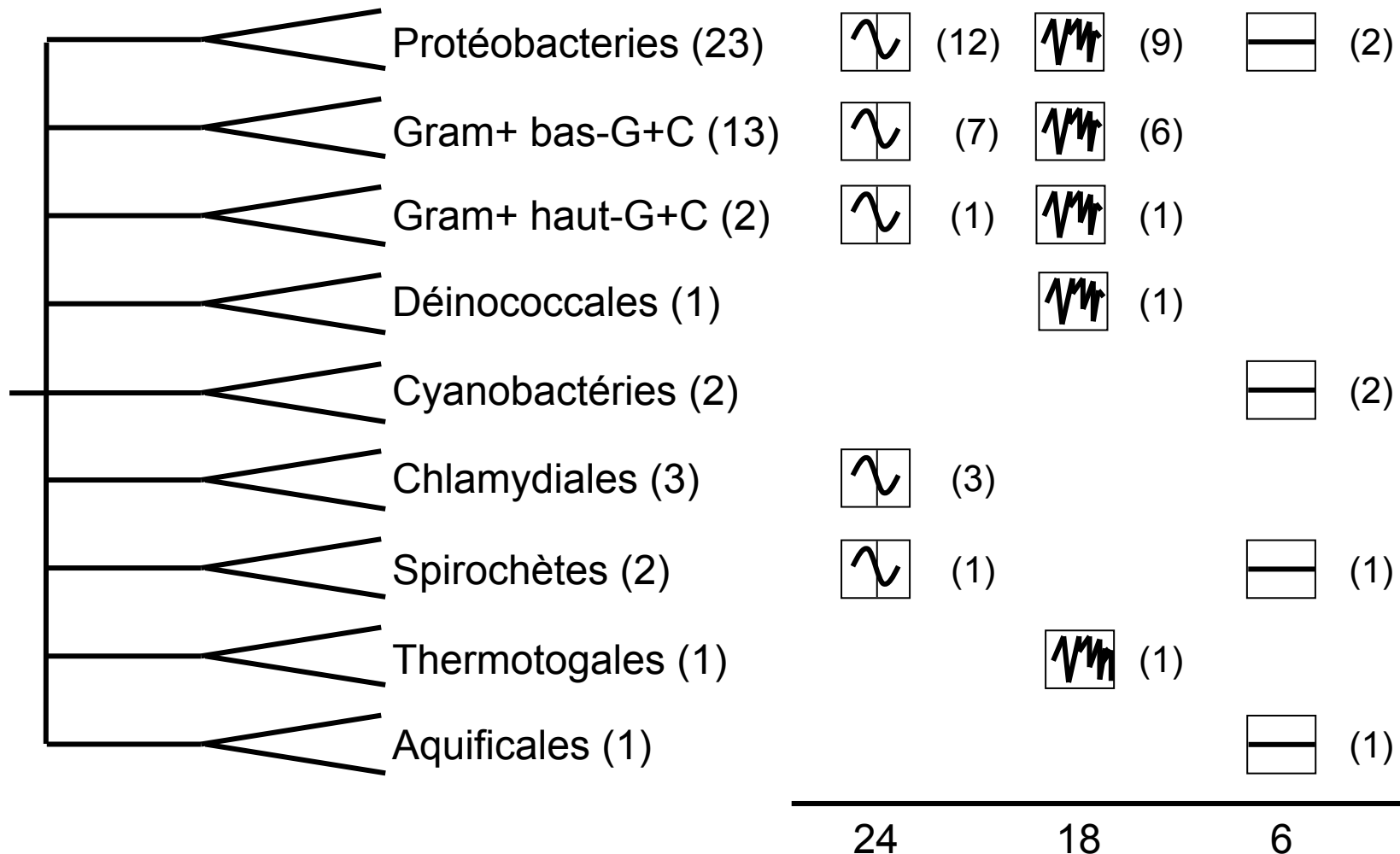
Effet de la localisation

La variation des valeurs « brutes » du G+C3 le long du chromosome d'*E. coli* n'est pas significative

L'utilisation des valeurs centrées / cumulées (G+C3c) révèle une structuration du génome

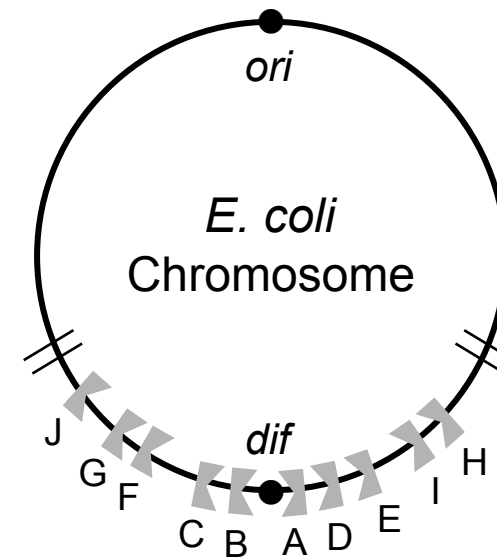


Répartition phylogénétique



Le terminus d'*E. coli*

- La région du terminus contient dix sites *ter* (A-J):
 - Combinaison à la protéine Tus :
 - Inhibition de l'action des hélicases de manière polaire.
 - Les deux fourches de réplication se rejoignent au niveau du site *dif*.
 - Résolution des dimères.



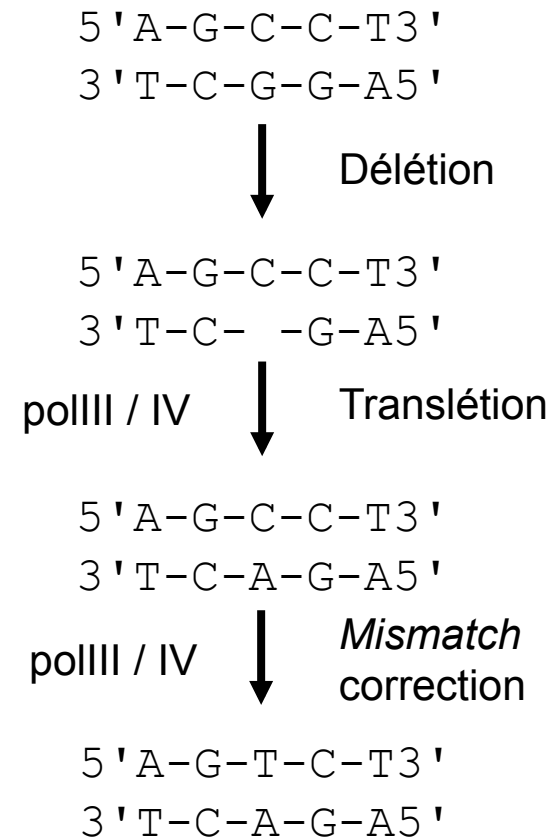
Localisation et polarité des dix sites *ter* d'*E. coli*

Réparation des lésions

- Quand le complexe atteint une lésion, la réplication s'arrête :
 - Recul de la fourche.
 - Formation d'une jonction de Holliday sous l'action des hélicases RecG et PriA :
 - Réparation par recombinaison homologue.
- Dans la région du terminus, les complexes *ter*/Tus inhibent l'action des hélicases !
 - Réparation par le mécanisme de translétion.

La translétion

- Réalisée par l'intermédiaire des polymérase SOS (polIII et polIV).
- Introduction systématique de dAMP (*A-rule*) aux sites abasiques.
- Enrichissement en A+T de régions n'utilisant pas la réparation par recombinaison :
 - Cas des séquences au voisinage du terminus de réplication.



Transferts horizontaux

- Se définissent comme étant le passage direct de l'ADN d'un génome à un autre.
- Semblent toucher l'ensemble des organismes vivants.
- La plupart des études ont été réalisées chez les bactéries :
 - Raisons historiques (Griffith, 1928).
 - Motivations d'ordre médical et économique :
 - Acquisition de pathogénicité et/ou de résistances.

Conditions nécessaires

- Proximité entre le donneur et l'accepteur.
- Stabilité de l'ADN dans l'environnement :
 - Utilisation d'un vecteur de transmission.
- Capture par l'hôte suivi d'une insertion.
- Maintient de l'ADN incorporé :
 - Avantage sélectif procuré.

