MASTER IN BIOLOGY

Unusual organization of the rRNA genes in the Rickettsia genus

ROMEDENNE , Magali

Award date:
1995

Awarding institution:
University of Namur

FACULTES UNIVERSITAIRES N.-D. DE LA PAIX
NAMUR
FACULTE DES SCIENCES

Unusual organization of the rRNA genes in the *Rickettsia* genus

Mémoire présenté pour l'obtention du grade
de Licencié en Sciences
biologiques

Romedenne Magali

Décembre 1995

Facultés Universitaires Notre-Dame de la Paix
FACULTE DES SCIENCES
Rue de Bruxelles 61 - 5000 NAMUR
Tél. 081/72.41.11 - Telex 59222 Facnam-b - Telefax 081/72.44.20

Unusual organization of the rRNA genes in the *Rickettsia* genus

ROMEDENNE Magali

Summary

Dans une première partie du travail, nous décrivons l'organisation des gènes codant pour le peptide déformylase (PDF) et méthyionyl-tARN$_f^{Met}$ formyltransférase (MTF) dans la bactérie *Rickettsia prowazeki*. L'analyse de la séquence de clones ainsi que l'analyse par PCR de l'ADN génomique grâce aux amorces, correspondant à chacun des deux gènes, confirme que les gènes *fms* et *fmt* sont étroitement liés dans le génome de *Rickettsia prowazeki*.

Dans un second temps, nous continuons de démontrer l'organisation inabituelle des gènes codant pour les ARN ribosomiaux déjà observée dans *Rickettsia prowazeki*. Nous montrons dans cette étude que *fmt* est aussi localisé en amont du gène rrl dans le génome du *R. sibirica*, *R. parkeri*, *R. rickettsii*, *R. amblyommii* tout comme dans l'agent ELB. Ces résultats suggèrent que le réarrangement génomique des gènes codant pour les ARN ribosomiaux précèdent la divergence des espèces TG (Typhus group) et des espèces SFG (Spotted Fever group).

Mémoire de licence en Sciences Biologiques
Décembre 1995
Promoteur : J. VANDENHAUTE
Co-promoteur : S. ANDERSSON

Nine months, two weeks and a few hours, I had just arrived in the snow of Sweden for a great adventure! Back in Belgium, I miss all these people that I met, but anyway I am happy to see again the rain of my native country and all these people who supported me in my "crazy" trip. Now, it is time to thank evrybody who encourage me during this work.

First I wish to express my gratitude to Professor Charles G. Kurland who welcomed me in his department and introduced me in the scientific world.

I am extremely grateful for my "swedish" supervisor Siv Andersson. Your advices, supports and patience were really helpful for the realisation of this repport. I thank you for the interresting subject.

I thank Professor J. Vandenhaute who accepted to be my supervisor.

I am grateful for the ERASMUS program which gives me the wonderful opportunity to discover another university, to meet a lot of different people, culture and languages. I specially thank Father Devos and Marita Wrigen who organised my stay in Uppsala.

In addition, I would like to thank all my friends from the Molecular Biology Department.
Ulrika, for her patience to share the same banch, desk, and PCR machine! I thank you for the precious advices and encouragement in difficult moments (N'oublie pas ton Français, ça soulage parfois!).

Asfar, for all the informations about science but also about Pakistan (food and culture!). Be carreful on the snow! I really thank you for your friendship.

Ali (for advices), Haley (for all the primers), Thomas and Morten (for helping me with the computers, and that is not easy!), Anne-Sophie (for your humour, joy and patience with ALF), Jan (for your marker), Kristina (for your discrete but friendly presence) and Surrinder (unfortunately, you arrived to late but soon enough for me to enjoy your conversation).

I thank also Vildan and Anders (for the turkish music, see you for the next movie), Pedro (be carreful with chocolate), Monika (I will think about you next time when I will by my yoghourt!), Tracy and Mark (for your friendship), David and Mattias (for the BC), Angie (to take over the organization of the BC with an american accent),....

But there is also evrybody from outside of the department who make my stay in Uppsala a great pleasure.
Pernilla, my great buddy (or bodygard!). I thank you for your presence at anytime (specially on sunday afternoon in Kalmar nation).....

I thank also all the swedish students in biology for the nice parties, cafés, folkmusic, Valborg,... And of course, I should not forget the innumerable exchange students from everywhere in the world.

Je remercie Natacha pour sa compagnie durant ces cinq années, pour son courage de me supporter au kot, aux TP et pendant ces très très longs blocus. Je te remercie tout particulièrement pour l'encouragement dans mon expédition suédoise.

Un tout grand merci à Xavier qui, en dernière minute, a eu la patience de lire et corriger ce travail.

Evidemment, je ne peux oublier de remercier Ialex pour ses nombreux e-mails tout au long de ces neuf mois. Ils m'ont été réconfort, soutien et humour belge (et ça fait du bien!). Merci aussi pour l'aide pour l'achèvement de ce mémoire.

Merci beaucoup aux membres du laboratoire de biochimie qui m'ont accueillie ces derniers jours. Je remercie surtout Marie qui, il y a plus d'un an, m'a décidée de partir à Uppsala. Ton avis a joué un grand rôle dans cette expérience innoubliable. Merci à Christine que je découvre bien tard.

Je n'oublie pas tous les bios et tous les autres.

Il y a bien sûr Sophie que je remercie pour tout, mais je ne m'étale pas, ça risquerait d'être trop long.

Ce travail ne pourrait être achevé sans remercier de tout coeur mes parents et ma soeur Laurence qui depuis 22 ans (bientôt 23) ont eu le courage de me supporter et soutenir. Je ne peux d'ailleurs que vous féliciter!

Tack *Thanks* *Merci*

Magali

Contents

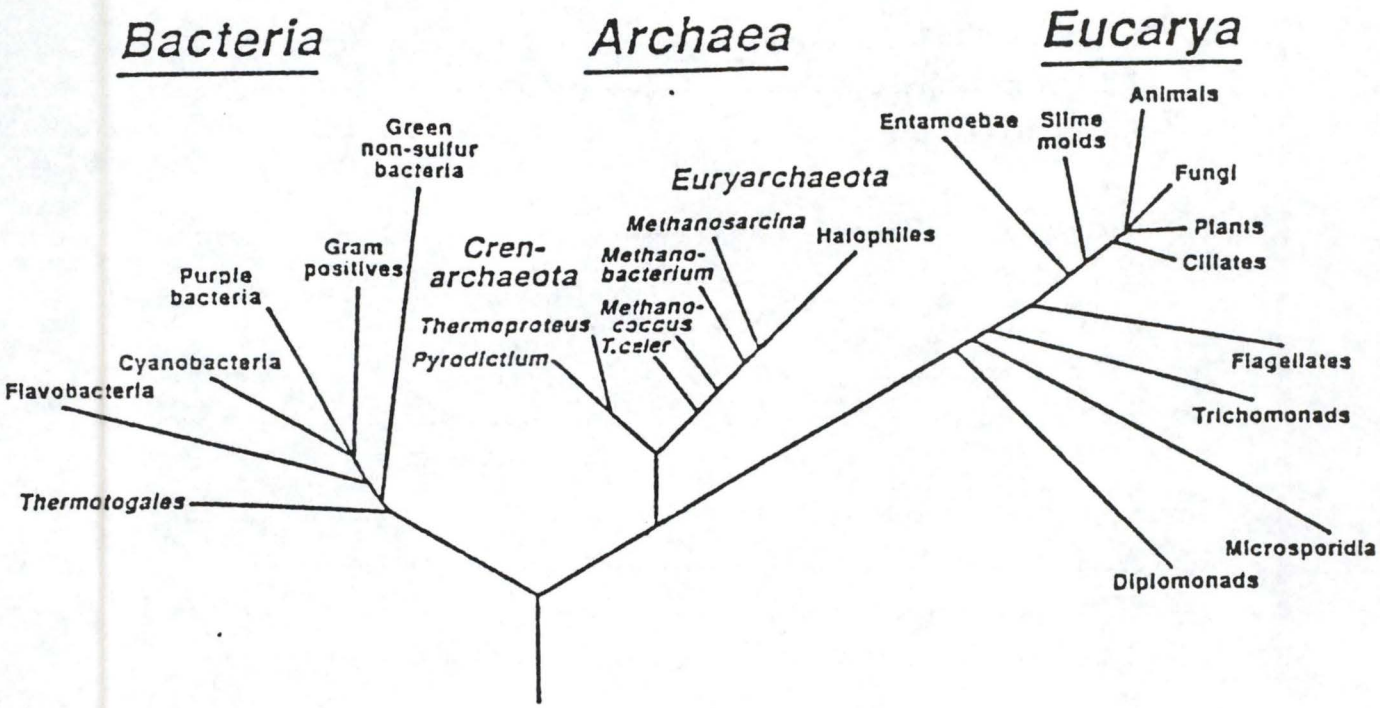


Figure 1: Universal tree in rooted form, showing the three domains: Archaea, Bacteria and Eucarya.

Introduction

1. Structure and evolution of bacterial genomes

1.1. Universal genealogical tree

All life forms could be accommodated in five kingdoms of organisms: the Animals, Plants, Fungi, Protistes and Monera, and in two superkingdoms with the eukaryotes and prokaryotes.

The prokaryotic cells are divided into the Bacteria and the Archaea domains. The Eukaryotes fall into the Eucarya domain (Fig. 1).

1.2. Comparative molecular data on bacterial genomes

Within this phylogenetic tree, the genomic evolutionary forces have had an impact on the genome size, nucleotide composition and general organisation, on the genetic code and in particular on the nature, the number and the disposition of the genes coding for the translation apparatus.

a. Genome size

The genome is the complete heritable set of genes of an organism. Bacterial genome sizes vary from 600 kb (*Mycoplasma genitalium*) to 12.8 Mb (in *Calothrix* strains). Organisms with values between these two extremes includes *Borrelia burgdorferi* with 950 kb, *Mycoplasma hypopneumoniae* (1.14 Mb), *E. coli* (4.7 Mb) and *Myxococcus*

xanthus (9.5 Mb) (Krawiec and Riley, 1990). The number of genes spans the range from 400 to 8000.

The genome of bacteria can be divided into three fractions: chromosomal DNA, plasmids and transposable elements (which are common components of the genome).

It is suggested that the larger genomes may have evolved from smaller genomes by genomic duplications as well as gene elongations, and that the smaller genomes may be the result of a reductive mode of genomic evolution (Andersson and Kurland, 1995).

b. Nucleotide composition

The nucleotide composition (in %) is another important genomic trait in procaryotes. It seems to be related to phylogeny with closely related species having the same G+C content. However, in the alpha proteobacteria for example, some species have high G+C contents (such as *Agrobacterium tumefaciens*) while others have low G+C contents (such as *Rickettsia prowazekii*). The genomic G+C content values vary from 25% to 75% (Muto and Osawa, 1987).

This variation in nucleotide composition among genomes can be explained by two hypotheses:

-the selectionists suggest that different environmental factors may influence the G+C content, the observed differences resulting from adaptations of bacteria to their environment (Li and Graur, 1991). For instance the thermophile bacteria have a high G+C content to stabilise the proteins at high temperature.

-the mutationists invoke biases in the mutation patterns as an explanation for differences in genomic G+C contents (Muto and Osawa, 1987). The redundancy of the genetic code permits changes in base composition which are neutral. A change in the third codon position most often do not alter the amino acid sequence of the protein and is therefore most strongly affected by biased mutation patterns (change in the second or the first position can also occur) (Andersson and Kurland, 1995).

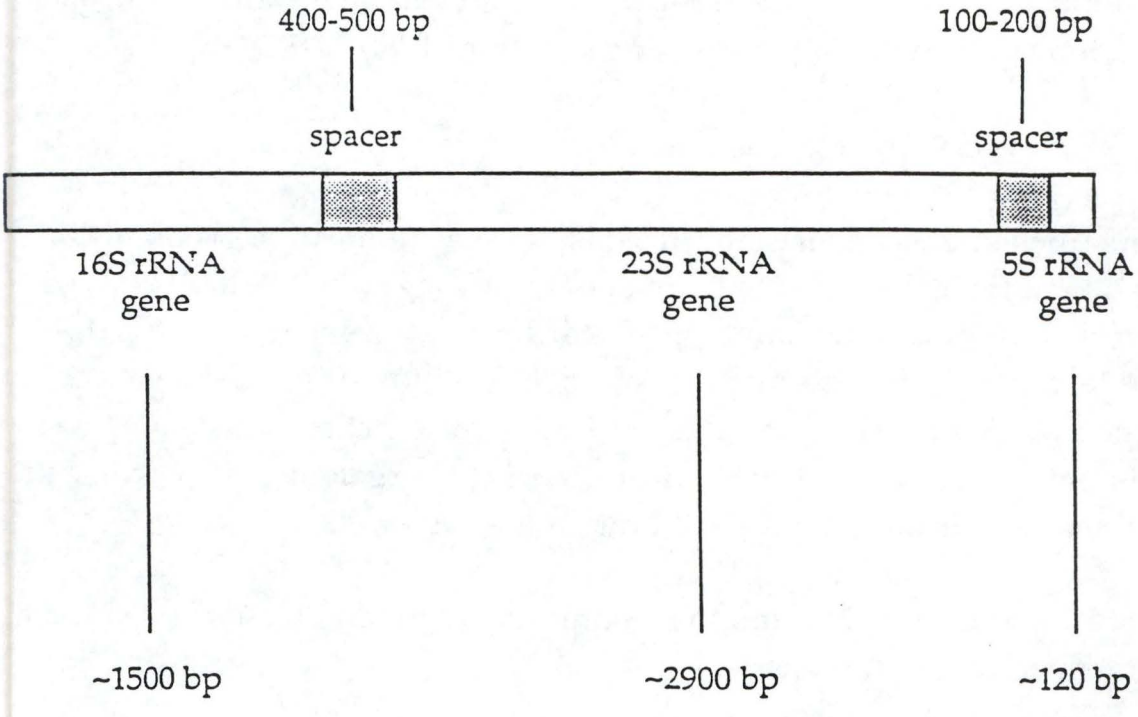


Figure 2: Schematic representation of a typical rRNA operon.

Arrangement of rRNA	Organisms	Number of copies
16S-23S-5S	*E. coli*	7
	Bacillus subtilis	10
	Mycoplasma capricolum	2
	Streptomyces coelicolor	6
23S-16S-5S	*Vibrio harveyi*	1
5S and 16S-23S	*Mycoplasma hyopneumoniae*	1
16S and 23S-5S-23S-5S	*Borrelia bugdorfei*	1
16S and 23S-5S	*Thermus thermophilus*	2

Table 1: Copy number and arrangement of the rRNA genes in some bacterial species

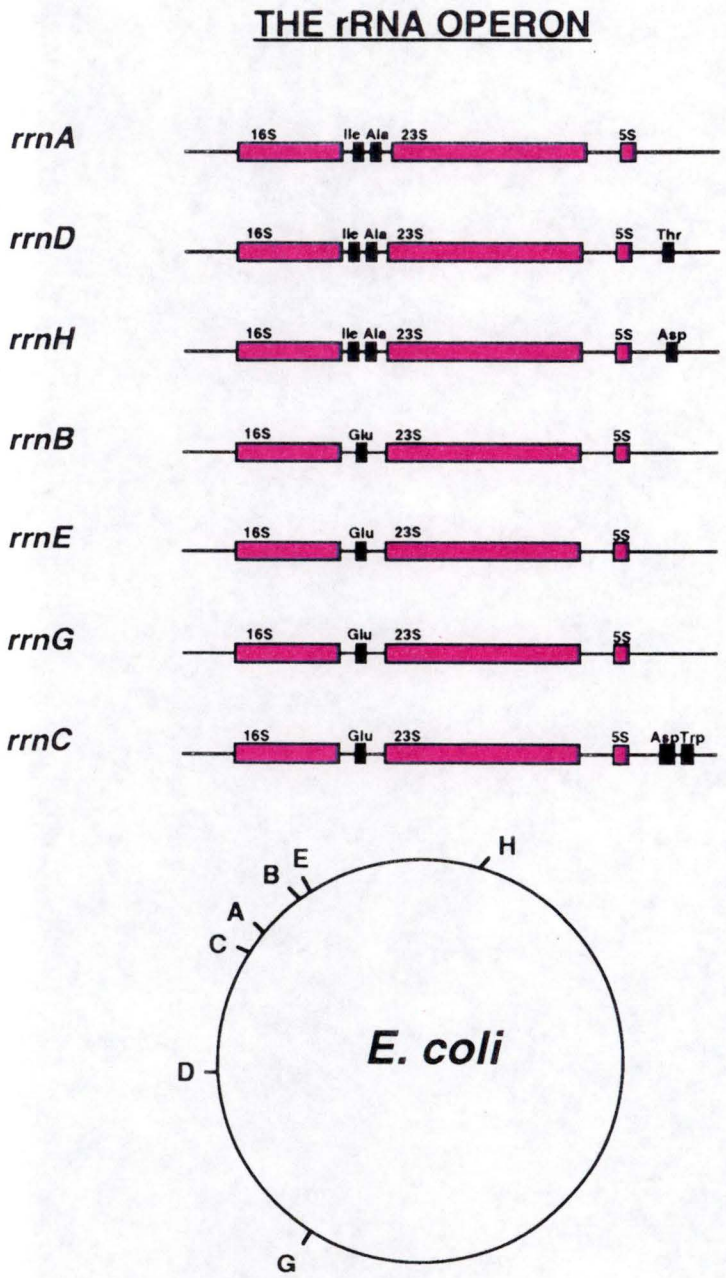


Figure 3: The seven rRNA operons in E. coli. Abbreviations (A, D, H, B, E, G, C) used to indicate the position of the different operons in the genome. 16S = rRNA small subunit, 23S = rRNA large subunit, 5S = rRNA subunit, Ile = tRNA for isoleucine, Ala = tRNA for alanine, Glu = tRNA for glutamic acid, Thr = tRNA for threonine, Asp = tRNA for aspartic acid, Trp = tRNA for tryptophane.

