Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 689

Patterns and Processes of Molecular Evolution in Rickettsia

BY

HALEH AMIRI

ACTA UNIVERSITATIS UPSALIENSIS
UPPSALA 2002

To my parents

Main references

The thesis is based on the following papers, which will be referred to in the text by roman numerals I-VII.

I. Syvänen, A.-C., **Amiri, H.**, Jamal, A., Andersson, S. G. E. & C. G. Kurland. 1996. A chimeric disposition of the elongation factor genes in *Rickettsia prowazekii. J. Bacteriol*. **178**:6192-6199.

II. **Amiri, H.**, Alsmark, U. C. M. & S. G. E. Andersson. 2002. Proliferation and deterioration of *Rickettsia* palindromic elements. *Mol. Biol. Evol.* In press.

III. **Amiri, H.,** Karlberg, O. & S. G. E. Andersson. 2002. In the phylogenetic footsteps of ATP- ADP translocases. *Manuscript*.

IV. **Amiri, H.,** Davids, W. & S. G. E. Andersson. 2002. Evolution of intergenic DNA in *Rickettsia*. *Manuscript.*

V. Davids, W., **Amiri, H.,** & S. G.E. Andersson. 2002. Loss of gene function: Clues from expressed gene fragments in *Rickettsia*. *Trends Genetics* In press.

VI. Frank, C., **Amiri, H.** & S. G. E. Andersson. 2002. Genome deterioration: Loss of repeated sequences and accumulation of junk DNA. *Genetica*. In press

VII. **Amiri, H.** & S. G. E. Andersson. 2001. *Rickettsia*: The highly rearranged cousin of mitochondria. *Recent Res. Dev. Microbiol*. **5**:321-329.

Table of contents

Introduction

Rickettsia prowazekii, the etiologic agent of body-louse typhus is a gram negative, rod-shaped bacterium that is incapable of growing outside of host cells (1). Epidemic or louse-borne typhus caused the death of millions people during the first and second world wars (2). As recently as 1995, there was a large outbreak of louse-borne epidemic typhus in Burundi. Due to the failure of public health programs and the appalling sanitation the disease subsequently swept across the higher and colder regions of the country (3).

Although there is a long history of human suffering due to microbes such as *Yersinia pestis* and *R. prowazekii*, the evolution of pathogenicity is a new challenge to scientists. Knowledge about the molecular mechanisms of pathogenicity reveals that the relationship between humans and bacteria is very dynamic. While scientists try to come up with new strategies to counter the challenges posed by microorganisms, bacterial genomes are exposed to intensive rearrangements, providing fast adaptations to environmental changes.

In addition to its medical interest, *Rickettsia* is an interesting model organism for evolutionary studies. The small genome size of *Rickettsia* is a product of several modes of reductive evolution and its close phylogenetic relationship to mitochondria makes this organism even more interesting. The transition to the obligate intracellular lifestyle has caused *Rickettsia* to undergo evolutionary processes such as genomic rearrangements, duplications and deletions.

In this thesis, I have focused on the origin and duplication of genes encoding ATP/ADP translocases as an example of the evolution of novel transport systems in *Rickettsia*. I have also studied genomic rearrangements and reductive evolution in *Rickettsia*, including both rapid and stepwise gene loss, evolution of palindromic repeat sequences, and finally, the influence of deletional bias on the genome size of bacteria.

Molecular phylogeny and clinical features of *Rickettsia*

The genus *Rickettsia* belongs to the α–subclass of the proteobacterial phylum (4) (figure 1). Based on the characteristics of vectors, hosts and antigenic cross-reactivity, the genus is further divided into three biotypes, namely the spotted fever group (SFG), the typhus group (TG) and the scrub typhus group (STG) (1). However, the tick-borne species *Rickettsia bellii* dose not fit into these three groups; these bacteria have characteristics common to both the

spotted fever group and the typhus group biotypes and also exhibit some unique features (1). In addition, phylogenetic analyses reconstructed from the 16S and 23S rRNA gene sequences indicate that *R. bellii* is not a member of either the SFG or the TG. Instead, this species appears to have diverged prior to separation of the TG and SFG, but subsequent to their divergent from the STG. In cell culture, *R. bellii* is capable of growing both inside of the cytoplasm and within the nucleus (5, 6).