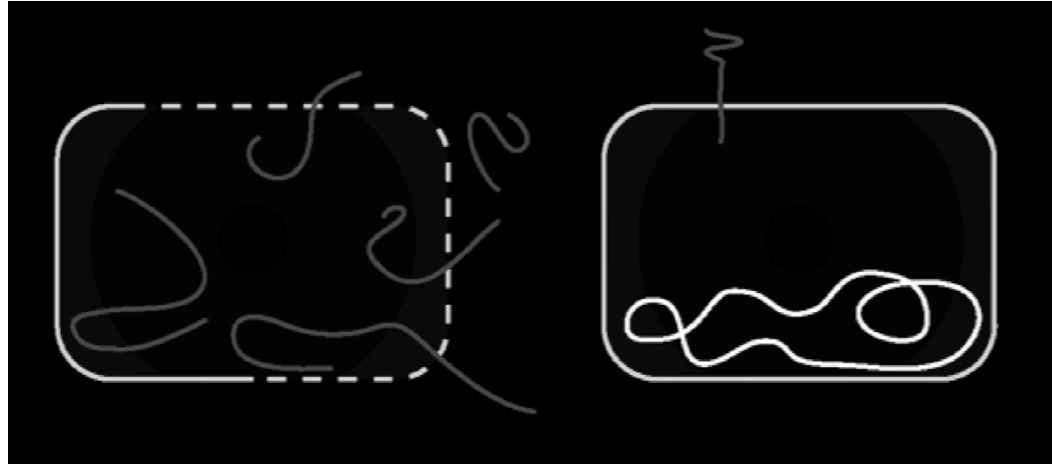
Protections / limitations

- Instabilité de l'ADN internalisé.
- Systèmes de restriction / modification.
- Incompatibilité compositionnelle (G+C, composition en codons).
- Pas de régions recombinantes suffisamment longues.

Cas connus

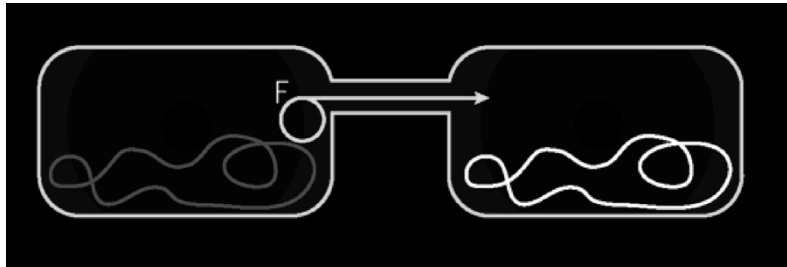
- Gènes plasmidiques de résistance aux antibiotiques et aux toxines.
- Séquences d'insertion (IS).
- Ilôts de pathogénicité.
- Plasmide Ti d'*Agrobacterium tumefaciens*.
- Phages et pseudo-phages.
- Transfert depuis des génomes d'organites (mitochondrie, chloroplaste) vers le noyau.

La transformation

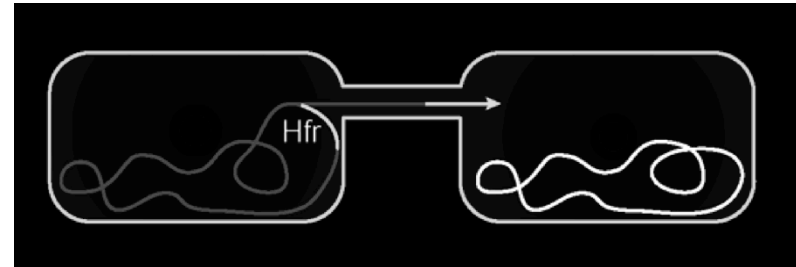


- Capture par la cellule de fragments d'ADN nu présents dans le milieu.
- La transformation naturelle est restreinte à certaines espèces bactériennes « compétentes » :
 - Présence des gènes *com* (≈ 10).

La conjugaison



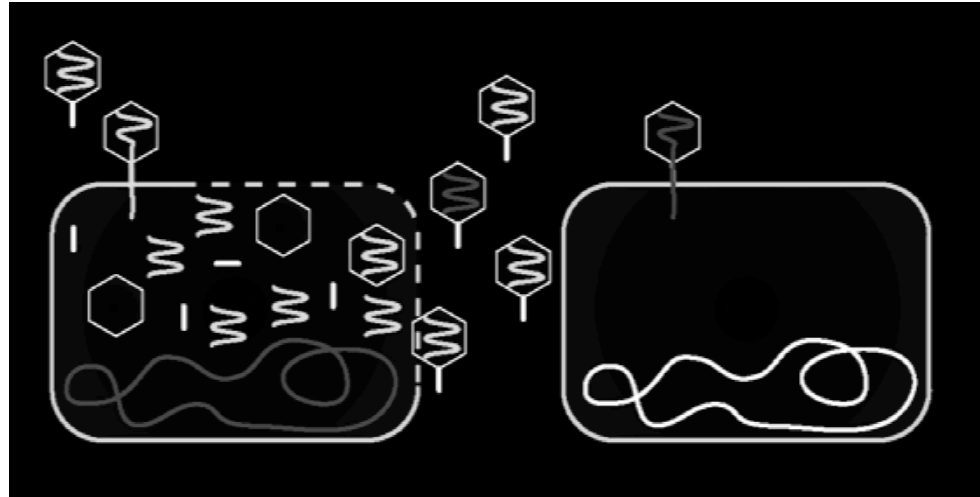
Évènement fréquent



Évènement rare

- Transfert par l'intermédiaire d'un plasmide conjugatif (*e.g.*, le plasmide *F* chez *E. coli*).
- La conjugaison serait le mécanisme le plus courant d'échange d'ADN entre organismes distants.