Changes in genomic global nucleotide composition involve differences in the frequencies of AT-rich or GC-rich codons, and thus produce compositional differences in amino acid sequence for homologous proteins.

Some bacterial proteins, tRNA or rRNA genes have a remarkably conservative amino acid composition and may be relevant to discuss the evolution of code.

c. Architectures

Surprisingly, the genes for components of the translation apparatus are often organized similarly among different bacteria and archaea (and even eukaryotes). But there are exceptions in their organization and it is suggested that such unusual genomic organizations may contain important information about the evolutionary history of these bacteria.

1.3. rRNA genes locus and flanking region

In most eubacterial species, the three rRNA genes are closely linked in the order 16S-23S-5S in one transcription unit, corresponding to the so called rRNA operon. This operon is typically composed of a promoter, a small subunit gene (SSU, 16S or rrs), a spacer usually containing tRNAs, a large subunit gene (LSU, 23S or rrl) and a spacer and a 5S gene (rrf). Other tRNAs may be found downstream of the 5S gene. (Fig. 2)

The rRNA operons are usually found in several copies in the genome, ranging from several hundred of copies in vertebrate genomes to a few copies in small bacterial genomes (Table 1). The different rRNA operons of *E. coli* are shown in the figure 3.

A very low rRNA copy number is often observed in obligate intracellular organisms together with a small genome size.
It is generally believed that obligate intracellular parasitic bacteria derive from extracellular free-living bacteria. The development of mechanisms for entering host cells, for multiplying, for survival and

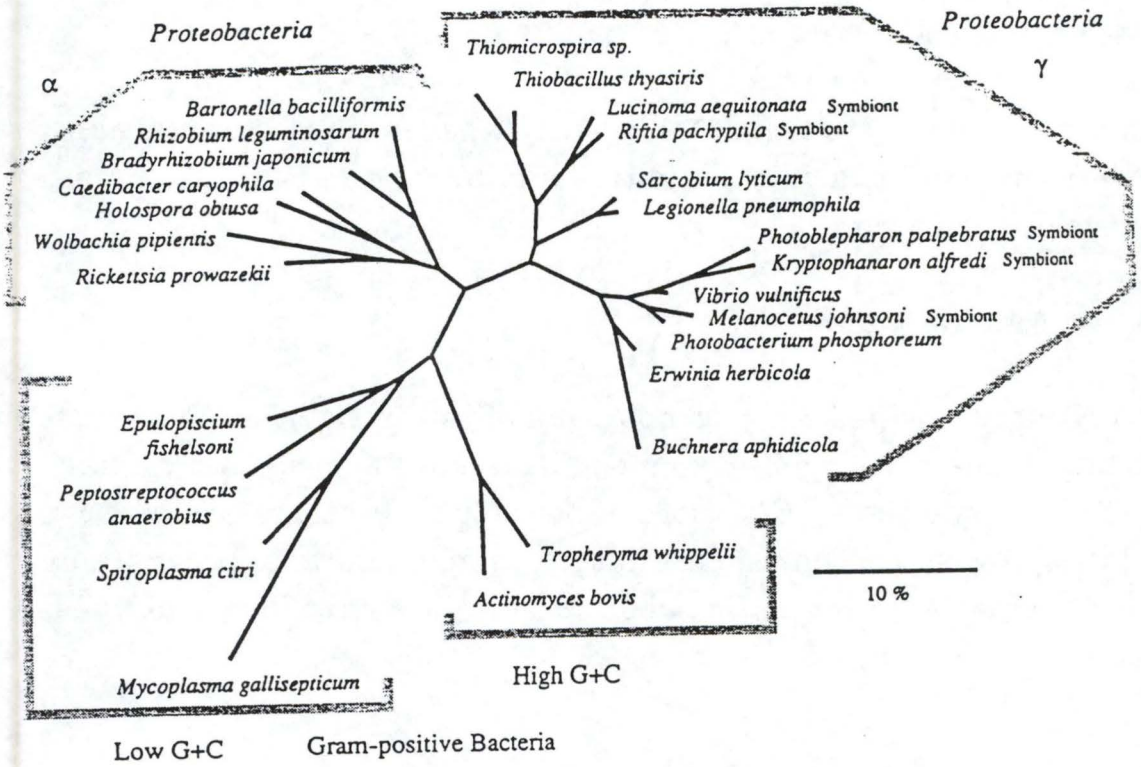


Figure 4: Phylogenetic tree showing *R. prowazekii* within the alpha-proteobacteria.

for infecting new cells are important for a parasitic life style. During the transition to this parasitic life style, the obligate intracellular bacteria may have gained new characters and lost old characters.
The very rich medium (cytosol) could explain the reduced number of genes resulting from a degenerative mode of evolution from a larger, bacterial genome. The objective behind a small genome could also be a quick replication and division.

2. Rickettsia characteristics

In this report we will study the unusual organisation of the rRNA operon in the Rickettsia genus.

Rickettsia is an obligate intracellular gram-negative eubacteria, member of the branch of the alpha-proteobacteria (purple bacteria) (Fig. 4). Rickettsia is interesting to study because of its small genome size (1.2 Mb), its close phylogenetic relationship to mitochondria and its obligate intracellular lifestyle.

2.1 *Rickettsia*, as a parasitic bacteria

Rickettsia have natural arthropod hosts (ticks, mites or insects). Man and other vertebrates may be affected by these organisms, most of which are pathogenic.

The penetration of the *Rickettsia* into host cells involves the phagocytosis of the microorganism after its attachment to the cell surface. Penetration of a *Rickettsia* into its host cell is accomplished within 10-20 minutes, but the intracellular multiplication process is very slow. It escapes quickly from the phagosome into the cytosol where the parasite grows. This very rich medium produces energy and every intermediates of most pathways that *Rickettsia* needs for its growth. Some species are also capable of growth and division in the nucleus.

Species	Human disease	Principal invertebrate host	Principal vertebrate host	Geographic distribution
Spotted fever group				
R. ricketsii	Rocky Mountain Spotted fever	Several genera of ticks	rabbit,hare, dog	Northand South America
R. conorii	Boutonneuse fever, Mediterraneanen fever	Several genera of ticks	rodent, dog	Mediterranean countries, Black Sea Coast of Russia, Africa
R. sibirica	Siberian tick typhus	Several genera of ticks	rodent	Eurasia, Asia
R. akari	Rickettsiapox	Mouse mite	rodent	worldwide
R. australis	Queensland tick typhus	*Ixodes* tick	rodent, marsupial	Australia
R. parkeri	No known disease	*Amblyomma*	unknown	USA
R. montana	No known disease	*Dermacentor* tick	unknown	USA
R. amblyommi	No known disease	*Amblyomma*	unknown	Missouri, USA
R. rhipicephali	No known disease	*Rhicephalus*	dog, meadow vole	USA
R. honei	Flinder´s Island spotted fever	*Ixodes* tick		Flinder´s Island (Tasmanie)
R. helvetica	No known disease	*Ixodes* tick	meadow vole	Switzerland
R. massiliae	No known disease	*Rhicephalus*	dog	France, Greece
R. africae	African tick-bite fever	*Amblyomma*	dog, ruminant, rabbit	South Africa
R. japonica	Japanese spotted fever	unknown	rodent, dog	Japan

Species	Human disease	Principal invertebrate host	Principal vertebrate host	Geographic distribution
Typhus group				
R. typhii	Murine typhus	Flea, louse	rat, mouse	worldwide
R. prowazekii	Epidemic typhus	Body louse	human	worldwide
	Recrudescent typhus		human	worldwide
	Sylvatic typhus	Flea and louse	flying squirrel	Eastern United States
R. canada	No known disease	*Haemophyalus*	rabbit, hare	Ontario, California
ELB agent	Murine-like typhus	Flea	opossum, dog	Texas, California
Scrub typhus group				
R. tsutsugamushi	Scrub typhus	Trombiculid mites	rodent	Japan, Eastern Asia to India, Australia, Pacific Islands
Undetermined				
R. belii	No known disease	Several genera of ticks	unknown	USA
AB bacterium	No known disease	Ladybird beetle	unknown	England

Table 2: *Rickettsia* genus

The *Rickettsia* genus is divided into three groups: The Spotted Fever Group (SFG), the Typhus Group (TG) and the Scrub Typhus Group (STG) (Table 2)

a. The Spotted Fever Group (or SFG)

The members of the SFG grow and divide in the cytosol, but they can also grow within the nucleus.
The average G+C content of the coding sequences has a value of 32 and 33 %.
R. ricketsii, R. sibirica, R. parkeri, R. montana and *R. amblyommi* will be discussed later in this study.

b. The Typhus Group (TG)

The characteristic hosts of the TG species are the insects.
The members of the Typhus Group are not able to invade the nucleus.
The G+C content has an average of 29-30 %.
R. prowazekii and *ELB agent* will be the object of this project.

c. The Scrub Typhus Group (STG)

The scrub typhus agent (*R. tsutsugamushi*) is very different from the species of the SFG and the TG in its morphology cell surface proteins. It is capable to grow in the cytosol and can be found in the nucleus.
This untill now unique member of the STG will not be considered in this study.

d. *R. bellii* and *AB bacterium* are not included within any of these three groups.

2.2. Genomic characteristics of *Rickettsia*

The small and unique chromosome of the bacterium *Rickettsia* has a size around 1.2 Mb (Eremeeva et al. 1993). Calculations based on a coding potential of 60-70%, a genome size of 1.2 Mb and an average

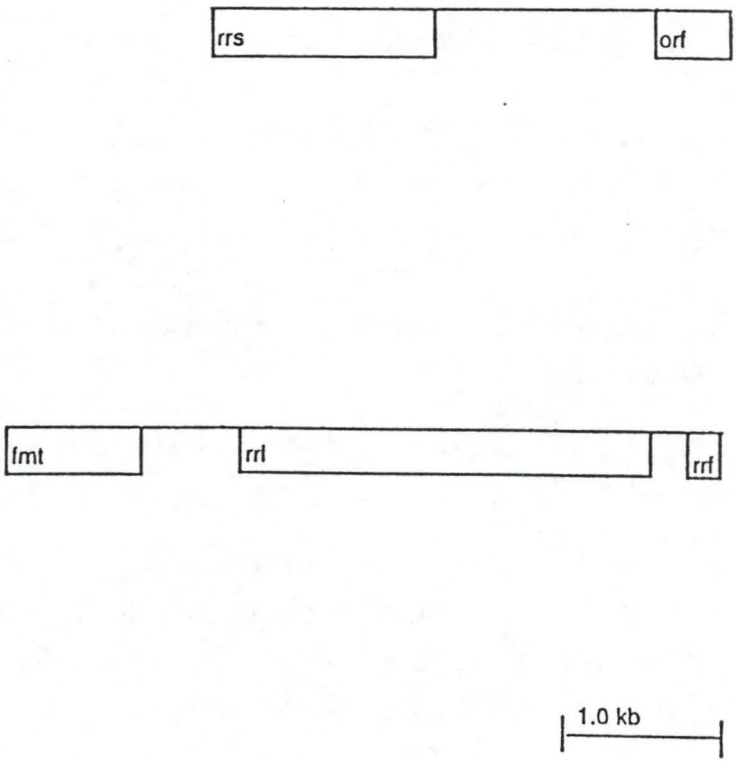


Figure 5: Organization of the rRNA genes in *Rickettsia prowazekii*. Thin lines indicate a noncoding region; Boxes indicate a coding region. Abbreviations used to indicate genes are as follows: rrs = rRNA small subunit gene, orf = open reading frame, fmt = methionyl-tRNA$_f^{Met}$ formyltransferase gene, rrl = rRNA large subunit gene, rrf = 5S rRNA gene.

gene length of 1 kb suggests that there may be around 800 genes in the *R. prowazekii* genome (Andersson and Sharp, 1995).
Rickettsia is a AT-rich bacteria with average values between 29 to 33% due to a mutational bias towards A+T instead of G+C base pairs. Protein coding and intergenic region have G+C content values of 32.5% and 21.6%, respectively. The functionally less important third codon position is more biased than the first two positions (Andersson and Sharp, 1995).

2.3. rRNA operon in *Rickettsia prowazekii*

In most eubacterial species, the three rRNA genes are closely linked in the order 16S-23S-5S grouped into one operon. The number of copies in bacteria varies from 1 (*Mycoplasma hypopneumaniae*) to 11 in some Bacillus species.
Studies of the rRNA genes in *Rickettsia prowazekii* shows that 16S rRNA gene is separated from 23S-5S rRNA gene cluster. One unidentified open reading frame (ORF) was detected downstream of the 16S rRNA gene. An ORF encoding the methionyl-tRNA$_f^{Met}$ N-formyltransferase gene (*fmt* gene) was found upstream of the 23S rRNA gene. No tRNAs have been found upstream from 23S gene nor in the region downstream from 16S gene (Andersson et al., 1995) (Fig.5). In *Rickettsia* there is only one copy of each of the rRNA genes.

2.4. Interest to study the organization of the rRNA genes in *Rickettsia*

It has been shown that *R. prowazekii* has one copy of the 16S rRNA gene, one copy of the 23S-5S rRNA gene cluster (Andersson et al., 1995) and also one copy of the *tuf* gene which encodes elongation factor Tu (EF-Tu) (Syvänen et al., 1995). Since these genes are found in several copies in others bacteria, and the organization seems to be unusual in *R. prowazekii*, it is suggested that the genome of *Rickettsia* has evolved from a lager genome.
Because of their ability to recombination, duplication, inversion and deletion, the sequences which are found in multiple copies (as the rRNA genes) are interesting to study the hypothesis of a reductive evolution.

Goals of this project

The study of the unusual organization of the rRNA genes in the genus *Rickettsia* is divided in two parts.

First, we will study the *fms-fmt* arrangement in *Rickettsia prowazekii* (*fms* encodes peptide deformylase and *fmt* encodes methionyl-tRNA$_f^{Met}$ N-formyltransferase gene).
The colocalisation of these two genes might be a conserved feature within the bacterial kingdom. Probably because of their related fonctions, the two genes *fms* and *fmt* are found linked to each other in *E. coli*, *Thermus thermophilus* and *Haemophilus influenzae*. The conservation of this linkage in these three species might be a conserved trait within the bacterial organisms (Meinnel and Blanquet, 1994).
Recently the complete sequencing of the genome of *Mycoplasma genitalium* has shown a different organisation in which they are separated from each other by 328 kb (Fraser et al., 1995).
In the present work, the *fms* gene has been sequenced in *Rickettsia prowazekii* and a PCR amplification shows the linkage of *fms* gene with the *fmt* gene which is situated upstream of the 23S rRNA gene.

Secondly, we will analyse and compare the sequences of the *fmt* gene and the sequences of the spacer between *fmt* and 23S genes for seven different species of the genus *Rickettsia*.
Noncoding regions are often altered by random mutation without any troubles in the genome. Thus, the study of the spacer is interesting as a fast clock of evolution and also for eventuel diagnostic of the different species of *Rickettsia*.

Conserved organization of the genes encoding Peptide Deformylase and Methionyl-tRNAf^{Met} Formyltransferase in *Rickettsia prowazekii*

Magali P. C. Romedenne*, Siv G. E. Andersson and Charles G. Kurland

Department of Molecular Biology, Uppsala University, Biomedical Center, Box 590, S751 24 Uppsala, Sweden

**Present address: Laboratory of Genetics, Facultés Universitaires Notre dame de la Paix, 61 rue de Bruxelles, 5000 Namur, Belgium*

Running title: *fms* and *fmt* in *Rickettsia prowazekii*

Keywords: *fms*, *fmt*, genomic organization, *Rickettsia prowazekii*

Summary

We describe here the organization of the genes encoding peptide deformylase (*fms*) and methionyl-tRNA$_f^{Met}$ formyltransferase (*fmt*) in *Rickettsia prowazekii*. Sequence analysis of cloned fragments as well as a PCR analysis of genomic DNA using primers from each of the two genes confirm that the *fms* and *fmt* genes are tightly linked together in the *R. prowazekii* genome.

1. Introduction

A characteristic feature of translation intitiation in bacteria, chloroplast and mitochondria is the N-formylation of the initiator Met-tRNA$_f^{Met}$, which is catalyzed by methionyl-tRNA$_f^{Met}$ formyltransferase (MTF) using formyltetrahydrofolate as a cofactor. After about 15-30 rounds of elongation the formyl group is removed by the peptide deformylase (PDF) and in about half of the proteins the methionine residue is removed by the methionine aminopeptidase.

Genes encoding functionally related products are sometimes organized into co-transcribed units in bacterial genomes. The genes encoding the enzymes involved in the formylation and deformylation of the Met-tRNA$_f^{Met}$, *fmt* and *fms*, are genetically linked in several different species of diverse phylogenetic affiliations such as *Escherichia coli* (Guillone et al., 1992), *Haemophilus influenzae* (Fleishman et a1., 1995) and *Thermus thermophilus* (Meinnel and Blanquet, 1994). In contrast, in the 0.6 Mb genome of *Mycoplasma genitalium* these two genes are separated from each other by 328 kb (Fraser et al., 1995). Disruption of otherwise conserved bacterial operons, such as for example the rRNA operon, has been observed in several small sized bacterial genomes and suggested to be a consequence of a reductive mode of evolution, where extensive deletion and intrachromosomal recombination events have rearranged an ancestral genome with a size and an organization that was more typical of the bacteria (Andersson and Kurland, 1995).

Since *R. prowazekii* has a small-sized genome of 1.2 Mb (Eremeeva et al., 1993) it was of interest to determine how the *fms* and *fmt* genes are arranged in this organism. It has previously been shown that the *fmt* gene is situated upstream of the 23S rRNA gene (Andersson et al., 1995). In this paper, we present evidence which suggests that the *fms* and *fmt* genes are closely linked to each other in the *R. prowazekii* genome.

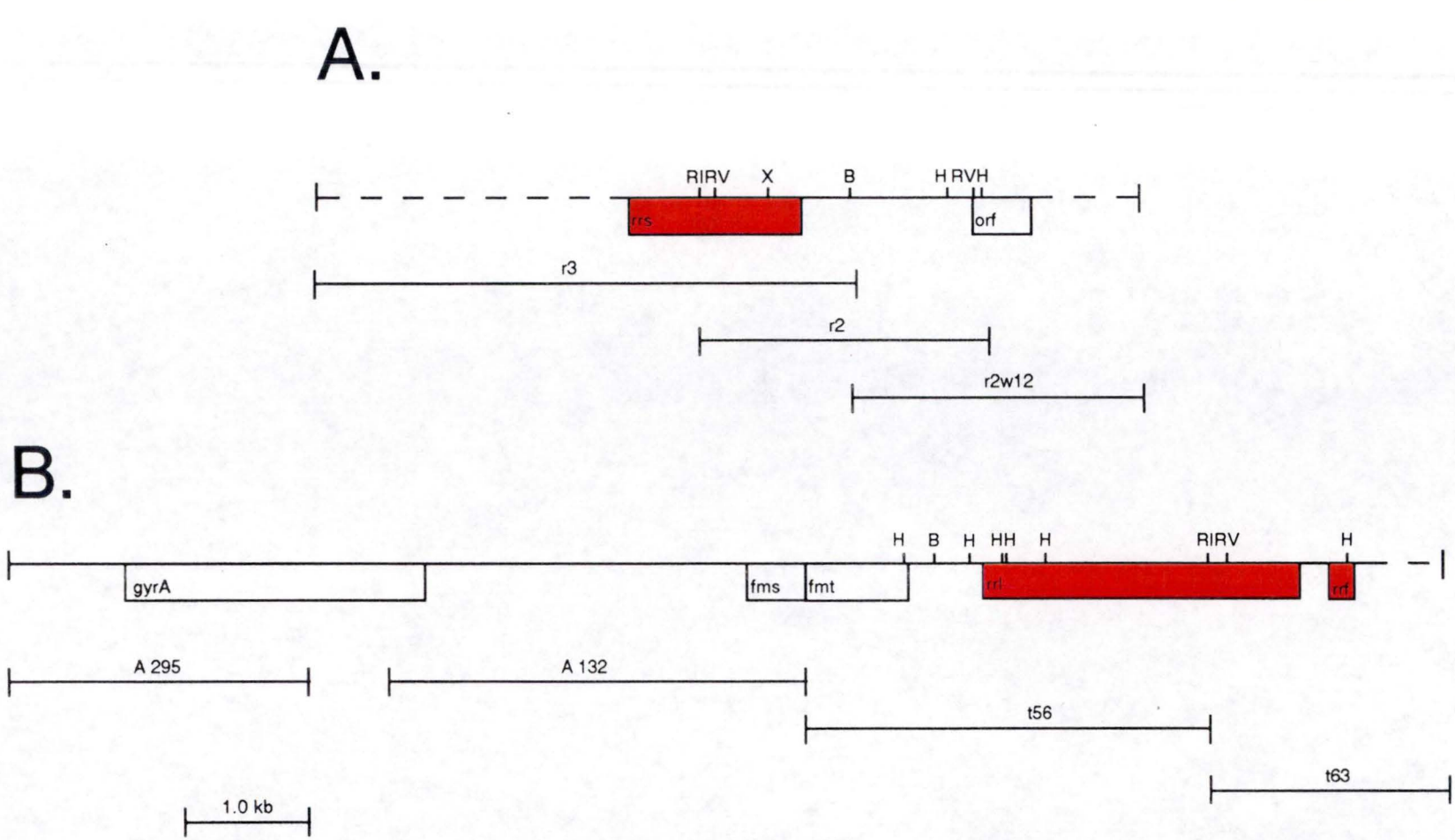


Figure 1a: Schematic presentation of the region surrounding the *fms* and *fmt* genes in *R. prowazekii*. (A) small subunit rDNA region. (B) large subunit rDNA region. The location of clone 132 that was used in sequencing is indicated. Thin lines indicate a noncoding region; Boxes indicate a coding region. Abbreviations used to indicate restriction enzyme sites are as follows: B = *Bam* HI, RI = *Eco* RI, RV = *Eco* RV, H = *Hind* III, X = *Xho* I. Abbreviations used to indicate genes are as follows: rrs = rRNA small subunit gene, orf = open reading frame, *gyr A* = gyrase A gene, *fms* = peptide deformylase gene, *fmt* = methionyl-tRNAf^{Met} formyltransferase gene, rrl = rRNA large subunit, rrf = rRNA subunit.

2. Material and method

2.1. Purification of plasmids.

DNA plasmid containing the clone 132 was prepared using Qiagen Maxi Kit (KEBU, Stockolm, Sweden).

2.2. DNA sequencing

Autoread TM sequencing kit (Pharmacia, Uppsala, Sweden) which used Sanger's method (dideoxy chain termination method) was used for sequencing the clone 132.
For the DNA sequencing of both strands of the insert we used modified T7 DNA polymerase (sequenase) and fluorescent forward and reverse primers. The sequence of the gene was performed by following the primer walking method and the non-fluorescent primers deduced are used by addition of fluorescent ATP (Pharmacia, Uppsala, Sweden).
The product of the sequencing reaction were separated and analysed with the aid of an ALF (Automated Laser Fluorescent) DNA sequencer (Pharmacia, Uppsala, Sweden).

2.3. Sequence analysis

The data banks were searched for homologous sequences using the BLAST (Basic Local Alignment Search Tool) program (Altsihul et al., 1990).
The predicted protein products of the *fms* and *fmt* genes were aligned (Fig.3a and 3b) using CLUSTAL W (Higgins et al., 1992).

2.4. PCR amplification

The polymerase chain reaction (PCR) amplification were carried out in 50 µl reaction mixture containing 10 nmoles of dATP, dCTP, dGTP and dTTP, 50 pmoles each of indicated primer, 20 to 50 ng of

```
ATG GAA TTC ATC TTG ACG GTT ACA GAG AAC GGT TTT GGT AAA AGA AGC TCG GCA TAT GGT
M   E   F   I   L   T   V   T   E   N   G   F   G   K   R   S   S   A   Y   G

TAT AGA ATT ACT GAT CGT GGT GGT AGT GGT ATA ATA AAT ATG GAT ATT AAC GAT AAA ACC
Y   R   I   T   D   R   G   G   S   G   I   I   N   M   D   I   N   D   K   T

GGT TTA GTC GTT GGT GTT ATG CCA GTT AAA ATG GAT GAT GAG TTA ATG CTA ATC ACA AAT
G   L   V   V   G   V   M   P   V   K   M   D   D   E   L   M   L   I   T   N

AGT GGT AAG TTA ATT CGC TGT AAA CTT GAG TCG GTA CGT ATT ACG GGT CGT AAC ACT AGT
S   G   K   L   I   R   C   K   L   E   S   V   R   I   T   G   R   N   T   S

GGT GTA ATT CTG TTT AAA CTA GAT GAT GAT GAA AAA GTA GTA TCT GTT TCT TTA ATA GCT
G   V   I   L   F   K   L   D   D   D   E   K   V   V   S   V   S   L   I   A

GAA ACT TCT GAA AGC GGA GAA GCT AGT GAA TTA GCG GAA GAG GGA TTA GAG AAT GAT GTT
E   T   S   E   S   G   E   A   S   E   L   A   E   E   G   L   E   N   D   V

AAG GTA TAG TAGTGGAAGTATATGCTGATTTTATTTAACCCATGGACATTTTTTATATATTATTCATTTAAAACCGG
K   V   *
CTCTTTTTGCAGCGCATGACTTGATATATAGTACTAAATCAATACTAATATTAGTATTTTTAGTCTCGATCATTGAATAG
TTATGATATTAATTAAAATGTCTTGTACTTTCATCTTAATTAAATAAGAACAAATTTATGTGCTTTTGTTGTTAAATTAC
TTTACCTATTTTTATAACCATGATGTTAATCAATGCATTTATCTGTTTGTTTGCGTTTAGCAATATCTCATATAAAATTC
TAGAAAACACTAATTTTTTATAAAATAGATATTCAATTTTTGAATATACAAATATTTACTTTTTAATCTTGTTACTACTA
TAATAGGTTTCAGCGAGTAATAATATTTCTACATTAATTAGATCAATCTATTCCAATGCACTAATTATTATTAAATCAGT
TTTATGTTTCGTAACATACAACTTGTGGCTATTAATTATAGCTGTTATTTTTGTTACACACA*CTCAATTACTTTCAAAT
CTTATTATTTTTTGTTTACCTTTTAAATGTTAGC*AAAaGAGCTAGCCcAAAGATTTCATTCATATATTTTATGT*TATC
TGGTATGTGTAGTTAGCGCATACTCCTATAATGATATCAATTTTATTTATCTA*TATAGTATAGGAGAAACGATAAATGA
AAGT*TTGATGTTTCTTTGATTTTGCTATCATTATTTCTAATGACTTTAGTATTTTTAAAAATTTTGTGTTTAAAGGAAT
TGCATTGTACTTCAAGATAGCAGCATTGTAATATTAAATTGTTTATCTATATAATCTATGATCTTTACCATTAAACTAAA
GTTTTAGAAAGTATTTAAAGATTATCCCTTTATTTTTTATGATAATTGCATATGCTGCAGTTTTGTTATTGCTCGTTTAG
ACTTCAGTAATAGCGGAGTTAAGAATATTAAAAAGTGTGATTTTAACGATAAATTCCGAGATATGATGAGATGATATAAC
TTGATCTAATATTAGTAGAAAATAATACAAAAAATAAAGCGCTTGCTCGTACTGTTATTATAGTAAAAGTAAATATTTAT
TCACTATTGAGTGTTTTACTCCGATTTTAGCGTAAGTAATCCTAAAAAGTTGTTAATTAGCTTTAATAGCGATAATAATA
TCAAGGTGGTGTAATTATTGTAATGCTTGCTATTCTGTTAAAAAGACATATATTTATATCTTATTTAATCCTCTTATTTT
TTATAAAATGATGTTTTTTTTAAGATTGATAGATAATCTATATATTCTATAACTCAGAAACCTTaAAATTAACGCAAATTT
TGTATGGATTAAATGCCATGT*AACACT*AATAAAATTTTAGCAATATTATTAAGTATCTCTTTAGGTTTCTGACAAAAA
TTCTCTTAACGCACTATATTTCAACTAAAAATTAATGATGATAATATCGTTCTTTTATCAATTAAGTATTACATTATTTA
TATTCAATAAATTAATCCATATATTATTATTTTTTTTATAGAAGAAATTTTTTGAATTGGAATTTTTTATTAATTAAGATAA
TTGCTCTTTAAAAGCACAAAAGCATATAATCTTGTGTAGAAGATTATTTTTAGTATAAATTAATGATGTGGTTATTAATT
GCCATTGGGTTAACCGTTTCAGTTCTTATTATAGGTATAGTGTCTATGGCTATAGGAGGAAAAtTTGATAAGAAATTGAG
CTTAAAACTAATGACTTTGAGAGTATTTTTTCAGGCCGTTTCTATATTTTTATTAATTTATTTTTATAAAGTTCATTTTT
TATACTAAAATTAAGTGTAATGATCTATTTTTATAGGTGAATTTTTATTTTTTAGTAAGATTTATATCTCTTAAACTATT
TACTTCATGAGGTATGTAAAATCAAAAGATATTATAACCCAAACTGTTTAATACCAATTCAGATCAATAACTACTTAATT
TAT*TCTATCATCTCTACAAATAATTTTTTCAATTTGTCTAGAGTTTTTTGTTATAACTTGTTCTACTTCATCTCTAGAA
CCTTTTAAA*TCTCATCAATTTCTCTAGCATTGATAATGCATGATATCGCTATAATGTGTTGAGTTATTCACAAATTTCT
CCATTATCACACTATAATATATTATTAATCTTCCTTAATTGATACATATATTTTATGCTGTCTGATATTTGTTATAATCA
TGCAAATAAATGATTGAAAGTAGGTGAGCTTATTATACAACTATTACCGGAATTAGCAAGATTTACTTTCTTTGCATTTA
ATTTTATTGTATATTATAAGTAAAAGATTAAGATTGAATT ATG TCA ATA TTC TCC ATA GTA ACC GCA CCT
                                          M   S   I   F   S   I   V   T   A   P
GAC GAA AGA TTA AAG CAA AAA TCT AAG CCT  GTT TTA GAA TGT ACC GAT CAA ACA CGA AAA
D   E   R   L   K   Q   K   S   K   P   V   L   E   C   T   D   Q   T   R   K

TTT ATG CAT GAT ATG CTT GAA ACT ATG TAC  AAT GCG GAT GGT GCA GGG CTT GCG GCA GTA
F   M   H   D   M   L   E   T   M   Y   N   A   D   G   A   G   L   A   A   V

CAA GTA GGA ATA TTG GCA CGT ATT TTA GTA  ATT GAT ATA AAA GAA TAT GAT CCA GTA GAA
Q   V   G   I   L   A   R   I   L   V   I   D   I   K   E   Y   D   P   V   E

AGA CCG AAA GAT TTT TAT CCA CTT TTT ATA  GTG AAC CCT GAA ATA ATA GAA AAA TCA ACA
R   P   K   D   F   Y   P   L   F   I   V   N   P   E   I   I   E   K   S   T

GAG TTA GTG ACT GCT AAT GAA GGC TGT ATC  TCA TTA CCA AAG CAA CGT ATT GAG GTT ACG
E   L   V   T   A   N   E   G   C   I   S   L   P   K   Q   R   I   E   V   T
```

```
AGA CCG GAG TCT GTA AAG ATA AGA TAT TTA  GAT TAT CAT GGT AAA TCA CAA GAG CTA AAA
R   P   E   S   V   K   I   R   Y   L    D   Y   H   G   K   S   Q   E   L   K

GCA AAT GAT TGG CTT GCA AGA GTT ATT CAG  CAT GAG TAT GAT CAT TTA GAA GGT AAG CTT
A   N   D   W   L   A   R   V   I   Q    H   E   Y   D   H   L   E   G   K   L

ATG GTT GAT TAT TTA AGT AAT TTA AAA CGA  GAT GTA GTA CTT CGT AAG CTT AAG AAA CTT
M   V   D   Y   L   S   N   L   K   R    D   V   V   L   R   K   L   K   K   L

AAA AAT AAT ATA GTG TGA AAGTAATTTTT ATG  GGA ACG CCT GAA
K   N   N   I   V   *               M    G   T   P   E
```

Figure 1b: Nucleotide sequence of the 5′ region of *fms* and the deduced amino acid sequence of the *fms* gene product. Asterisks signify stop codons and, in the spacer, asterisks signify ambiguities.

a b λ

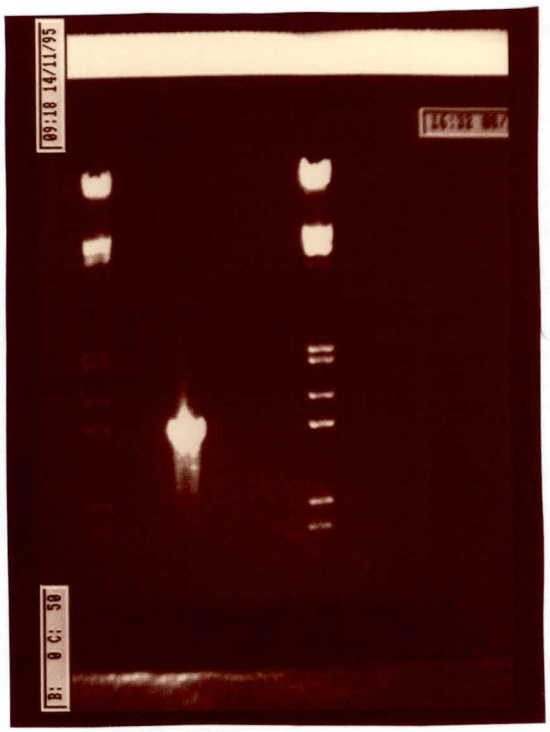


Figure 2. PCR analysis. Primers 132+2 and ot56-7 with genomic DNA from R. prowazekii isolated in H. Winklers laboratory, Mobile, Alabama (lane a). Lane b are water controls. Molecular marker were based on digests of lambda DNA with the enzymes EcoR I and Hind III, producing the following sizes of bands in kb: 21.2, 5.1, 4.9, 4.2, 3.5, 2.0, 1.9, 1.584, 1.375, 0.974, 0.831, 0.563 and 0.125.

chromosomal DNA, 1.25 units of Taq polymerase (Promega Biotech, Madison, WI) and 5 µl 10X buffer (0.5M Tris-HCl pH 8.8, 0.15M $(NH_4)_2SO_4$, 15mM $MgCl_2$, 1% Triton X-100 and 0.1% gelatin). After a hot start, 30 cycles of amplification was performed in a DNA thermal cycler (AMS Biotechnology, Taby, Sweden), with a standard profile of denaturation at 94°C for 1min, annealing at 53°C for 1 min, and extension at 72°C for 1 or 2 min. Following amplification, 5 to 10 µl of the amplified material was fractionated by electrophoresis on 1% agarose gels.