La transduction

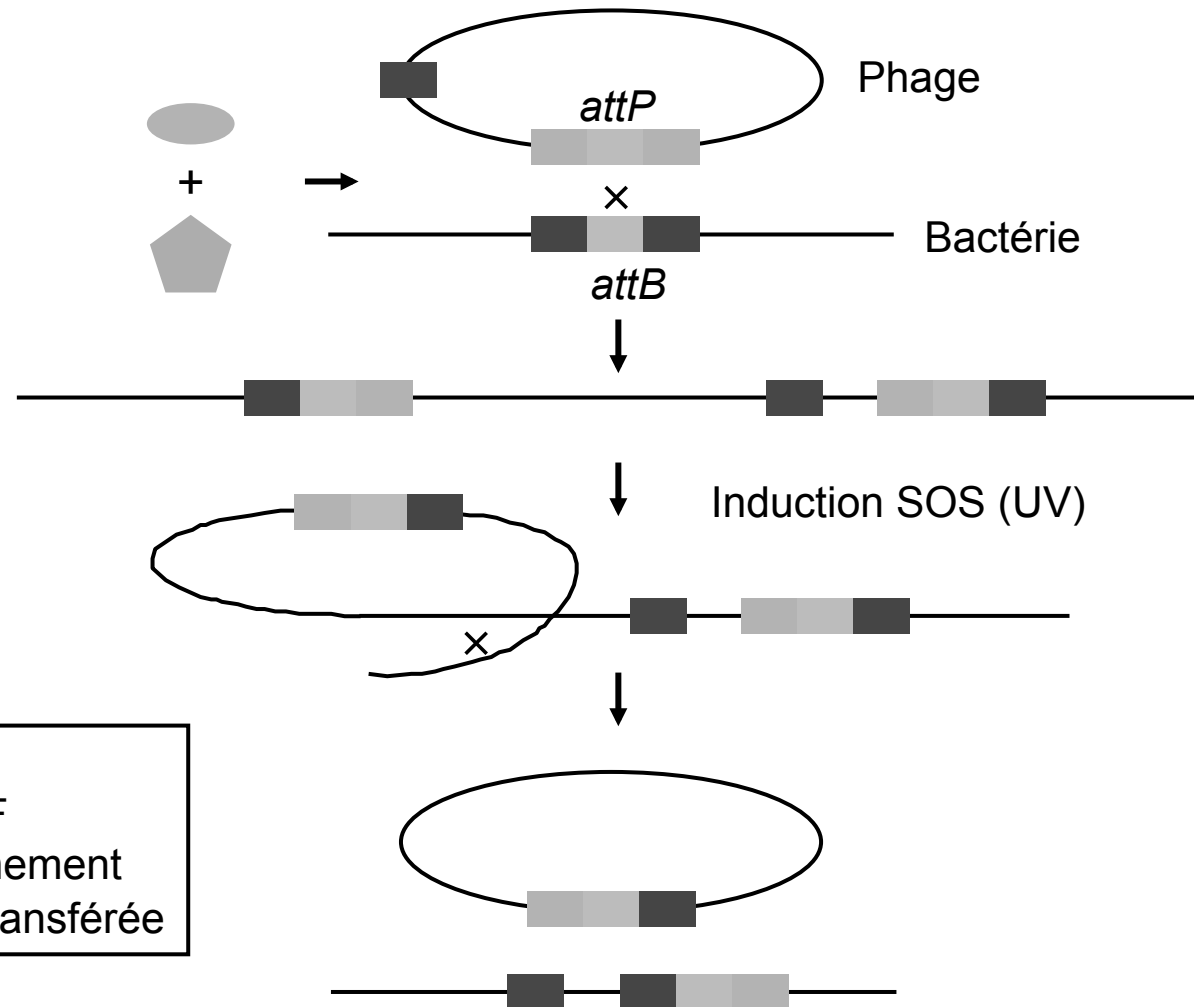


- Transfert d'ADN par l'intermédiaire de bactériophages.
- Implique généralement des bactéries appartenant à des espèces proches (spécificité d'hôte) :
 - Certains phages (*e.g.*, μ) possèdent un large spectre.

Intégration des séquences

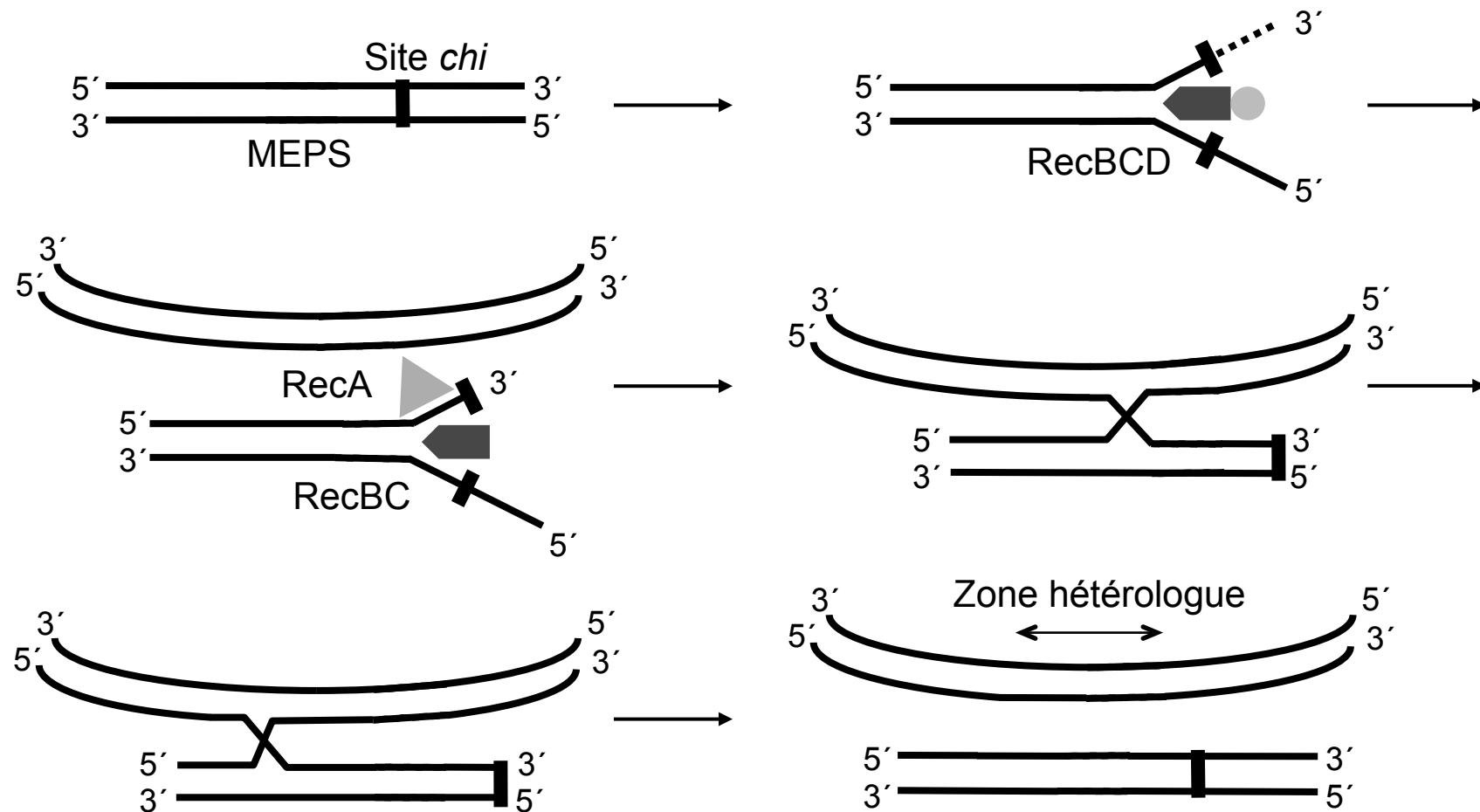
- L'intégration des fragments exogènes est rendue possible par différents mécanismes :
 - Recombinaison site-spécifique.
 - Recombinaison générale homologue :
 - Recombinaison réciproque.
 - Recombinaison non réciproque.
- Contribution à la divergence des génomes :
 - Chez *E. coli*, effet 50 fois plus important que la mutagenèse.

Recombinaison site-spécifique



Protéine Int
Protéine IHF
Site d'attachement
Séquence transférée

Recombinaison réciproque



Les zones MEPS

- Zones homologues minimales nécessaire à l'amorçage de la recombinaison (*Minimum Efficient Processing Segment*) :
 - Doivent présenter une forte similarité entre l'ADN endogène et exogène.
 - Longueur dépendante des mécanismes de réparation des mésappariements existant chez l'hôte :
 - *E. coli* : 23-27 pb.
 - *B. subtilis* : ≈ 70 pb.

Rôle des sites *chi*

- Séquences d'ADN courtes (10 pb) stimulant la recombinaison.
- Permettent le démarrage de la recherche des zones MEPS par la protéine RecA.
- Une région comprenant une zone MEPS et un site *chi* peut recombiner.
- Leur présence peut permettre de confirmer l'existence d'un transfert horizontal.
- Connus dans un petit nombre d'espèces.