3. Results

3.1. Nucleotide sequence of the *fms* gene and upstream regions.

In order to study the arrangement of the *fms* and *fmt* gene, we selected for sequence analysis clone 132 (Fig. 1a) from a set of 1000 clones randomly isolated from a genomic library constructed in lambda Zap II (Andersson et al., 1995). The nucleotide sequence of the 3 272 base pairs long insert revealed one complete open reading frame of 525 base pairs between positions 2718 and 3243 (Fig. 1b). The overall G+C content of this ORF is 33.7%, with the first and the second codon positions having higher G+C content values (50.3% and 29.1%, respectively) than the third (21.7%). The observed position-specific differences in base frequency patterns are in perfect agreement with the results obtained from an analysis of 27 protein coding genes (Andersson and Sharp, 1995), suggesting that the open reading frame identified corresponds to an actively transcribed gene. Translation of this gene produces a protein of 175 amino acids; a search in the database for homologous sequences using the BLAST program (Altschul et al., 1990) gives strongest homology to the *E. coli* peptide deformylase, encoded by the gene *fms* (Fig. 3a)

A short open reading within the first 363 nucleotides of clone 132 was found to correspond to the previously sequenced gene *gyrA*, encoding DNA gyrase (Wood and Waite, 1994). An homology search of the intergenic region (2355 bp) of *gyrA* and *fms* failed to identify any homologs in the databanks. The average G+C content of this regions is only 23.3%, which is close to 21.4%, previously calculated from a large

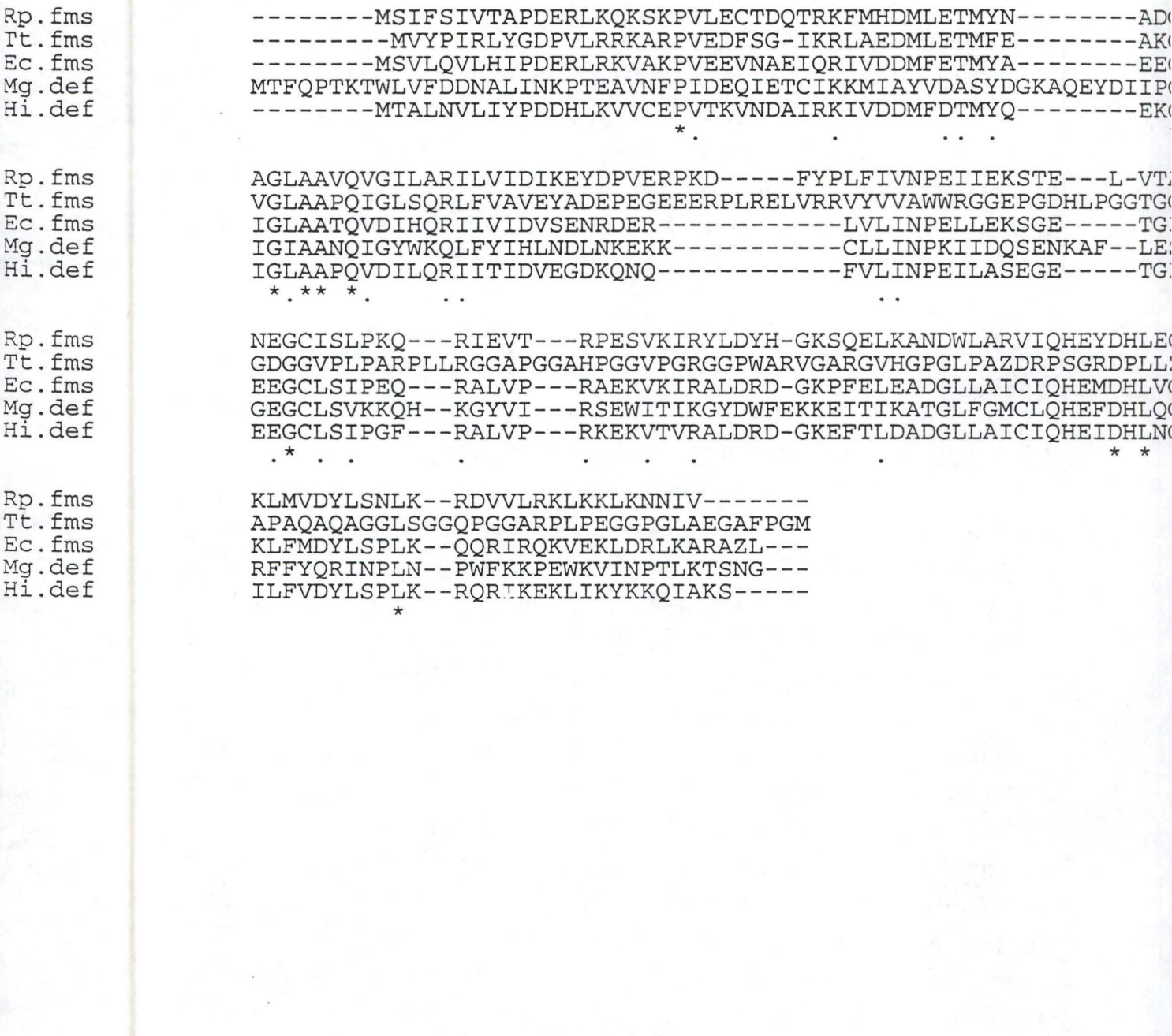


Figure 3a: Multiple alignment of the deduced amino acid sequences of the *fms* gene in 5 bacterial species. Symbols beneath the aligned sequences indicate identical residues (*) and sites with conservative replacements (.).
The 34 amino acids in the 5'end of the fms gene in M. genitalium are not represented.
Abbreviations used are: Rp = R. prowazekii, Tt = T. thermphilus, Ec = E. coli, Mg = M. genitalium, fms = def = peptide deformylase gene.

```
Rp.fmt   ---------MGTPEFAVPTLKKLIIH--HEVKAVFTQQPKAKGR-GLHLAK-SPIHQLAF
Tt.fmt   ----MRVAFFGTPLWA-PCWTPYASA---TRWSWWS-PSPTSPRAGAZGPPRAPWPATPR
Ec.fmt   VSESLRIIFAGTPDFAARHLDALLSSG-HNVVGVFTQPDRPAGR-GKKLMP-SPVKVLAE
Mg.fmt   ---MFKIVFFGTSTLSKKCLEQLFYDNDFEICAVVTQPDKINHR-NNKIVP-SDVKSFCL
Hi.fmt   -MKSLNIIFAGTPDFAAQHLQAILNSQ-HNVIAVYTQPDKPAGR-GKKLQA-SPVKQLAE
                    **   .                    .       * .     .

Rp.fmt   EHQIPVYSPSTLRNDETINLIKKIDADIIVVIAYGFIVPKAILEAKKYGCLNIHPSDLPR
Tt.fmt   RRGFPFCARRGLGRRPSSRPCARPRPRWRWWRPTGSSSPRRPWTSPHTASSTSTPPSSPS
Ec.fmt   EKGLPVFQPVSLRPQENQQLVAELQADVMVVVAYGLILPKAVLEMPRLGCINVHGSLLPR
Mg.fmt   EKNITFFQPK--QSISIKADLEKLKADIGICVSFGQYLHQDIIDLFPNKVINLHPSKLPL
Hi.fmt   QNNIPVYQPKSLRKEEAQSELKALNADVMVVVAYGLILPKAVLDAPRLGCLNVHGSILPR
                                          *    .         .      *

Rp.fmt   HRGAAPLQRTIIEGDRKSSVCIMRMDSGLDTGDILLKEDLNLERRITLDELSNKCAHLGA
Tt.fmt   TGGRPPSSGPSSPGNGRPGSPSCAWTRAWTPVPSTPSGGRPSCRTRTPWPWGTGSGTRGW
Ec.fmt   WRGAAPIQRSLWAGDAETGVTIMQMDVGLDTGDMLYKLSCPITAEDTSGTLYDKLAELGP
Mg.fmt   LRGGAPLHWTIINGFKKSALSVIQLVKKMDAGPIWKQQDFLVNNDWNTGDLSIYVEEHSP
Hi.fmt   WRGAAPIQRSIWAGDVQTGVTIMQMDEGLDTGDMLHKVYCDILPTETSTSLYNKLAELAP
           *  *       *                                .

Rp.fmt   ELLIKTLANIDN--IVPIKQSSNGITYAHKLTKAEGKINWYESAYSIDCKIRG--MNPWP
Tt.fmt   SSFLRCWSASRS--SPPGPRKGRSPTHRPFPRR---RGGWTSGRAPRPSTAATGPSSPGP
Ec.fmt   QGLITTLKQLADGTAKPEVQDETLVTYAEKLSKEEARIDWSLSAAQLERCIRA--FNPWP
Mg.fmt   SFLIECTKEILN-KKGKWFEQIGEPTFGLNIRKEQEHLDLNQIYKSFLNWVKG--LAPKP
Hi.fmt   SALIDVLDNLENGKFIAEKQDGSQSNYAEKLSKEEAQLNWSLSAMQLERNIRA--FNPWP
            .                    .    .   .                    * *

Rp.fmt   GAYFSYNDKI-----IKILRAEYFNYNHH-FIPGTVINN---KLEIACGSG-ILRVKKLQ
Tt.fmt   GATSSTGAGG----SRPZGSAPSPGKGSPGWWPGWGRRAWRWARPRASSSS--WRS-SLR
Ec.fmt   MSWLEIEGQP-----VKVWKASVIDTATN-AAPGTILEANKQGIQVATGDG-ILNLLSLQ
Mg.fmt   GGWLSFEGKN-----IKIFKAKYVSKSNYKHQLGEIVNISRKGINIALKSNEIISIEKIQ
Hi.fmt   IAYFSTEDKDGNAHTLKVYQAKVLPHQD--KPAGTILSADKNGIQIATVDG-VLNLLQLQ
                           *            *           *            ..

Rp.fmt   QESKKALNIEEFLRGTN--ILKDTILKZ
Tt.fmt   EGGPCPPPTGPGAMAWP--RGPGSARY-
Ec.fmt   PAGKKAMSAQDLLNSRREWFVPGNRLVZ
Mg.fmt   IPGKRVMEVSEIINGKHP-FVVGKCFK-
Hi.fmt   SAGKKPMSAQDLLNGRAEWFTIGKVLA-
```

Figure 3b: Multiple alignment of the deduced amino acid sequences of the *fmt* gene in 5 bacterial species. Symbols beneath the aligned sequences indicate identical residues (*) and sites with conservative replacements (.).
Abbreviations used are: Rp = R. prowazekii, Tt = T. thermphilus, Ec = E. coli, Mg = M. genitalium, fmt = methionyl-tRNA$_f^{Met}$ formyltransferase gene

number of intergenic regions (Andersson and Sharp, 1995). This is consistent with the interpretation that the region between *gyrA* and *fms* corresponds to noncoding DNA.

3.2. The *fms* gene is linked to the *fmt* gene in *Rickettsia prowazekii.*

Sequence analysis of the end-fragments of clone 132 and t56 suggests that these may share a common genomic breakpoint (Fig. 1a). To determine whether *fms* is indeed linked to *fmt* in the genome, we amplified by PCR the region in between these two genes by using a primer from the upstream region of the *fms* gene, 132+2 (ATTAGCAAGATTTACTTTCTT) together with a primer from within the *fmt* gene, ot56-7 (CACTGCCACAAGCGATT). According to sequence data we expect a genomic fragment of 1 361 base pairs to be amplified in this reaction. Indeed, a fragment of the expected size was obtained (Fig. 2), confirming that the *fms* gene is located immediately upstream of the *fmt* gene in *R. prowazekii*.

3.3. Comparative analysis of the *fms* and *fmt* gene products.

The amino acid sequences of the peptide deformylase (PDF) and methionyl-tRNAf^{Met} formyltransferase (MTF) from five different bacterial species (including *Rickettsia prowazekii* sequences) have been analysed. Three of these belong to the Proteobacteria; *Rickettsia prowazekii* to the alpha proteobacteria, *Haemophilus influenza* and *Escherichia coli* to the gamma proteobacteria. *Mycoplasma genitalium* belongs to the low G+C Gram-positive lineage. *Thermus thermophilus* has been placed in a lineage near to the green non-sulfur bacteria (Olsen et al., 1994).

The sequence alignments of the *fms* and *fmt* gene products were performed using CLUSTAL W (Higgins et al., 1992) and are presented in Figs. 3a and 3b. The two proteins show approximately the same degree of functional conservation, as inferred from the percent identity values in a direct comparison of the *R. prowazekii* and *E. coli* homologs; 48.5% and 39.8% for PDF and MTF, respectively.

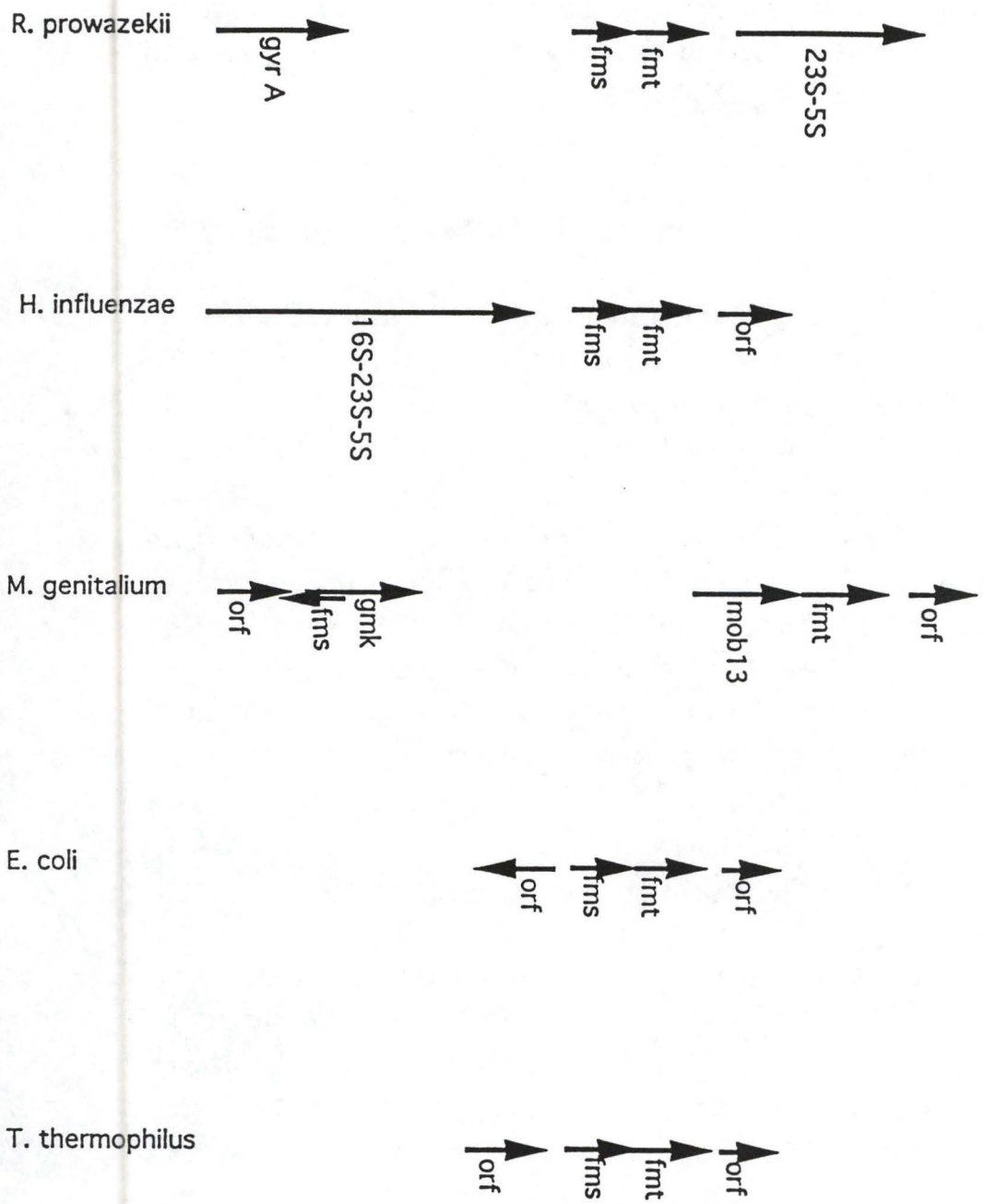


Figure 4. Organization of the region surrounding the genes *fms* and *fmt* in five bacterial species. Abbreviations used to indicate genes are as follows: *fms* = peptide deformylase gene, *fmt* = methionyl-$tRNA_f^{Met}$ formyltransferase gene, *gyr A* = gyrase A gene, 16S = rRNA small subunit, 23S = rRNA large subunit, 5S = rRNA subunit, orf = open reading frame, *gmk* = 5' guanylate kinase gene and *mob 13* = mobilisation protein gene.