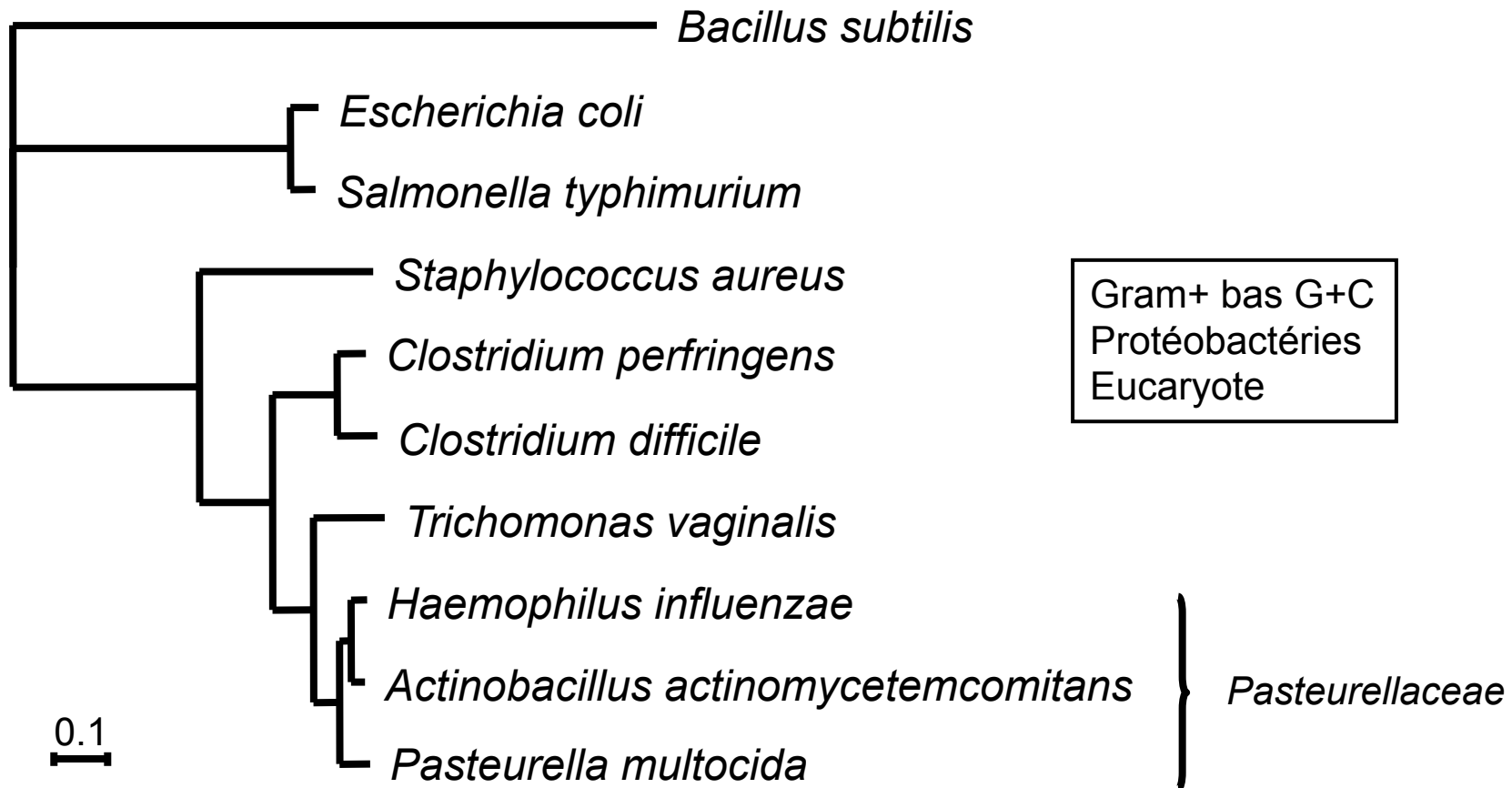
Transferts et pathogénicité

- Transposons :
 - Gènes de l'entérotoxine ST d'*E. coli*.
- Prophages :
 - Toxines de l'*E. coli* entéro-hémorragique.
 - Gènes de la toxine diphtérique ou du choléra.
 - Toxines botuliniques.
- Plasmides :
 - *Shigella*, *Salmonella*, *Yersinia*.
- Ilôts de pathogénicité :
 - *E. coli* uropathogène, *S. typhimurium*, *Yersinia*, *H. pylori*, *V. cholerae*.

Méthodes de détection

- Plus grande similitude des séquences avec un organisme distant qu'avec un organisme proche du point de vue taxonomique.
- Composition en nucléotides des gènes ou des génomes :
 - Une région transférée présentera des valeurs significativement différentes de celle des régions « natives » :
 - Nécessité de prendre en compte les facteurs non liés aux transferts intervenant dans la plasticité des génomes.

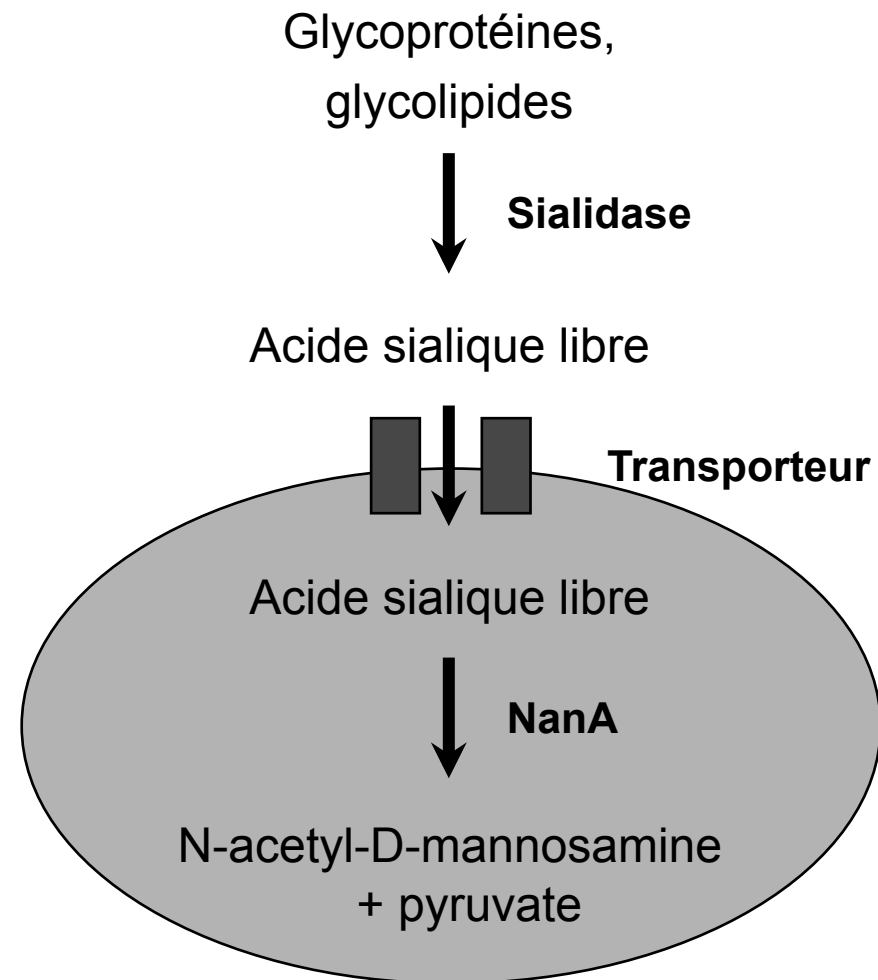
Transfert bactérie-eucaryote



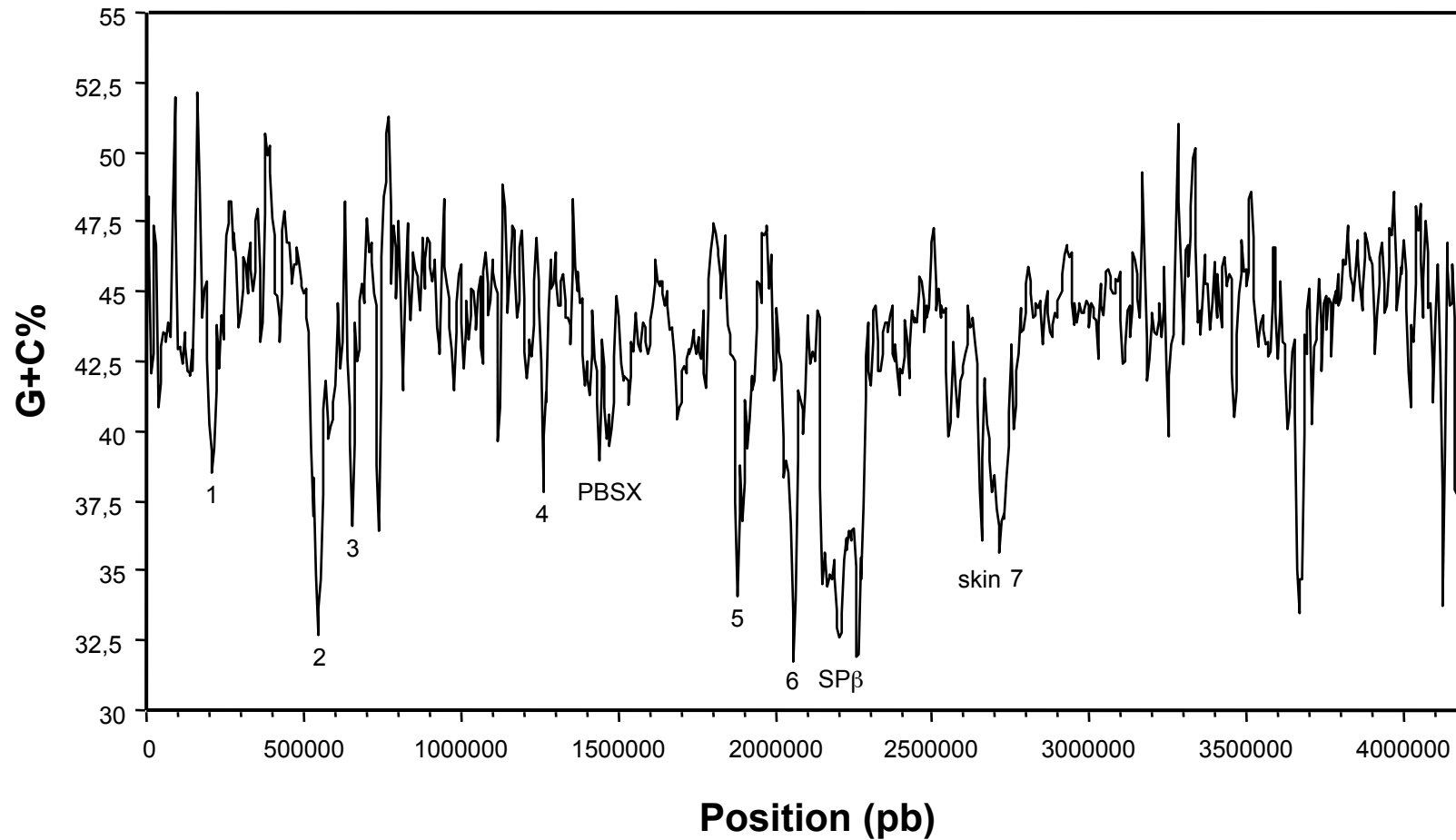
Phylogénie de la *N*-acetylneuraminase lyase (NanA)