The peptide deformylase from *M. genitalium* is about 41 amino acids longer at the N-ternimal end than its homologs in the other species. Interestingly, there is a Methionine codon at amino acid position 34 that if used as the intiation codon would produce a protein of approximately the same length as observed in the other species. The gene encoding PDF in *M. genitalium* is located in the opposite direction of its surrounding genes and is overlapping with as much as 269 bases at its 5'end with the *gmk* gene and 23 bases with it other neighbouring gene. The overlap with the *gmk* (5'guanylate kinase) gene is such that the third codon is shared by the two genes. The G+C content values at the first two positions are more similar to the *gmk* gene upstream of the Met-34 codon, but more similar to the *fms* gene downstream of the Met-34 codon. A translation intiation start site at codon position 34 in *M. genitalium* would produce a protein that is only 8 amino acids longer than its homologs. Most of the gaps observed in the alignment appear to be caused by the *T. thermophilus* sequence which has a few extra amino acids at a number of different positions. The region between positions 60-69 appear to be particularly well conserved among these five bacterial species.

The *fmt* gene product is of approximately the same length in all species, but with only a few positions being conserved among all of the species analysed.

3.4. Organization of the *fms-fmt* locus in bacterial genomes.

The organization of the *fms* and *fmt* genes as well as their genomic neighbourhoods is presented in Fig. 4. Here it is shown that *fms* and *fmt* are genetically linked in *R. prowazekii*, *E. coli*, *T. thermophilus* and *H. influenzae,* but that the two genes are separated from each other in *M. genitalium*.

The conservation of gene order appears to be restricted to the *fms-fmt* locus since no similarities between any of the species can be observed with respect to their flanking sequences. It is interesting to note however that the *fms-fmt* locus is located immediately downstream of the rRNA operon in *H. influenzae* but immediately upstream of the 23S-5S rRNA gene cluster in *R. prowazekii*.

4. Concluding remarks.

In this report, we have shown that in *R. prowazekii* the *fms* gene which encodes peptide deformylase (PDF) is positioned immediately upstream of the *fmt* gene which encodes methionyl-tRNAf^{Met} formyltransferase (MTF). These two enzymes are involved in the formylation and subsequent deformylation of the formyl group of the intiatior Met-tRNAf^{Met} during protein synthesis. That genes encoding functionally related enzymes have also been organized into the same operons appear to be a common theme in bacterial evolution. Colocalisation of *fms* and *fmt* has been demonstrated in organisms as different as *H. influenzae* (Fleischman et al., 1995), *E. coli* (Guillone et al., 1992) and *T. thermophilus* (Meinnel and Blanquet, 1994). Since it seems unlikely that *fms* and *fmt* should have become linked to each other independently in different bacterial lineages, the colocalization of these two genes probably represents the organization of these two genes in a bacterial ancestor. The sole exception observed so far is *M. genitalium* where *fms* and *fmt* are several hundred kb apart on the chromosome. Since the evolution of *M. genitalium* is believed to have been associated with a large number of deletion and rearrangement events, the disruption of the*fms* and *fmt* linkage probably represents a recent, derived event. Despite the fact that *R. prowazekii* similarly has a small genome size and is thought to have undergone a reductive mode of evolution as reflected in low copy numbers and rearranged operon structures for the rRNA and elongation factor genes (Andersson et al., 1995; Syvänen et al., 1995), the ancestral arrangement of *fms* and *fmt* has not been disrupted. Thus, the rearranged 23S-5S rRNA gene cluster is flanked by an operon that is highly preserved in structure.

Rearrangement of rRNA Genes in *Rickettsia* Preceeded the Divergence of the Typhus and the Spotted Fever Group Rickettsia

Magali P. C. Romedenne[1*], Diane Stothardt[2], Siv G. E. Andersson[1] and Charles G. Kurland[1]

[1]*Department of Molecular Biology, Uppsala University, Biomedical Center, Box 590, S751 24 Uppsala, Sweden*
[2] *Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA*

**Present address: Laboratory of Genetics, Facultés Universitaires Notre dame de la Paix, 61 rue de Bruxelles, 5000 Namur, Belgium*

Running title: rRNA genes in *Rickettsia*

Keywords: genomic organization, molecular evolution, *Rickettsia*

Summary

An unusual arrangement of the rRNA genes has previously been observed in *Rickettsia prowazekii*, where the 23S rRNA gene is preceeded by *fmt* which encodes methionyl-tRNAf^{Met} formyltransferase (Andersson et al., 1995). In this study we show that *fmt* is located upstream of *rrl* also in the genomes of *R. sibirica, R. parkeri, R. rickettsii, R. amblyomii* as well as of the ELB agent. These findings suggest that the genomic rearragement of the rRNA genes preceeded the divergence of the typhus and the spotted fever group rickettsia.

1. Introduction

Rickettsia are small, gram-negative bacteria that live in close association with eukaryotic cells (Weiss and Moulder, 1984). They are members of the family *Rickettsiaceae*, which includes the three tribes *Rickettsiaea*, *Ehrlichiaea* and *Wolbachia*. The genus *Rickettsia* can be further divided into the spotted fever group (SFG), the typhus group (TG) and the scrub typhus group (STG).

Members of the spotted fever group Rickettsia are distributed world-wide. These bacteria grow and divide primarily within the host cell cytoplasm, but are capable of multiplying within the cell nucleus as well. Many species are pathogens causing diseases such as for example Siberian Tick Typhus (*R. sibirica*), Rickettsialpox (*R. akari*) and Queensland Tick Typhus (*R. australis*). The typhus group Rickettsia grow and divide exclusively within the host cell cytoplasm and include species such as *R. prowazekii* and *R. typhi,* the etiological agents of epidemic and murine typhus, respectively. The scrub typhus group of Rickettsia is composed of several heterogenous strains of *R. tsutsugamushi*, the causative agent of scrub typhus. A few species can not easily be classified into either of these three categories. The ELB agent for example was first classified into the typhus group rickettsia based on the observation that it causes a disease similar to murine typhus in human populations (Azad et al., 1992; Williams et al., 1992). However, sequence comparisons of the 17 kDa antigen gene and the 16S rRNA gene have subsequently shown that this species may be more closely related to the spotted fever group rickettsia than to the typhus group rickettsia (Azad et al., 1992; Williams et al., 1992).

Features in common of all the different species of Rickettsia are their intracellular growth environment, their small genome sizes (1.2 Mb) (Eremeeva et al., 1994) and their low genomic G+C content values (29-33%). It is generally believed that the Rickettsia have evolved from free-living bacteria with larger size genomes (Andersson and Kurland, 1995). From this perspective it seems plausible that the single copy number of the rRNA genes and the disruption of the traditional rRNA operon structure as observed in *R. prowazekii* is the result of a degenerate mode of evolution (Andersson et al., 1995). In this study, we examine the genus *Rickettsia* for the occurrence and

distribution of the unusual arrangment of rRNA genes previously observed in *R. prowazekii* (Andersson et al., 1995). We report here the sequence of the intergenic region between *fmt* and the 23S rRNA gene in six different Rickettsia species.

2. Materials and methods

2.1. PCR amplification and purification

The polymerase chain reaction (PCR) amplification were carried out in 100 µl reaction mixture containing 20 nmoles of dATP, dCTP, dGTP and dTTP, 100 pmoles each of indicated primer, 40 to 100 ng of chromosomal DNA, 2.5 units of Taq polymerase (Promega Biotech, Madison, WI) and 10 µl 10X buffer (0.5M Tris-HCl pH 8.8, 0.15M $(NH_4)_2SO_4$, 15mM $MgCl_2$, 1% Triton X-100 and 0.1% gelatin). After a hot start, 30 cycles of amplification was performed in a DNA thermal cycler (AMS Biotechnology, Taby, Sweden), with a standard profile of denaturation at 94°C for 1min, annealing at 53°C for 1 min, and extension at 72°C for 1 or 2 min. Following amplification, 5 to 10 µl of the amplified material was fractionated by electrophoresis on 1% agarose gels.

The PCR products were purified by using Wizard PCR Preps DNA purification System (Promega). The concentration of the final product was measured by running a 5 µl on a 1% agarose gel side by side with lanmbda DNA of known concentration.

2.2. DNA sequencing

Thermo Squenase fluorescent labelled primer cycle sequencing kit (Amersham) which used dideoxy chain termination method was used to sequence PCR products. Cycle sequencing uses repeat cycles of thermal denaturation, annealing and elongation/termination.
For the DNA sequencing we used Thermo sequenase (Amersham) and fluorescent primers.

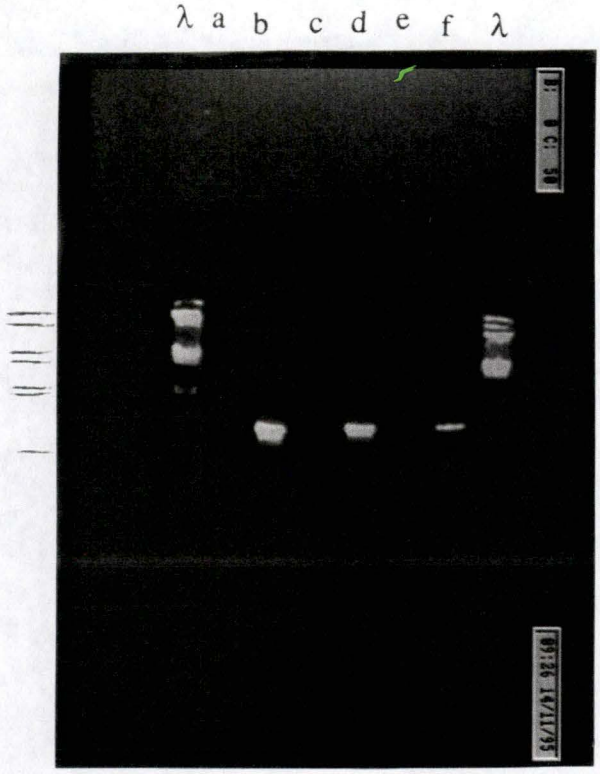


Figure 1a: PCR analysis. Primers ot56+2 and ot56-2 with genomic DNA from *R. prowazekii* (lane b), *R. sibirica* (lane d), *R. parkerii* (lane f). Lanes a, c and e are water controls. Molecular marker were based on digests of lambda DNA with the single enzyme *Hind* III, producing the following sizes of bands in kb: 23.1, 9.4, 6.6, 4.4, 2.3, 2.0 and 0.6.

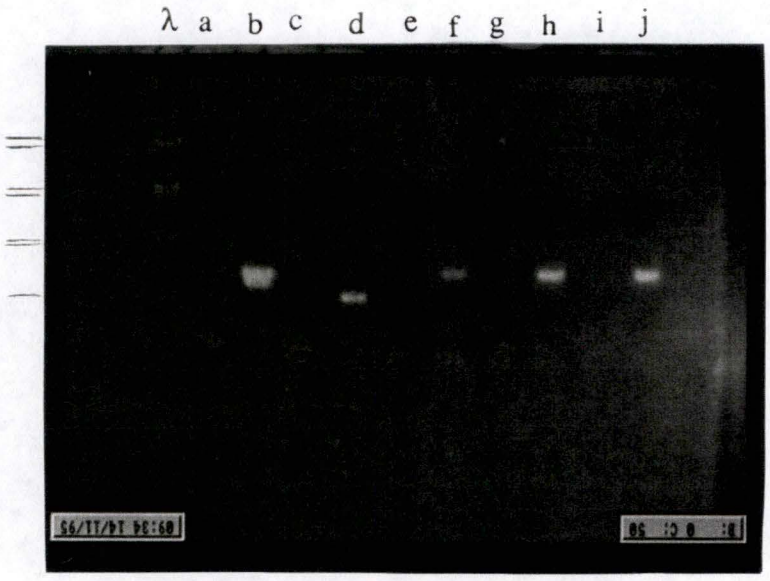


Figure 1b: PCR analysis. Primers ot56+2 and ot56-2 with genomic DNA from *R. prowazekii* (lane b), ELB agent (lane d), *R. rickettsii* (lane f), *R. montana* (lane h) and *R. amblyomii* (lane j). Lanes a, c, e, g and i are water controls. Molecular marker were based on digests of lambda DNA with the single enzyme *Hind* III, producing the following sizes of bands in kb: 23.1, 9.4, 6.6, 4.4, 2.3, 2.0 and 0.6.

The product of the sequencing reaction were separated and analysed with the aid of an ALF (Automated Laser Fluorescent) DNA sequencer (Pharmacia, Uppsala, Sweden).

2.3. Sequence analysis

The aligment of the *fmt* gene and the spacer were performed by using CLUSTAL W (Higgins et al., 1992) as well as the tree deduced following the Neighbor-joining method.

The Ks and Ka were calculated with the Li's method (Li, W-H., 1993)

3. Results

3.1. PCR amplifications with primers from *fmt* and 23S rRNA

To study the genomic neighbourhoods of the *fmt* gene and the 23S rRNA gene we performed PCR reactions using one primer from within the *fmt* gene, ot56+2 (CTAAAGCAGAAGGAAAAATT) and one primer from within the *rrl* gene, ot56-2 (GCTTCTAGTGCCAAGGCATC). This primer pair yielded a product of 1.0 kb in size when genomic DNA from *R. prowazekii* was used to drive the PCR reaction (Fig. 1), as previously observed (Andersson et al., 1995). Successful PCR amplifications were also obtained when these primers were used in PCR reactions together with genomic DNA from the following species: *R. sibirica, R. parkeri, R. rickettsii, R. montana, R. ablyommii* and the ELB agent (Fig. 1). This suggests that the *fmt* and the 23S rRNA genes are near to each other in all of these genomes. Some size variation of the amplified fragments was observed, with genomic DNA from the ELB agent producing a PCR product of only 500 bp in length.

		1	2	3	4	5	6	7	8
1	P1;Rsib fmt	0.0	1.3	0.5	1.0	1.3	1.6	8.9	6.6
2	P1;Rrick fmt	5.0	0.0	0.7	1.3	1.5	1.8	9.2	6.9
3	P1;Rpark fmt	1.3	5.0	0.0	0.5	0.7	1.0	8.4	6.0
4	P1;Ramb fmt	0.0	5.0	1.3	0.0	1.3	1.6	8.1	5.8
5	P1;ELB fmt	3.9	9.1	5.3	3.9	0.0	1.8	7.5	5.2
6	P1;Rmont fmt	1.3	6.1	2.6	1.3	5.2	0.0	9.5	7.2
7	P1;Rp fmt	26.5	22.6	26.5	26.5	33.1	27.6	0.0	4.2
8	P1;Rty fmt	23.2	20.5	23.2	23.2	29.2	25.3	11.7	0.0

Table 1: Evolutionary distances based on the *fmt* nucleotide sequences. Substitutions at synonymous sites (Ka) are shown above the diagonal. Substitutions at synonymous sites (Ks) are shown below the diagonal.

```
.b_fmt          YFSYNDKIIKILEAEYLNADHHCTSGTVISDKLEIACGSGILRVKKLQQESKKALNIEEF
.ck_fmt         YFSYNDKIIKILEAEYLNADHHFTSGTVISDKLEIACGSGILRVKKLQQESKKALSIEEF
ırk_fmt         YFSYNDKIIKILEAEYLNADHHFTSGTVISDKLEIACGSGILRVKKLQQESKKALNIEEF
ıb_fmt          YFSYNDKIIKILEAEYLNADHHFTAGTVISDKLEIACGSGILRVKKLQQESKKALNIEEF
ı_fmt           YFSYNDKIIKILEAEYLNADHHFTSGTVISDKLEIACGSGILRVKKLQQESKKALNIEEF
ınt_fmt         YFSYNDKIIKILEAEYLNAAHHFTSGTLISDKLEIACGSGILRVKKLQQESKKALNIEEF
ow_fmt          YFSYNDKIIKILRAEYFNYNHHFIPGTVINNKLEIACGSGILRVKKLQQESKKALNIEEF
r_fmt           YFSYNDKIIKILEAEYLNYNHHFIPGTVISNKLEIACGSGILRVTKLQQESKKALNIEAF
                ************ *** *  **   **.*  ************* ********** ** *

.b_fmt          LRGTNILKDTVLKZ
.ck_fmt         LRGTNILKDTVLKZ
ırk_fmt         LRGTNILKDTVLKZ
ıb_fmt          LRGTNILKDTVLKZ
ı_fmt           LRGTNIL-DTILKZ
ınt_fmt         LRGTNILKDTVLKZ
ow_fmt          LRGTNILKDTILKZ
r_fmt           LRGTNILKDTILKZ
                ******* **.***
```

Figure 2. Multiple alignment of the deduced amino acid sequences of the *fmt* gene in eight Rickettsia species. Rsib = *R. sibirica* ; Rrick = *R. rickettsii* ; Rpark = R. *parkeri* ; Ramb = *R. amblyomii* ; ELB = *ELB agent* ; Rmont = *R. montana* ; Rprow = *R. prowazekii* ; Rty = *R. typhi*. Symbols beneath the aligned sequences indicate identical residues (*) and sites with conservative replacements (.).