NanA et pathogénicité ?

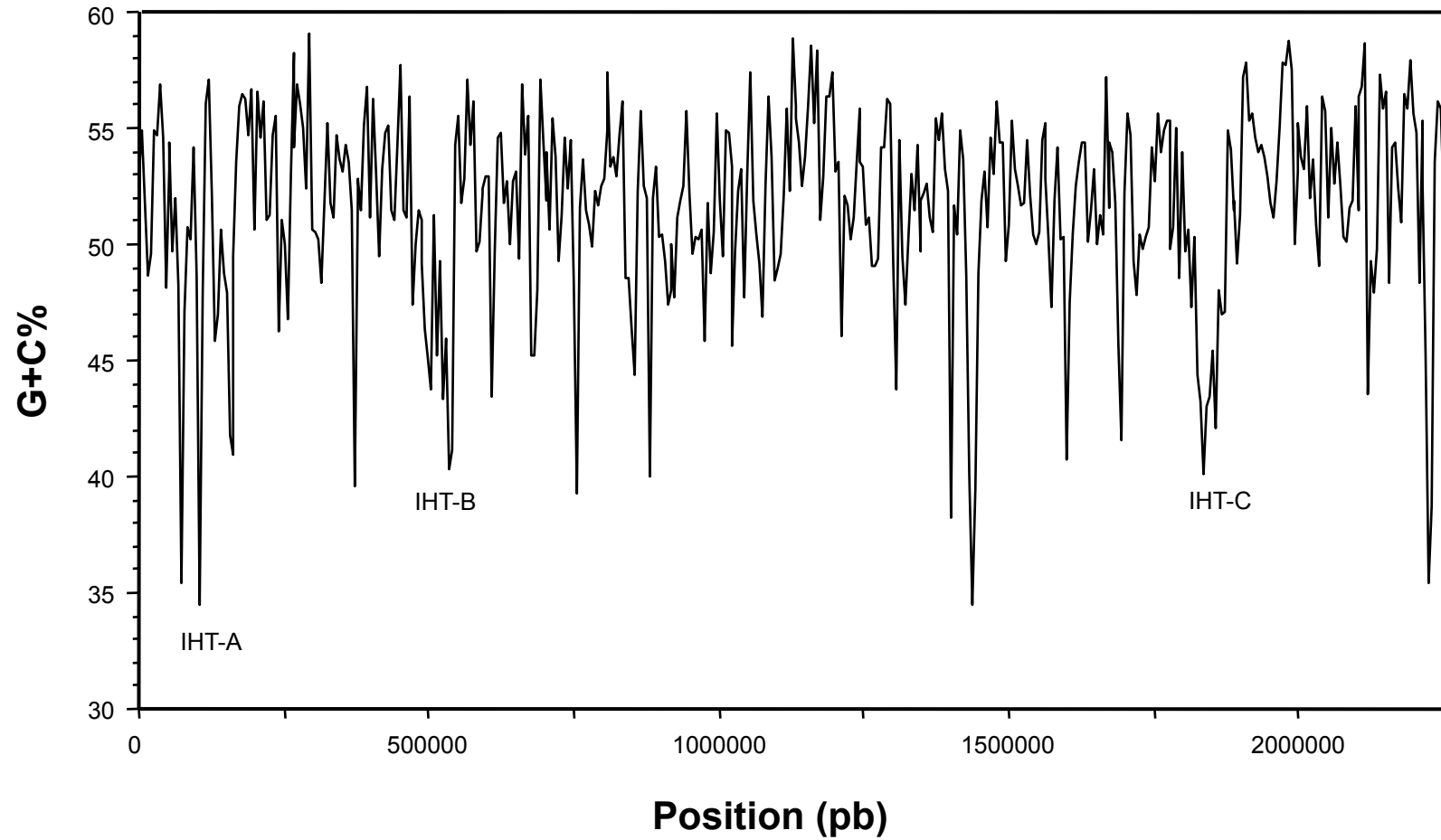
- Chez les Pasteurellacées :
 - Parasitage des muqueuses des animaux :
 - Proximité physique entre *T. vaginalis* et les bactéries considérées.
- *T. vaginalis* connue pour sa capacité à phagocyter les bactéries.
- Transfert responsable de l'acquisition de la pathogénicité ?



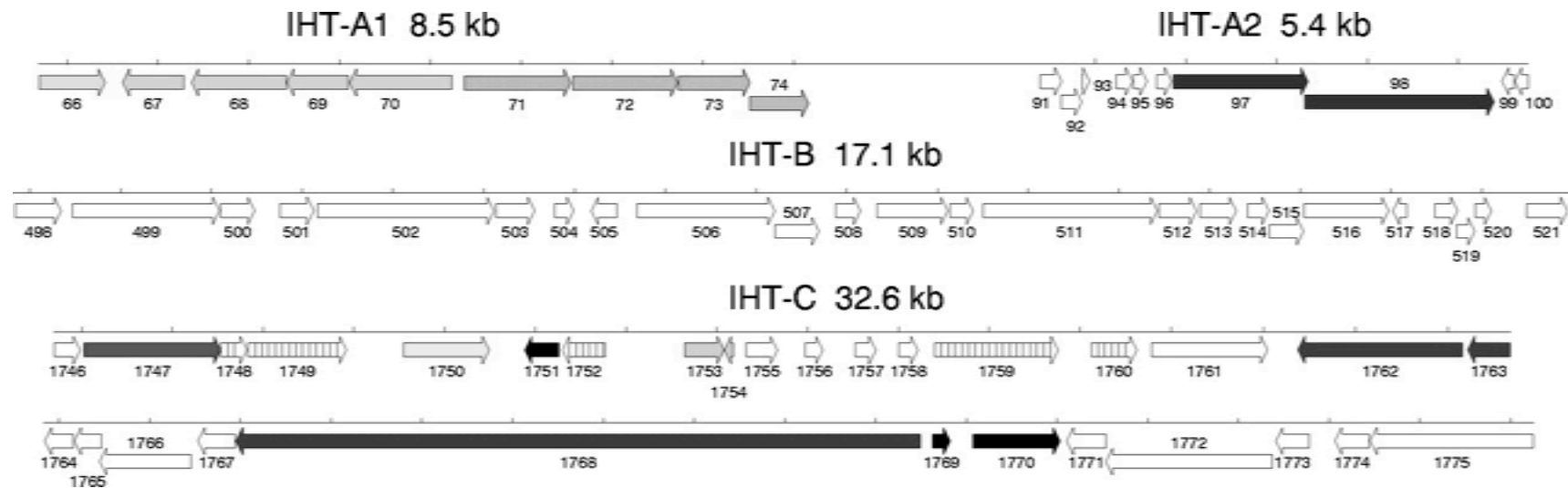
Profil en G+C chez *B. subtilis*



Profil chez *N. meningitidis*



Les îlots de *N. meningitidis*



IHT-A1

66 : adénine rRNA méthylase, 67-70 : protéines de biosynthèse de la capsule, 71-74 : protéines d'exportation de la capsule

IHT-A2

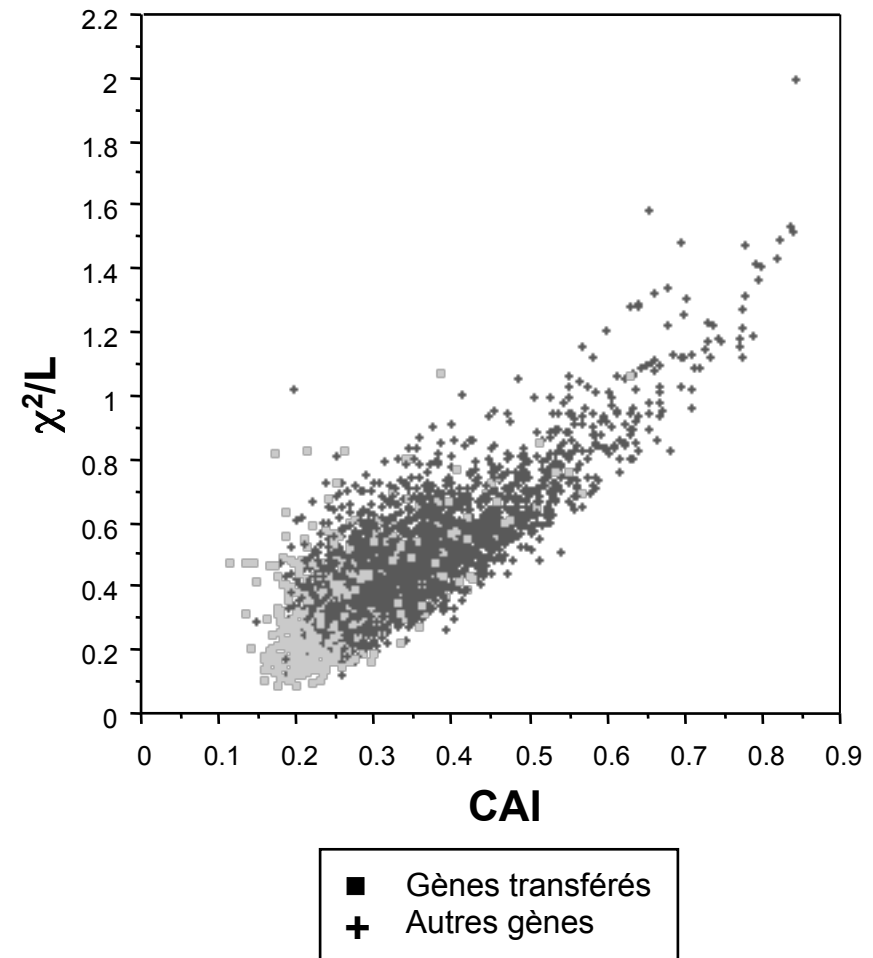
97 : protéine de sécrétion, 98 : transporteur ABC

IHT-C

1747 : protéine TspB, 1750 : PivNM-2, 1751, 1769-70 : transposases, 1753-54 : protéines de phages, 1762-63, 1768 : toxines ou assimilées

Lawrence et Ochman (1998)

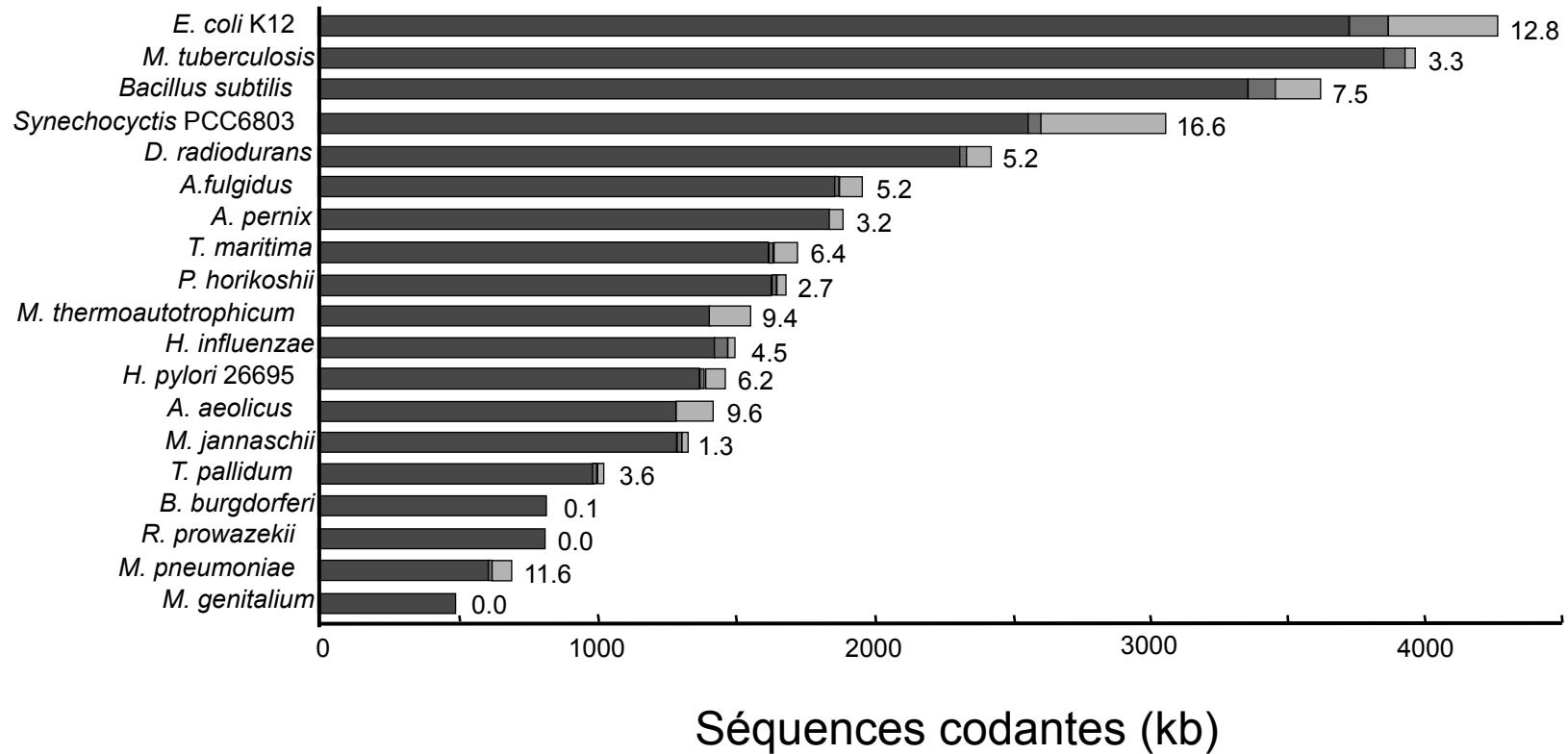
- Utilisation d'indices compositionnels (en particulier χ^2).
- Proximité de structures impliqués dans les transferts :
 - Éléments transposables (séquences IS).
 - Régions recombinantes (sites *chi*).