3.2. Comparative nucleotide sequence analysis

The identity of the PCR products obtained were verified by a complete sequence analysis. The sequence data obtained corresponded to 219 basepairs of the 3´end of the *fmt* gene in addition to a spacer region of 303 to 886 base pairs ending 3 base pairs upstream of the 23S rRNA gene. The length of the 3´end of the *fmt* gene was found to be identical in all species, with the exception of the ELB agent in which 3 bases had been deleted. The deletion was found to correspond to exactly one codon and was therefore not associated with any shift in reading frame.

The alignment of the amino acid sequences of the partial *fmt* gene products is presented in Fig. 2. Based on this alignment we have calculated the number of synonymous (K_S) and nonsynonymous (K_a) nucleotide substitutions in the *fmt* gene among the eight species analysed (Table 1). As expected, for most pairs the K_S values are higher than the K_a values. The observed substitution frequencies clearly show that the two members of the typhus group rickettsia (*R. prowazekii* and *R. typhi*) are distinct from the members of the spotted fever group rickettsia. On the average, the nonsynonymous and synonymous substitution frequencies are 7.4% and 25.6%, respectively, between the two groups of species. For comparison, it can be noted that the synonymous substitution frequencies are less than 5% among the different species within the spotted fever group rickettsia. The substitution frequencies observed for the ELB agent suggest that it is more closely related to the spotted fever group rickettsia ($K_S < 0.09$, $K_a < 0.02$) than to the typhus group rickettsia ($K_S > 0.29$, $K_a > 0.05$).

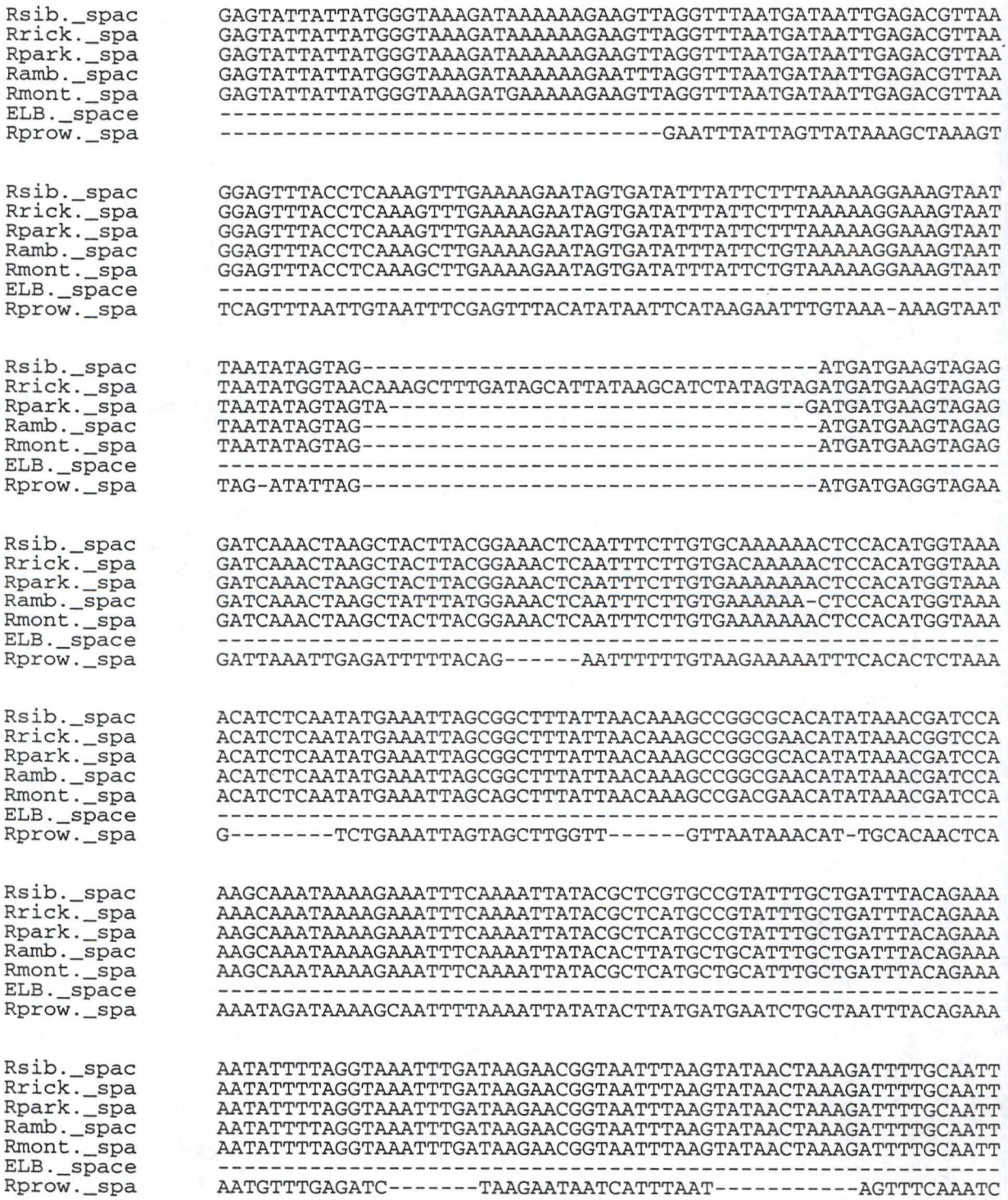
The lenghts and the G+C content values of the spacer regions between *fmt* and *rrl* are presented in Table 2. In *R. prowazekii* the G+C content of this spacer sequence is 22.5%, close to an average value of 21.4% calculated from a large number of *R. prowazekii* intergenic regions (Andersson and Sharp, 1995). This suggests that the region in between *fmt* and *rrl* corresponds to noncoding DNA, as also supported by the lack of homologous sequences in the databanks. The other six species analysed were found to have slightly higher G+C content values ranging from 26.4% (ELB agent) to 28.7% (*R. montanta*). The three

Species	Length (bp)	G+C content (%)
SFG		
R. sibirica	854	28.5
R. ricketsii	886	28.7
R. parkii	857	28.1
R. ambirica	844	27.8
R. montana	845	27.8
TG		
ELB-agent	303	26.4
R. prowazekii	716	22.5

Table 2: Size and G+C content of the spacer region in seven Rickettsia species.

1	vs.	2:	DIST =	0.0235;	p =	0.0176;	q =	0.0059;	length =	851
1	vs.	3:	DIST =	0.0082;	p =	0.0070;	q =	0.0012;	length =	854
1	vs.	4:	DIST =	0.0308;	p =	0.0225;	q =	0.0083;	length =	844
1	vs.	5:	DIST =	0.0284;	p =	0.0213;	q =	0.0071;	length =	845
1	vs.	6:	DIST =	0.0660;	p =	0.0528;	q =	0.0132;	length =	303
1	vs.	7:	DIST =	0.2420;	p =	0.1399;	q =	0.1021;	length =	715
2	vs.	3:	DIST =	0.0234;	p =	0.0187;	q =	0.0047;	length =	854
2	vs.	4:	DIST =	0.0380;	p =	0.0285;	q =	0.0095;	length =	841
2	vs.	5:	DIST =	0.0380;	p =	0.0297;	q =	0.0083;	length =	842
2	vs.	6:	DIST =	0.0627;	p =	0.0462;	q =	0.0165;	length =	303
2	vs.	7:	DIST =	0.2416;	p =	0.1404;	q =	0.1011;	length =	712
3	vs.	4:	DIST =	0.0296;	p =	0.0225;	q =	0.0071;	length =	844
3	vs.	5:	DIST =	0.0296;	p =	0.0237;	q =	0.0059;	length =	845
3	vs.	6:	DIST =	0.0660;	p =	0.0528;	q =	0.0132;	length =	303
3	vs.	7:	DIST =	0.2420;	p =	0.1413;	q =	0.1007;	length =	715
4	vs.	5:	DIST =	0.0178;	p =	0.0154;	q =	0.0024;	length =	843
4	vs.	6:	DIST =	0.0596;	p =	0.0464;	q =	0.0132;	length =	302
4	vs.	7:	DIST =	0.2325;	p =	0.1317;	q =	0.1008;	length =	714
5	vs.	6:	DIST =	0.0660;	p =	0.0528;	q =	0.0132;	length =	303
5	vs.	7:	DIST =	0.2392;	p =	0.1371;	q =	0.1021;	length =	715
6	vs.	7:	DIST =	0.1294;	p =	0.0734;	q =	0.0559;	length =	286

Table 3: Evolutionary distances based on the spacer sequence between *fmt* and *rrl*. Abbreviations are as followed: 1 = *R. sibirica*, 2 = *R. ricketsii*, 3 = *R. parkeri*, 4 = *R. amblyommi*, 5 = *R. montana*, 6 = *ELB agent,* 7 = *R. prowazekii*, DIST = percentage of divergence, p = rate of transition (A<->G; C<->T), q = rate of tranversion, Length = number of sites used in comparison.


Rsib._spac GAGTATTATTATGGGTAAAGATAAAAAAGAAGTTAGGTTTAATGATAATTGAGACGTTAA
Rrick._spa GAGTATTATTATGGGTAAAGATAAAAAAGAAGTTAGGTTTAATGATAATTGAGACGTTAA
Rpark._spa GAGTATTATTATGGGTAAAGATAAAAAAGAAGTTAGGTTTAATGATAATTGAGACGTTAA
Ramb._spac GAGTATTATTATGGGTAAAGATAAAAAAGAATTTAGGTTTAATGATAATTGAGACGTTAA
Rmont._spa GAGTATTATTATGGGTAAAGATGAAAAAGAAGTTAGGTTTAATGATAATTGAGACGTTAA
ELB._space
Rprow._spa GAATTTATTAGTTATAAAGCTAAAGT
Rsib._spac GGAGTTTACCTCAAAGTTTGAAAAGAATAGTGATATTTATTCTTTAAAAAGGAAAGTAAT
Rrick._spa GGAGTTTACCTCAAAGTTTGAAAAGAATAGTGATATTTATTCTTTAAAAAGGAAAGTAAT
Rpark._spa GGAGTTTACCTCAAAGTTTGAAAAGAATAGTGATATTTATTCTTTAAAAAGGAAAGTAAT
Ramb._spac GGAGTTTACCTCAAAGCTTGAAAAGAATAGTGATATTTATTCTGTAAAAAGGAAAGTAAT
Rmont._spa GGAGTTTACCTCAAAGCTTGAAAAGAATAGTGATATTTATTCTGTAAAAAGGAAAGTAAT
ELB._space
Rprow._spa TCAGTTTAATTGTAATTTCGAGTTTACATATAATTCATAAGAATTTGTAAA-AAAGTAAT
Rsib._spac TAATATAGTAG-----ATGATGAAGTAGAG
Rrick._spa TAATATGGTAACAAAGCTTTGATAGCATTATAAGCATCTATAGTAGATGATGAAGTAGAG
Rpark._spa TAATATAGTAGTA-----GATGATGAAGTAGAG
Ramb._spac TAATATAGTAG-----ATGATGAAGTAGAG
Rmont._spa TAATATAGTAG-----ATGATGAAGTAGAG
ELB._space
Rprow._spa TAG-ATATTAG-----ATGATGAGGTAGAA
Rsib._spac GATCAAACTAAGCTACTTACGGAAACTCAATTTCTTGTGCAAAAAACTCCACATGGTAAA
Rrick._spa GATCAAACTAAGCTACTTACGGAAACTCAATTTCTTGTGACAAAAACTCCACATGGTAAA
Rpark._spa GATCAAACTAAGCTACTTACGGAAACTCAATTTCTTGTGAAAAAAACTCCACATGGTAAA
Ramb._spac GATCAAACTAAGCTATTTATGGAAACTCAATTTCTTGTGAAAAAA-CTCCACATGGTAAA
Rmont._spa GATCAAACTAAGCTACTTACGGAAACTCAATTTCTTGTGAAAAAAACTCCACATGGTAAA
ELB._space
Rprow._spa GATTAAATTGAGATTTTTACAG------AATTTTTTGTAAGAAAAATTTCACACTCTAAA
Rsib._spac ACATCTCAATATGAAATTAGCGGCTTTATTAACAAAGCCGGCGCACATATAAACGATCCA
Rrick._spa ACATCTCAATATGAAATTAGCGGCTTTATTAACAAAGCCGGCGAACATATAAACGGTCCA
Rpark._spa ACATCTCAATATGAAATTAGCGGCTTTATTAACAAAGCCGGCGCACATATAAACGATCCA
Ramb._spac ACATCTCAATATGAAATTAGCGGCTTTATTAACAAAGCCGGCGAACATATAAACGATCCA
Rmont._spa ACATCTCAATATGAAATTAGCAGCTTTATTAACAAAGCCGACGAACATATAAACGATCCA
ELB._space
Rprow._spa G--------TCTGAAATTAGTAGCTTGGTT------GTTAATAAACAT-TGCACAACTCA
Rsib._spac AAGCAAATAAAAGAAATTTCAAAATTATACGCTCGTGCCGTATTTGCTGATTTACAGAAA
Rrick._spa AAACAAATAAAAGAAATTTCAAAATTATACGCTCATGCCGTATTTGCTGATTTACAGAAA
Rpark._spa AAGCAAATAAAAGAAATTTCAAAATTATACGCTCATGCCGTATTTGCTGATTTACAGAAA
Ramb._spac AAGCAAATAAAAGAAATTTCAAAATTATACACTTATGCTGCATTTGCTGATTTACAGAAA
Rmont._spa AAGCAAATAAAAGAAATTTCAAAATTATACGCTCATGCTGCATTTGCTGATTTACAGAAA
ELB._space
Rprow._spa AAATAGATAAAAGCAATTTTAAAATTATATACTTATGATGAATCTGCTAATTTACAGAAA
Rsib._spac AATATTTTAGGTAAATTTGATAAGAACGGTAATTTAAGTATAACTAAAGATTTTGCAATT
Rrick._spa AATATTTTAGGTAAATTTGATAAGAACGGTAATTTAAGTATAACTAAAGATTTTGCAATT
Rpark._spa AATATTTTAGGTAAATTTGATAAGAACGGTAATTTAAGTATAACTAAAGATTTTGCAATT
Ramb._spac AATATTTTAGGTAAATTTGATAAGAACGGTAATTTAAGTATAACTAAAGATTTTGCAATT
Rmont._spa AATATTTTAGGTAAATTTGATAAGAACGGTAATTTAAGTATAACTAAAGATTTTGCAATT
ELB._space
Rprow._spa AATGTTTGAGATC-------TAAGAATAATCATTTAAT----------AGTTTCAAATC

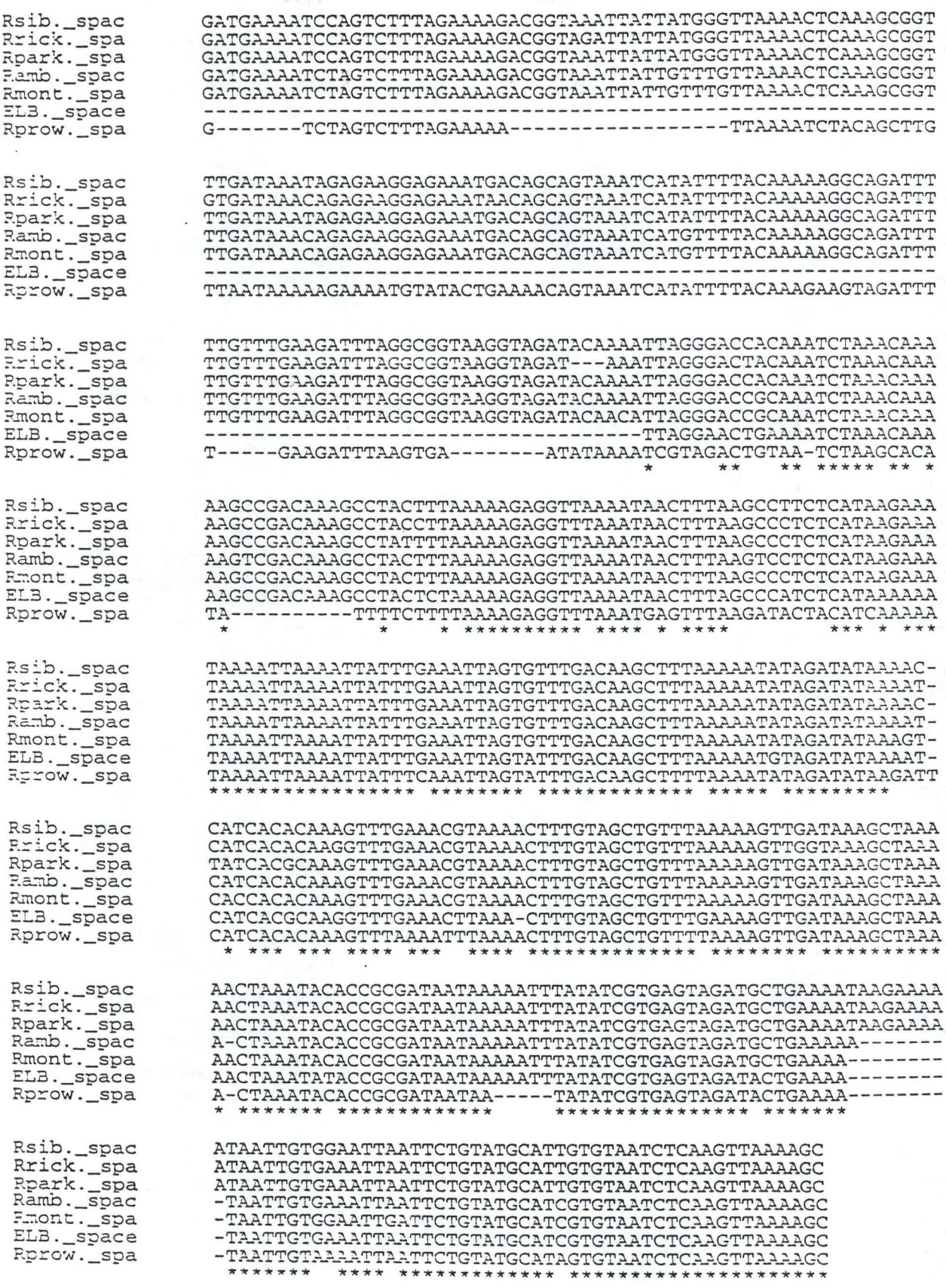


Figure 3: Multiple alignment of the nucleotide sequences of the spacer regions between *fmt* and *rrl* in seven Rickettsia species. Rsib = *R. sibirica* ; Rrick = *R. rickettsii* ; Rpark = R. *parkerii* ; Ramb = *R. amblyomii* ; ELB = *ELB agent* ; Rmont = *R. montana* ; Rprow = *R. prowazekii* . Symbols beneath the aligned sequences indicate identical residues (*).

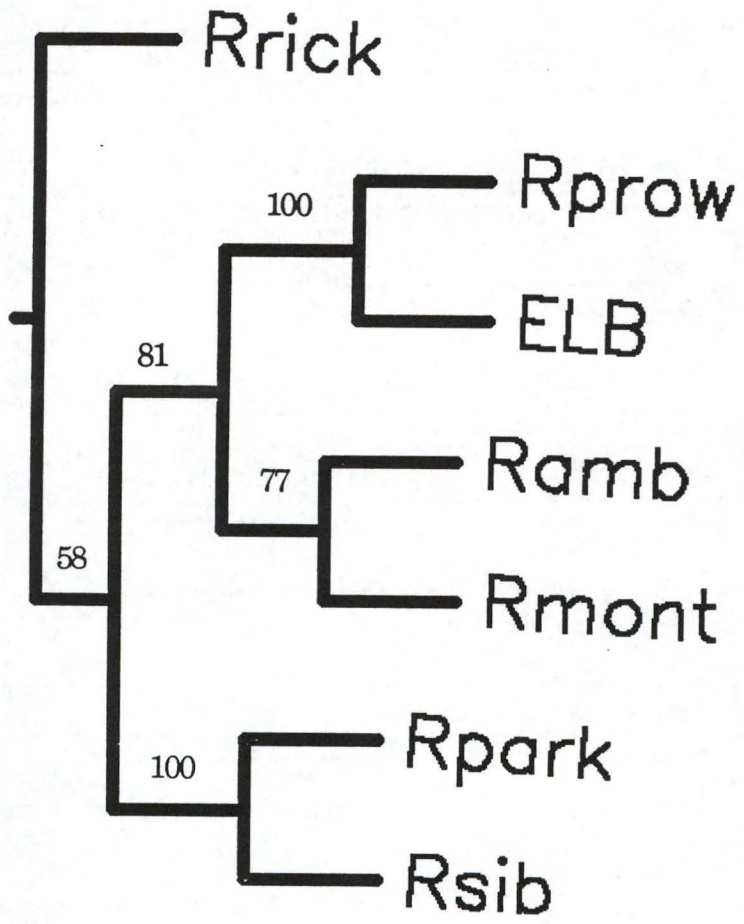


Figure 4: Relationship among the Rickettsia inferred from the nucleotide sequence of the spacer region between *fmt* and *rrl*. The phylogeny was estimated by the neighbor-joining method applied to pairwise sequence differences corrected for multiple replacements. The number at each node indicates the percentage of bootstraps in which the taxa to the right were clustered.

most closely related spotted fever group rickettsia, *R. sibirica, R. rickettsii* and *R. parkerii* were observed to have virtually identical G+C content values within the spacer regions (from 28.1% to 28.7%).

The alignments of the spacer sequences are presented in Fig. 3. Here it can be seen that the reduced size of the PCR product obtained with genomic DNA from the ELB agent can be entirely explained by a contiguous deletion of almost 500 base pairs within the spacer region. The deletion extends from the region immediately next to the *fmt* termination codon to 500 base pairs into the spacer region. The 303 base pairs of the remaining spacer region show strong homology to the other *Rickettsia* sequences (Table 3). The corrected nucleotide distance values among the different species of the spotted fever group rickettsia are very small, less than 5%, in accordance with the low K_S values obtained within this group of species (<5%). An average distance value of 23.9% is observed between *R. prowazekii* and the five representatives of the spotted fever group rickettsia (*R. parkerii*, *R. sibirica*, *R. montana* and *R. amblyomii*). This value is similar to the average substitution frequency at synonymous sites for the same pairs of species (25.6%).

Phylogenetic trees based on these distance values are presented in Fig. 4. Neighbour joining as well as parsimony analysis produce three main clusters; *R. prowazekii* and the ELB agent, *R. parkerii* and *R. sibirica*, *R. montana* and *R. amblyommii*.The low bootstap values associated with *R. rickettsii* suggests that the branching pattern of this specie can not be determined with certainty in this analysis.

4. Concluding remarks

It has been shown previously that the *fmt* gene encoding methionyl-$tRNA_f^{Met}$ formyltransferase is situated upstream of the 23S-5S rRNA cluster in *R. prowazekii* (Andersson et al., 1995). Here we have shown that the same organization is found in *R. sibirica*, *R. ricketsii, R. parkeri, R. amblyommi, R. montana* and ELB agent.
Analyses of the alignment of the *fmt* gene and of the spacer between *fmt* and 23S rRNA genes show a large difference between the species from the spotted fever group (SFG) (*R. sibirica, R. ricketsii, R. parkeri,*

R. amblyommi and R. montana) and the species from the typhus group (TG) (*R. prowazekii*). Small difference among the spotted fever group has been found.

The ELB agent which was classified into the typhus group because of its disease similar to murine typhus, seems to be closer to the spotted fever group. Anyway the phylogenetic tree deduced from distance values produce a cluster with ELB agent together with *R. prowazekii* (TG). Because of the small number of positions being compared (the length of the ELB agent is 303bp), it is difficult to evaluate the significance of the tree obtained. Analyses of longer spacers or genes may help to deduce a phylogenetic tree of the Rickettsia genus. Studies of the other species from the Rickettsia genus may also be helpful specially for *R. typhi* which is from the typhus group and for *R. akari* and *R. australis* whish are close to ELB agent (the three bacteria are often called AAE group).

Since the rearrangement of the rRNA operon is conserved in the seven species, it is believed that the reorganization of the 16S-23S-5S gene cluster has occured before the diversification of the genus Rickettsia.

General conclusion

Deletions and recombination events between duplicated genetic sequences lead to genomic shrinkage as well as to rearrangements of genomic neighborhoods (Krawiec and Riley, 1990). Genomic rearrangment are frequent events in bacteria, but free living bacterial populations are quite resistant to the inroads of these variations. In contrast, mitochondrial populations have quite small and highly variable genome (Gray, 1992). We expect the intracellular lifestyle to provide an evolutionary setting within wich the genomes of endoparasitic bacteria would tend to have characteristics that are intermediate between those of free living bacteria and organelles.

Deletions and rearrangements of the ribosomal RNA genes in the Rickettsia genus were to be expected because genomic rearrangements are often associated with repeated sequences, which serve as sites of intrachromosomal homologous recombination.

It has been shown in this work that the highly preserved arrangment of the *fms* gene which encodes peptide deformylase (PDF) and the *fmt* gene encoding methionyl-tRNAf^{Met} formyltransferase (MTF) has been found immediately upstream of the rearranged 23S-5S rRNA gene cluster.

Futhermore, we show that the unusual organisation of the rRNA genes are found in six other species of the Rickettsia genus.

All these observations suggest that the rearrangement of rRNA genes in Rickettsia preceeded the divergence of the typhus group and the spotted fever group.

Materials and methods

1. Preparation of Plasmids

(QIAGEN Plasmid Kit) (KEBO, Stockholm, Sweden)

1.1. Introduction

The Clone 132 sequenced was one of thousand clones selected randomly for universal and reverse sequencing from lambda Zap library.
The lambda Zap II phage is disigned with pBluescript plasmid DNA, flanking the foreign insert in the phage linear genome. E. coli is infected by lambda Zap II phage and by a helper phage. The helper phage recognises the flanking region of the insert in the lambda DNA. The single strand phagemid circularizes and contains the clone 132. Then the phagemid infect another E. coli cells (XLI) and the cells containing the double strand plasmid will be ampiciline resistant. The plasmid is stored at -80°C in LB with 30% of glycerol.

The clone 132 which contains the *fms* gene was produced and purified by the folowing protocol.

1.2. Components

Resuspension Buffer : 100 ug/ml RNase A
50 mM Tris-HCl
10 mM EDTA (pH 8)

Lysis Buffer: 200 mM NaOH

1% SDS

Neutralisation Buffer : 3.0 M KAc (pH 5.5)

Equilibration Buffer : 750 mM NaCl
50 mM MOPS
15% ethanol (pH 7.0)
0.15% Triton X-100

Wash Buffer : 1.0 M NaCl
50 mM MOPS
15% ethanol (pH 7.0)

Elution Buffer: 1.25 M NaCl
50 mM Tris-HCl
15% ethanol (pH 8.5)

TE : 10 mM Tris-HCl
1mM EDTA (pH 8.0)

5x LB Broth (per liter) : 500 g NaCl
(Stock solution) 50 g Bacto-Tryptone
25 g Yeast Extract

1 x LB (working solution) : 200 ml 5x LB
800 ml dd H2O
10 ml 20% glucose
1 ml Ampiciline (100ug/ml)

LB/Ampiciline plate: 14g agar
1 liter 1x LB

1.3. Protocol

Streak transformes cell within a LB/Ampiciline plate and incubate overnight at 37°C. Colonies who appear on the plate, contain the double stranded plasmid with the cloned DNA insert.

Take a single isolated colony and started a culture in 20 ml of working solution. When the mixture becomes cloudy, start an overnight culture with 2ml of the mixture and 200ml of 1x LB.

Centrifuge at 4,000 RPM during 10 minutes. Add 10 ml of Resuspension Buffer to the bacterial pellet and 10 ml of Lysis Buffer. Mix gently.

After incubation at room temperature for 5 minutes, the lysis should occur, and the solution appears viscous. Add 10 ml of Equilibration Buffer. Place the solution on ice for 20 minutes. Mix the solution and centrifuge it at 4°C for 30 minutes at 10,000 RPM.

The supernatant should be clear after centrifugation. If not a second centrifugation is performed.

Add 10 ml of Equilibration Buffer to a Qiagen-tip 500. Let the liquid going out without applying pressure. Apply the clear supernatant onto the column, let the liquid enter the resin by gravity flow.

Wash two times with 30 ml of Wash Buffer .

Elute the DNA by adding 15 ml of Elution Buffer and collect it in 30 ml glass tube (corex).

Add 0.7 volumes of isopropanol for precipitation of DNA. Centrifuge directly at 10,000 RPM for 30 minutes at 4°C. Remove carefully the supernatant.

Wash carefully the DNA pellet with 15 ml of cold 70% ethanol. Discard the ethanol and let dry for 5 minutes and dissolve the pellet in a small volume of TE.

The DNA concentration was measured by running it and a aliquot of lambda DNA of known concentration on a 1% agarose gel.

2. Polymerase Chain Reaction (PCR)

2.1. Introduction

The goal of this technique is to amplify short segments between two defined primers.
First the DNA is denatured. Two short primers (around 20 bases) are annealed with the two single-stranded DNA, at a complementary site.
The enzyme Taq DNA polymerase is used to synthesize a complementary single-strand from the 3´end of each primer.
The cycle of 3 steps is repeated 20 to 30 times to have an increasing of the number of copies.

PCR is required to amplify the spacer between *fmt* gene and 23S gene for *R. prowazakii, R. ricketsii, R. sibirica, R. montana, R. parkii, ELB-agent and R. ambilyommi.*
We use two primers, one from the *fmt* gene (5´-CTAAAGCAGAAGGAAAAATT-3´) (position 575) and one primer from the beginning of the the 23S gene (5´-GCTTCTAGTGCCAAGGCATC-3´).
These segments will be sequenced later.

A PCR of *R. prowazaki* genomic DNA was also performed in order to amplify the locus containing both *fms* and *fmt* genes.
The amplification by PCR is using two primers : one from the *fms* gene (5´-ATTAGCAAGATTTACTTTCTT-3´) and the second one from the *fmt* gene (5´-CACTGCCACAAGCGATTTCTAGTTT-3´).

2.2. Components

Taq DNA Polymerase in Storage Buffer containing :

50% glycerol
50mM Tris-HCl (pH8.0)
100mM NaCl
0.1mM EDTA
1mM DTT
1.0% Triton X-100

10X Thermophilic Buffer (Magnesium Free) :
500mM KCl
100mM Tris-HCl (pH 9.0)
1%Triton X-100

Magnesium Chloride 25mM

dNTP 2.5 mM : 5 µl of 100 mM of dATP
5 µl of 100 mM of dCTP
5 µl of 100 mM of dGTP
5 µl of 100 mM of dTTP
180 µl H2O

2.3. Protocol

The reaction buffer was: 135 µl of sterilised water
25 µl of 10 x Thermophilic Buffer
15 µl of MgCl2
25 µl of dNTP (2.5 mM)
12.5 µl of each primers (25 pmoles).

The total reaction volume was 225 ul.

Two aliquots of 90 µl of this reaction are prepared in two tubes. Add in one tube 5 µl of water as a control and add 5 µl of genomic DNA in the second tube.
Add one drop of mineral oil.

Hot Start at 94° C for two minutes, then 80° C for three minutes. Add 5 µl of diluted Taq DNA polymerase (2.5 units).

The cycle started at 94° C for one minute, followed by one minute at 53° C and finished at 72° C for three minutes for 30 cycles.
The program finished by ten minutes at 72° C.

2.4. Control

A control of the PCR product on a 1% agarose gel was performed to observe eventual contaminations and to estimate the yield of the PCR.

Components :

5 x TBE (stock solution): 54 g Tris base
27.7 g boric acid
20 ml 0.5 M EDTA (pH 8)

0.5 M EDTA (pH 8) : 35 g of disodium ethylen diamine tetraacetate
160 ml H_2O Adjust to pH to 8
Adjust the volume to 200 ml

The components of the agarose gel 1% :
0.5 g of agarose
50 ml 0.5 x TBE
5 µl EtBr (ethidiumbromide)

10 µl of each reaction were loaded together with 2 µl of dye.
The samples were electrophoresed at 100 volts for thirty minutes.

3. Direct Purification from PCR Products
(Wizard PCP Preps DNA purification System) (Promega)

3.1. Introduction

Before sequencing, it is important to purify the PCR products to remove the mineral oil and the excess of components. It is necessary to remove all the oil before charging the minicolumn. After washing the minicolumn, the DNA was collected in TE.

3.2. Components

Direct Purification Buffer: 50 mM KCl
10 mM Tris-HCl (pH 8.8)
1.5 mM MgCl2

Magic PCR Preps Resin (Promega)

Magic Minicolumn (Promega)

TE buffer : 10 mM Tris-HCl (pH7.5)
1 mM EDTA

Chloroform 100%

Isopropanol 80%

3.3. Protocol

Add 100 µl of water and 50 µl of chloroform to each completed PCR reaction. Mix carefully and spin one minute at 12,000 RPM.