Résultats chez *E. coli*

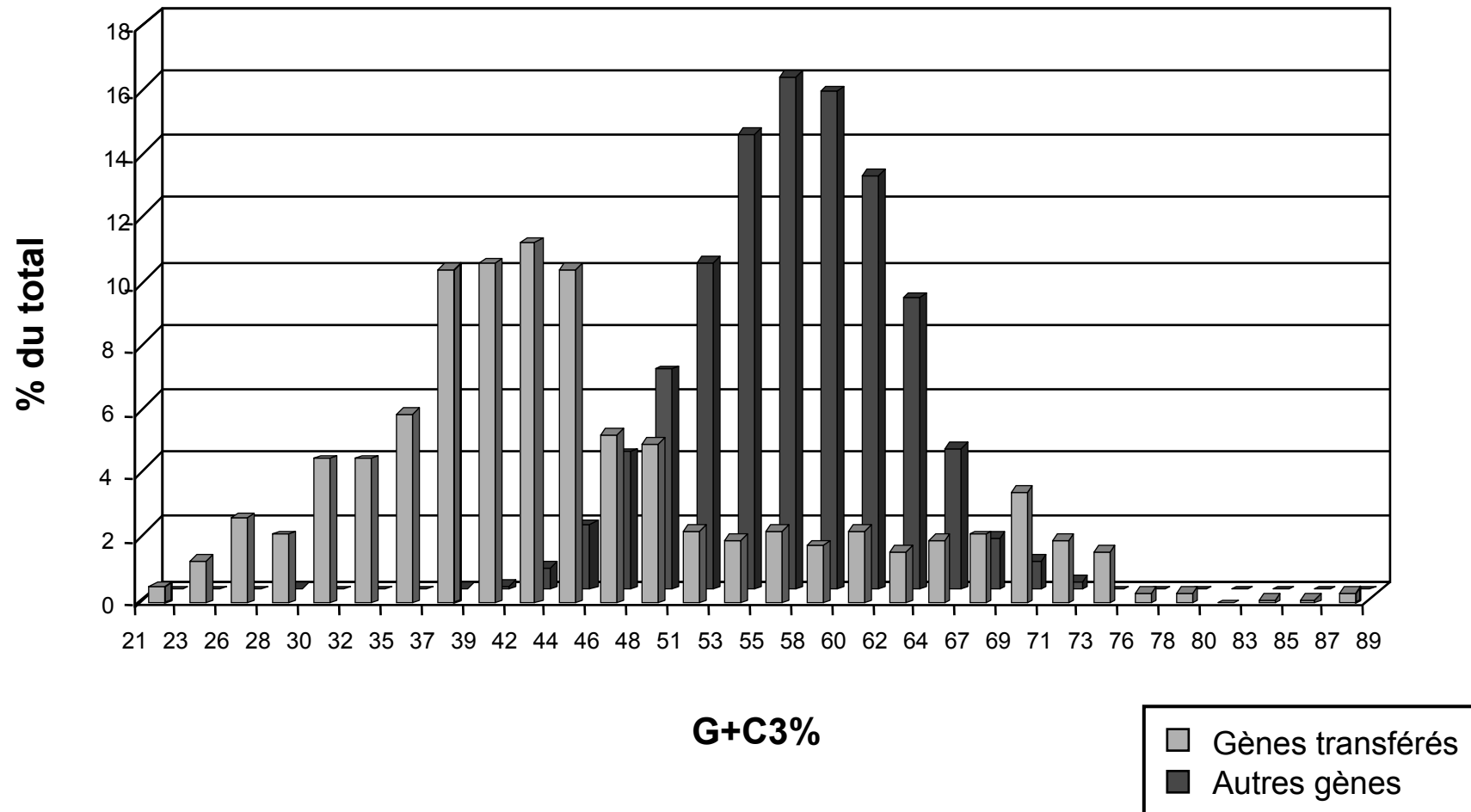
- 755 gènes sur 4286 (17,6 %) sont prédits comme transférés chez *E. coli*.
- 67 % des séquences IS du génome sont associées à ces gènes :
 - Ces séquences IS constituent fréquemment des frontières entre zones endogènes et exogènes.
- 36 % des gènes détectés se situent au voisinage du terminus de réplication (24 % du chromosome).

Autres espèces

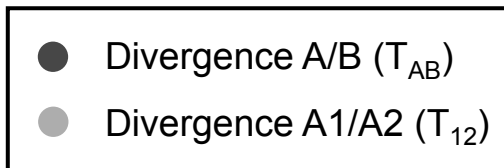
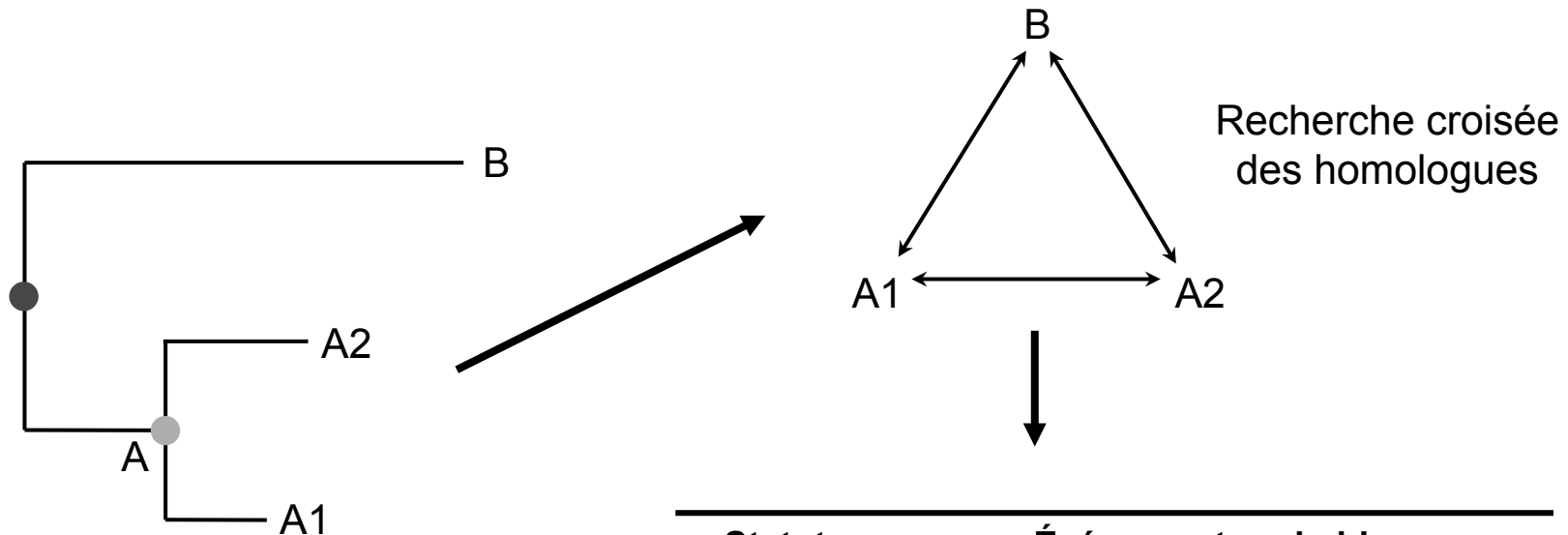


- Gènes « natifs » présents dans le génome
- Gènes transférés associés à des éléments mobiles
- Gènes transférés non associés à des éléments mobiles