Remove 150 µl from the upper layer of the sample.
Add 100 µl Direct Purification Buffer, vortex.
Add 1 ml of magic PCR Preps resin and vortex briefly 3 times over a one minute period.

Prepare one Magic Minicolumn for each PCR product and connect it to a 2 ml syringe.

Apply the DNA mix into the barrel and push slowly the plunger, after that, place 2 ml of isopropanol 80% and push gently the liquid out to wash the minicolumn.

Remove the syringe and place the minicolumn into a 1.5 ml microcentrifuge tube. Spin the minicolumn for twenty seconds at 12,000 RPM to remove all the liquid out.

Let the minicolumn for five minutes at room temperature to dry the resin.

Add 50 µl of warm TE buffer (or water) to the Magic Minicolumn and wait ten minutes. Centrifuge the minicolumn for twenty seconds at 12,000 RPM to elute the DNA fragment.

4. DNA sequencing

4.1. Introduction

The Automated Laser Fluorescent ALF DNA sequencer from Pharmacia LKB has been conceived in accordance with Sanger´s method (chain termination technique).
Sequencing requires hybridization of the single stranded DNA to be sequenced with a short complementary primer.
The DNA polymerase starts at the 3´ hydroxyl end of the primer and synthesises a complementary copy of the single strand DNA.

DNA sequencing by enzymatic synthesis employs four different reaction mixtures, each contains all four deoxyribonucleoside triphosphates (dNTP) and one of the four dideoxyribonucleoside triphosphate (ddNTP). The ddNTP has no 3´ hydroxyl group and so, when incorporated, causes the chain termination.
When the sequencing reaction is fractionated by electrophoresis on gel, the pattern of bands shows the distribution of synthesised DNA.

The use of fluorescent primers or fluore dATP are required for the analyse of the fluorescent curves by the A.L.F. DNA sequencer which will translate them into a DNA sequence, exported as a computer file.

4.2. Auto Read Sequencing kit (Pharmacia)

Auto Read Sequencing kit is designed for use with Automated Laser Fluorescent ALF DNA sequencer from Pharmacia LKB.
This method is to sequence a clone by using fluorescent M13 Universal Primer and fluorescent M13 Reverse Primer or primers no fluorescent in combination with Fluore-dATP Labelling mix.

4.2.1. Sequencing with fluorescent primer

The reversal and universal fluorescent primers start elongation at specific priming sites allowing sequencing of both strands.

4.2.1.1. Components

A mix : 5 uM ddATP, 1 mM dATP, 1 mM dCTP, 1 mM dTTP, 1 mM c7dGTP, 50 mM NaCl and 40 mM Tris-HCl (pH 7.6)

C mix : 5 uM ddCTP, 1 mM dATP, 1 mM dCTP, 1 mM dTTP, 1 mM c7dGTP, 50 mM NaCl and 40 mM Tris-HCl (pH 7.6)

G mix : 5 uM ddGTP, 1 mM dATP, 1 mM dCTP, 1 mM dTTP, 1 mM c7dGTP, 50 mM NaCl and 40 mM Tris-HCl (pH 7.6)

T mix : 5 uM ddTTP, 1 mM dATP, 1 mM dCTP, 1 mM dTTP, 1 mM c7dGTP, 50 mM NaCl and 40 mM Tris-HCl (pH 7.6)

T7 DNA polymerase : 8 units/ul in 25 mM Tris-HCl (pH 7.5), 0.25 M NaCl, 5 mM DTT and 50% glycerol

M13 Universal Primer : 5´-fluorescein-d(CGACGTTGTAAAACG ACGGCCAGT)-3´ in aqueous solution, 1.5 uM (2.1 pmol/ul; 0.55 A_{260} unit/ml)

M13 Reverse Primer : 5´-fluorescein-d(CAGGAAACAGCT ATGAC)-3´ in aqueous solution, 2.1 uM (2.1 pmol/ul; 0.42 A_{260} unit/ml)

Annealing Buffer : 1 M Tris-HCl (pH 7.6) and 100 mM $MgCl_2$

Extension Buffer : 304 mM citric acid, 324 mM DTT and 40 mM $MnCl_2$ (pH 7.5)

Enzyme Dilution Buffer : 20 mM Tris-HCl (pH 7.5), 5 mM DTT, 100 ug/ml BSA and 5% glycerol

DMSO : dimethyl sulfoxide 100%

Stop solution : 100% deionized formamide and Dextran Blue 2000 (5 mg/ml)

1M NaOH for denaturation

4.2.1.2. Protocol

a. Denaturation

Adjust the concentration of the template so that 20 µl contains 5 to 10 ug of DNA.
Add 5 µl of 1M NaOH. Incubate at 65° C for five minutes. Spindown the tube briefly.
The mixture is loaded on a microspincolumn and centrifuge at 12,000 RPM during two minutes.
Two aliquots (F and R) of 10 µl of the solution containing denatured template DNA are prepared in a 1.5 ml microcentrifuge tube.

b. Annealing of primers (Forward and Reverse)

Add 2 µl of annealing buffer in both tube (F and R).
Add 2 µl of fluorescent M13 Universal Primer in one tube (F) and 2 µl of fluorescent M13 Reverse Primer in the second tube (R).

The total volume in each tube is 14 ul. Vortex and centrifuge gently.
Preheat the annealing reaction mixtures at 65° C for five minutes. and then incubate immediately in a water bath at 37° C for ten minutes. Remove the tubes and leave them at room temperature for five minutes. Spindown the content of the tubes briefly.

c. Termination :

Preliminaries :

- For sequencing with a double stranded templates the T7 DNA polymerase is diluted in Enzyme dilution Buffer to a concentration of 6 to 8 units per 2 ul.

- Pipette 2.5 µl of the A mix, C mix, G mix, T mix respectively in four wells of a microsample plate. Keep it on ice.

Termination reaction :

In a 1.5 ml microcentrifuge tube add for each annealing reaction :

1 µl of extension buffer
3 µl of DMSO
2 µl of diluted T7 DNA polymerase.

Remove 6 µl of this mixture and add it in each annealing reactions (F and R), the total volume in each tube being now 20 ul.

Pipette 4.5 µl of this mixture into each of the four wells of the microsample plate. Mix carefuly and incubate for five minutes at 37°C.

Add 5 µl of Stop Solution to each reaction and mix gently.
Heat the reactions at 95° C for two minutes and then keep them on ice.

Load 6 to 8 µl into the appropriate wells of a sequencing gel.

4.2.2. Sequencing with no fluorescent primer

This technique uses Fluore-dATP Labelling Mix in combination with Auto Read Sequencing kit and an Automated Laser Fluorescent ALF DNA sequencer from Pharmacia.
This fluorescein-labelled dATP is incorporated during the labelling reaction.

4.2.2.1. Protocol

a. Denaturation and Annealing reactions

The denaturation reaction and annealing reaction are the same than B.1.2.a. and B.1.2.b. with a concentration of the primer between 10 to 100 pmoles.

b. Labelling reaction

To the tube containing the annealed template and the primer add 2 µl of fluor-dATP and 2 µl of diluted T7 DNA polymerase (6-8 units/2 ul). Mix gently.
Incubate the reaction mixtures at 37°C for ten minutes .

c.Termination

After incubation, add 1 µl of Extension Buffer and 3.5 µl of DMSO to each labelling reaction. Mix gently

.

Pipette 3 µl of each of the A mix, C mix, G mix, and T mix into four wells of a microsample plate. Add 5.4 µl of the reaction mixture to each of hte appropriate wells. Mix carefully.

Incubate for five minutes at 37° C. Then stop the reactions by adding 6 µl of Stop Solution.
Load 6 to 8 µl into the appropriate wells of the sequencing gel.

4.3. Thermo Sequenase fluorescent labelled primer cycle sequencing kit (Amersham)

The Thermo Sequenase is a thermostable DNA polymerase and then many cycles can be perfomed without the need to add new enzyme. Uniform sequence band patterns are generated and are easier to read.

4.3.1. Components

A reagent	Tris-HCl (pH9.5), magnesium chloride, Tween 20, Nonidet P-40, 2-mercaptooethanol, dATP, dCTP, dGTP, dTTP, ddATP, thermostable pyrophosphatase and Thermo Sequenase DNA polymerase.
C reagent	Same composition with ddCTP
G reagent	Same composition with ddGTP
T reagent	Same composition with ddTTP
Formamide loading dye	Formamide, EDTA and methyl violet

Fluorescent primers (5pmol/ul)

Template DNA (0.5-5ug)

4.3.2. Protocol

For each template DNA, prepare four microcentrifuge tubes containing respectively 2 µl of the A reagent, C reagent, G reagent and T reagent. Store on ice until required.

In a 1.5ml microtube add 5 µl of template DNA, 19 µl of water and 2 µl of the fluorescent primers. Mix carefully.

Add 6 µl of this mixture into each tubes containing the dNTP. Vortex. Add one drop of mineral oil and centrifuge the tubes at high speed for 20 seconds.

Start the PCR program by a pre-denaturation at 94°C for 2 minutes followed by 25 cycles : denaturation for 30 seconds at 95°C, annealing for 30 secondes at 60°C and elongation for 1 minute at 72°C.

When the program is completed, add 3 µl of Formamide loading dye to stop the elongation. Mix the reactions and centrifuge the tubes at high speed for 20 seconds.

Remove the 11 µl or each reaction from under the mineral oil, and put them on the wells of a microsample plate.

Warm the plate at 95°C for two minutes and place it directly on ice. Load 6 to 8 µl Of the reactions into the appropriate wells of the gel.

References

Ahern, H., 1994. *Does a molecular clock tick for bacteria?* ASM news 60: 600-603.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. 1990. *Basic local alignment search tool.* J. Mol. Biol. 215:403-410.

Amman, R. I., Ludwig, W., and Schleifer, K-H. 1995. *Phylogenetic identification and in situ detection of individual microbial cells without cultivation.* Microbial. Rev. 59: 143-169.

Andersson, S. G. E. and Kurland C. G. 1995. *Genomic evolution drives the evolution of the translational system.* Biochem. Cell Biol. In press.

Andersson, S. G. E. and Sharp, P. M. 1995. *Codon usage and base composition in Rickettsia prowazekii.* J. Mol. Evol. Submitted.

Andersson, S. G. E., Zomorodipour, A., Winkler, H. H. and Kurland, C. G. 1995. *Unusual organization of the rRNA genes in Rickettsia prowazekii.* J. Bacteriol. 177: 4171-4175.

Azad, A. F., Sacci, J. B., Jr., Nelson, W. M., Dasch, G. A., Schmidtmann, E. T., and Carl, M. 1992. *Genetic characterization and transovarial transmission of a typhus-like rickettsia found in cat fleas.* Proc. Natl. Acad. Sci. 89: 43-46.

Baylis, H. A., and Bibb, M. J., 1988. *Organisation of the ribosomal RNA genes in Streptomyces coelicolor A3(2).* Mol. Gen. Genet. 211: 191-196.

Chen, X, and Frinch, L. R., 1989. *Novel arrangement of rRNA genes in Mycoplasma gallisepticum: separation of the 16S gene of one set from the 23S and 5S genes.* J. Bacteriol. 171: 2876-2878.

De Rijk, P., Van de Peer, Y., Van den Broeck, I., De Wachter, R., 1995. *Evolution according to large ribosomal subunit RNA.* J. Mol. Evol. 41: 366-375.

Drancourt, M., and Raoult, D., 1994. *Taxonomic position of thr Rickettsiae: current knowledge.* FEMS Microbiol. Rev. 13: 13-24.

Eremeeva, M. E., Roux, V. and Raoult, D. 1993. *Determination of genome size and restriction pattern polymorphism of Rickettsia prowazekii and Rickettsia typhi by pulsed field gel electrophoresis.* FEMS Microbiol. Lett. 112: 105-112.

Fleischman, R. D., Adms, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Tomb, J-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., Venter, C. 1995. *Whole-genome random sequencing and assembly of Haemophilus influenzae.* Rd.Science 269: 496-512.

Fraser, C. M., Gocayne, J. D., White, O., Adms, M. D., Clayton, R. A., Fleischman, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, J. L., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J-F., Dougherty, B. A., Bott, K. F., Hu, P-C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison III, C. A., Venter, C. 1995. *The minimal gene complement of Mycoplasma genitalium.* Science 270: 390-403.

Goffeau, A. 1995. *Life with 482 genes.* Science 270: 445-446.

Gray, M. W., 1993. *Origin and evolution of organnelle genomes.* Current Opinion in genetics and development 3: 1-7.

Guillon J-M., Mechulam, Y, Schmitter, J-M., Blanquet, S., and Fayat, G. 1992. *Disruption of the gene for Met-tRNA$_f^{Met}$ Formyltransferase severely impairs growth of Escherichia coli.* J. of Bacteriol. 174: 4294-4301.

Higgins, D. G., Bleasby, A. J. and Fuchs, R. 1992. *ClustalW: improved software for multiple sequence alignment.* CABIOS 8: 189-191.

Krawiec, S. and Riley, M. 1990. *Organisation of the bacterial chromosome.* Microbiol. Rev. 54: 502-539.

Kurland C. G., 1993. *Evolution of mithochondrial genomles and the genetic code.* BioEssays 14, 10: 709-714.

Li, W-H., 1993. *Unbiased estimation of the rates of synonymous and nonsynonymous substitution.* J. Mol. Evol. 36: 96-99.

Meinnel, T. and Blanquet, S. 1994. *Characterisation of the Thermus thermophilus locus encoding peptide deformylase and Methionyl-tRNA$_f^{Met}$ Formyltransferase.* J. Bacteriol. 176: 7387-7390.

Muto, A. and Osawa, S. 1987. *The guanine and cytosine content of genomic DNA and bacterial evolution.* Proc. Natl. Acad. Sci. 84: 166-169.

Ochman, H., and Wilson, C. 1987. *Evolution in bacteria: evidence for a universal substitutuion rate in cellular genomes.* J. Mol. Evol. 26: 74-86.

Ohashi, N., Fukuhara, M., Shimada, M., Tamura, A., 1995. *Phylogenetic position of Rickettsia tsutsugamushi and the relationship among its antigenic variants by analyses of the 16S rRNA gene sequences.* FEMS Micro. Letters 125: 299-304.

Olson, G. J., Woese, C. R., and Overbeek, R. 1994. *The winds of (Evolutionary) change: breathing new life into microbiolgy.* J. Bacteriol. 176: 1-6.

Sanger, F., Nicklen, S., and Coulson A. R., 1977. *DNA sequencing with chain-terminating inhibitors.* Proc. Natl. Acad. Sci. 74: 5463-5467

Stothard, D. R., Clark, J. B. and Fuerst, P. A., 1994. *Ancestral divergence of Rickettsia belii from the spotted fever and typhus group of Rickettsia and antiquity of the genus Rickettsia.* Inter. J. of Syst. Bacteriol. 44, 4/ 798-804.

Stothard, D. R., 1995. *The evolution history of the genus Rickettsia as inferred from 16S and 23S ribosomal RNA genes and the 17 Kilodalton cell surface antigen gene.* Dissertation, Ohio State University, Columbus, OH.

Syvänen, A-C., Amiri, H., Jamal, A., Andersson, S. G. E., and Kurland, C. G., 1995. *Disposition of elongation factor genes reflects the genomic evolution of Rickettsia prowazekii.* PNAS submitted.

Tamura, A., 1988. *Invasion and intracellular growth of Rickettsia tsutsugamushi.* Microbiol. Sci. 5: 228-232.

Taschke, C., Klinkert, M-Q., Wolters, J., and Hermann, R., 1986. *Organisation of the ribosomal RNA genes in Mycoplasma hyopneumoniae: the 5S rRNA gene is separated from the 16S and 23S rRNA genes.* Mol. Gen. Genet. 205: 428-433.

Weisburg W. G., Dobson, M. E., Samuel, J. E., Dasch, G. A., Mallavia, L. P., Baca, O., Mandelco, L., Sechrest, J. E., Weis, E., Woese, C. R., 1989. *Phylogenetic diversity of the Rickettsiae*. J. Bacteriol. 171: 4202-4206.

Weiss, E. and Moulder, J. M., 1984. In Bergeys's manual of Systematic Bacteriology, N.R. Kreig and J.G. Holt, eds. (Williams and Wilkins, Baltimore) p.687.

Williams, S. G., Sacci, J. B., Jr., Schriefer, M. E., Andersen, E. M., Fujioka, K. K., Sorvillo, F. J., Barr, A. R., and Azad, A. F., 1992. *Typhus and typhus-like rickettsiae associated with opossums and their fleas in Los Angeles County, California.* J. Clin. Microbiol. 30°: 1758-1762.

Winkler H. H., 1995. *Rickettsia prowazekii, ribosomees ans slow growth.* Trends in Microbiol. 3: 196-198.

Winkler H. H., 1990. *Rickettsia species (as organisms).* Annu. Rev. Microbiol. 44: 131-153.

Woese, C. R., 1994. *There must be a prokaryote somewhere: microbiology's search for itself.* Microbiol. Rev. 58: 1-9.

Wood, D. O. and Waite, R. T. 1994. *Sequence analysis of the Rickettsia prowazekii gyrA gene.* Gene 151: 191-196.