G+C3 des gènes

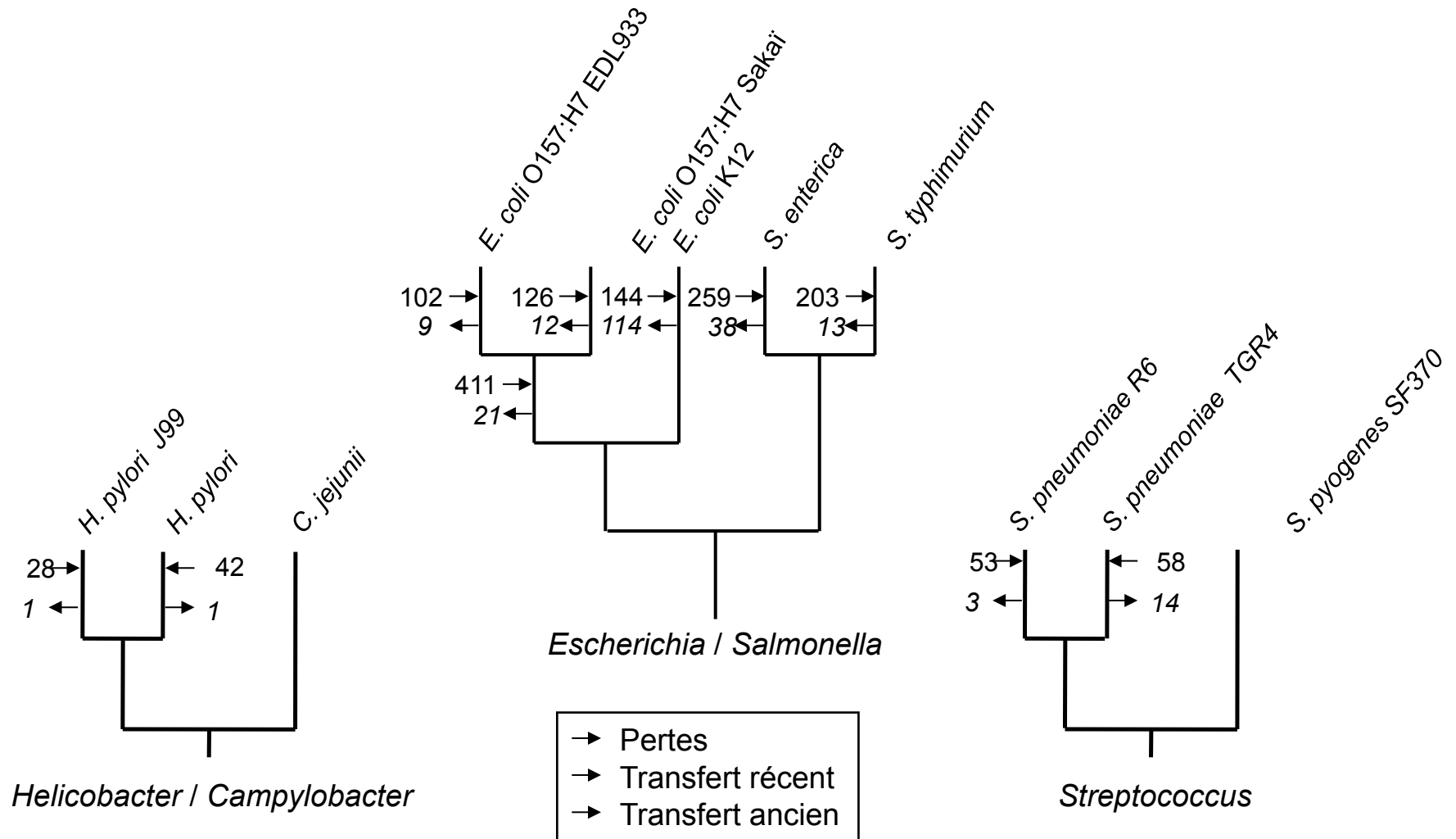


Approche comparative



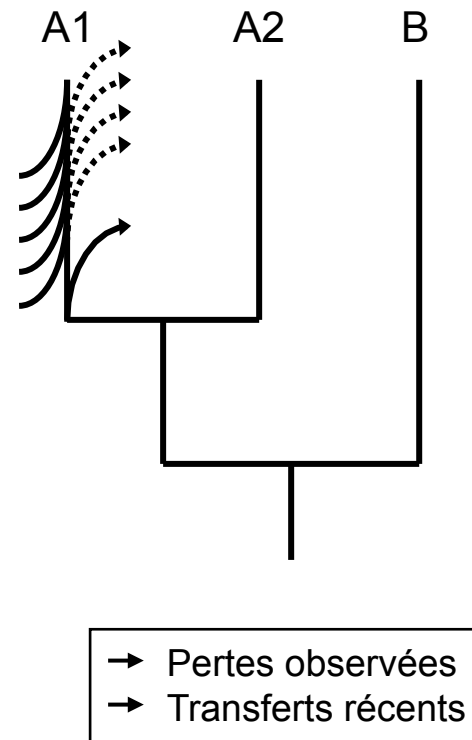
Statut	Événement probable
+A1 -A2 -B	Importation dans A1 après T_{12}
+A1 +A2 -B	Importation dans A après T_{AB} ou délétion dans B
+A1 -A2 +B	Délétion dans A2 après T_{12}
+A1 +A2 +B	Aucun événement
-A1 +A2 +B	Délétion dans A1 après T_{12}
-A1 +A2 -B	Importation dans A2 après T_{12}
-A1 -A2 +B	Délétion dans A après T_{AB}

Analyses effectuées

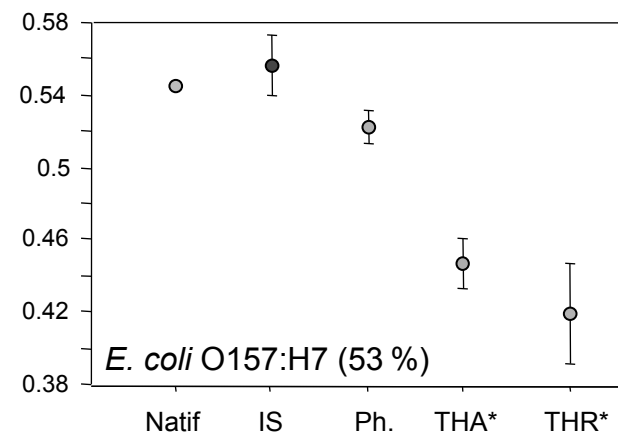
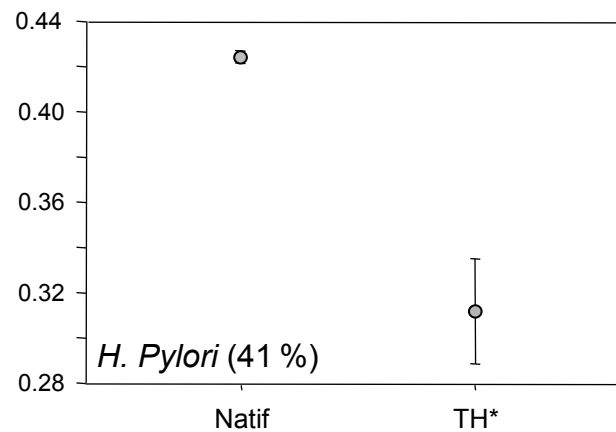
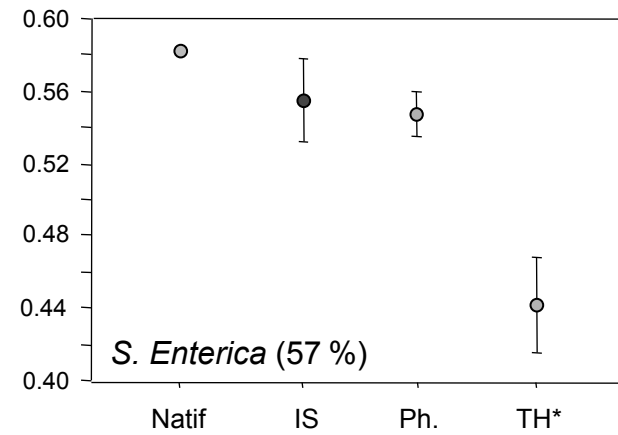
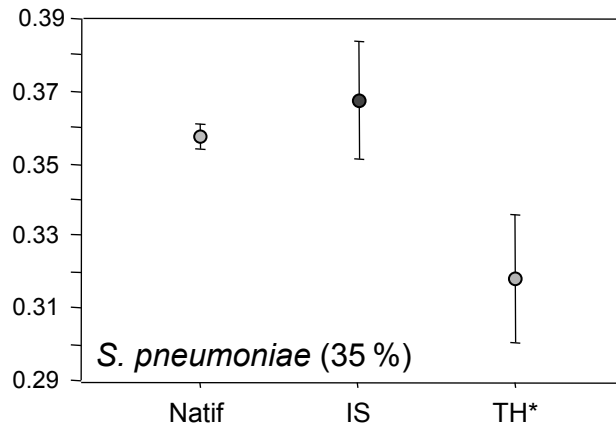


Gains et pertes

- Nb. d'acquisitions systématiquement supérieur au nb. de pertes :
 - Augmentation de la taille des génomes :
 - Cas des deux *E. coli* O157:H7 pathogènes.
 - Il existe un *turnover* rapide des gènes transférés :
 - Délétion peu de temps après transfert.



Contenu en G+C3



* Test de Mann-Whitney significatif à $P < 10^{-4}$

Hypothèses possibles

- Biais dans les séquences reconnues par les systèmes de restriction :
 - Celles-ci apparaissent comme étant globalement riches en G+C (70 % dans REBASE).
- Biais au niveau des vecteurs impliqués dans les transferts (phages et éléments IS) ?
- Les transferts horizontaux, un sous-produit des besoins en Adénine des bactéries ?

Compétition métabolique

- Parmi les ribonucléosides triphosphates, l'ATP est le plus abondant :
 - Au cœur de toutes les voies métaboliques :
 - Besoins importants en Adénine.
- Les récepteurs membranaires codés par les gènes *com* ont une affinité plus grande pour l'ADN riche en Adénine :
 - La transformation, un mécanisme visant à récupérer l'Adénine dans le milieu ?