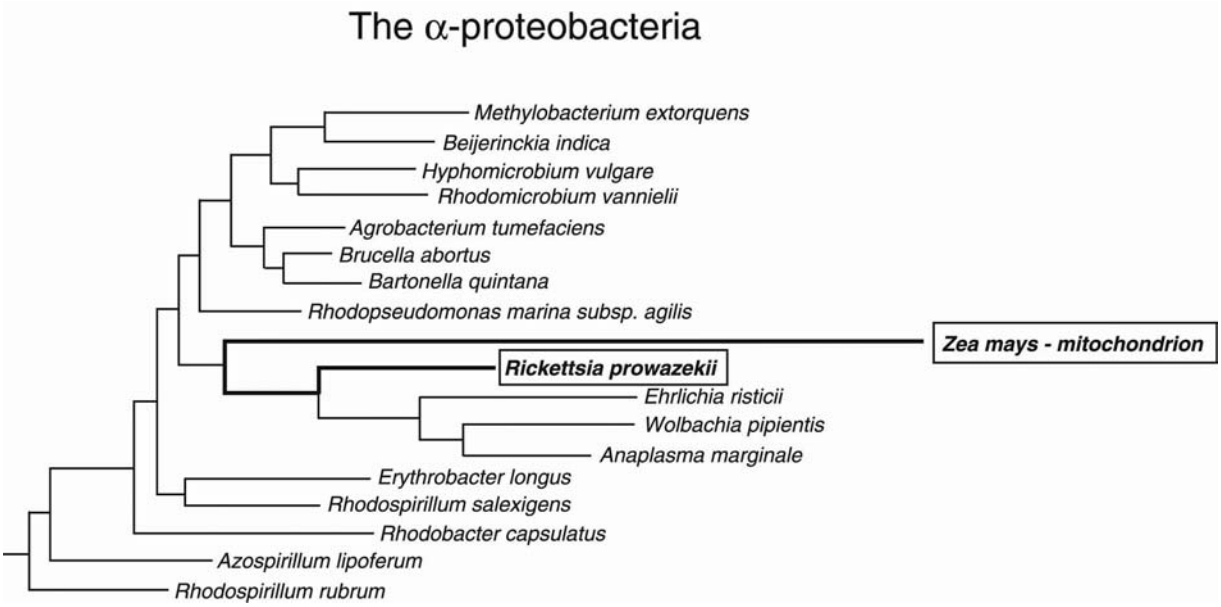

Fig. 1. Phylogenetic tree based on rRNA genes shows the position of *Rickettsia prowazekii* within the α–proteobacteria and its close phylogenetic relationship to mitochondria. The tree is modified and taken from Ref. 4.

More than 20 different members of SFG are known so far and new species continue to be recognized in different geographical regions. The SFG whose representatives are globally distributed consists of both pathogenic and non-pathogenic species. The non-pathogenic group includes for example *R. montana, R. parkeri, R. rhipicephali,* and *R. massiliae* (7,8). The most common diseases caused by the pathogenic members of SFG are typhus-like rickettsial diseases such as Rocky Mountain spotted fever (RMSF), African tick typhus and rickettsial pox.

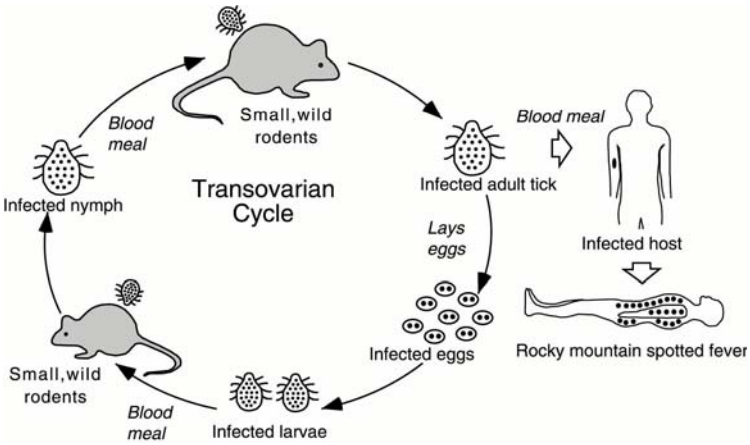

Fig. 2. Transovarian passage plays an important role in maintaining the *Rickettsia* infection in nature. The figure is modified and taken from Ref. 9.

The *Rickettsiae* are recycled in nature by transovarial transmission from infected female ticks to infected ova. These hatch into infected larval offspring that then infect small mammals (rodents). This so-called "transovarian passage" is critical for the maintenance of *Rickettsia* in nature from one generation of tick to another. Humans become an incidental host after infection by *Rickettsia* via an adult tick (9) (figure 2). *Rickettsia akari*, the causative agent of rickettsial pox is transmitted by house mice to humans. In cell culture, species of the SFG grow and multiply in cytoplasm, but they can also grow within the nucleus (10).

The TG consists of two insect born species, *R. typhi* and *R. prowazekii* . Epidemic typhus, caused by *R. prowazekii*, is transmitted in the faeces of human body lice and has historically been a major killer in wartime. *R. typhi* a causative agent of endemic murine typhus, primarily infects rodents and is spread to humans by rat flees, but the symptoms are milder than those of epidemic typhus. In cell culture, TG species grow and multiply in the cytoplasm and are unable to infect the cell nucleus (1).

The only member of the STG, *Orientia tsutsugamushi* causes scrub typhus (tsutsugamushi fever) in human. *O. tsutsugamushi* is capable of growing in both the nucleus and the cytoplasm and is transmitted to humans by infected mites (11).

The size and G + C composition of *Rickettsia* genomes

The G+C content of bacterial genomes vary from approximately 25% to 75% (15), with tRNA and rRNA genes being highly conserved in base composition features (4). In protein coding genes the G+C content values vary for different sites of the codon. The variation in G+C content at the first codon position is about 50% in known bacterial genomes. This value is decreased to 10% at the second codon position. However, nucleotide composition at the third codon position can vary over a 10-fold range from 10% in *Mycoplasma* to 92 % in *Streptomyces* (12,13). This is explained by the redundancy of genetic code, i.e. most of the changes in base composition at the third codon position result in silent substitutions (14).

In general, obligate intracellular bacteria have small genome sizes. The genomes of known *Rickettsia* species vary in size from 1.1 Mb in *R. prowazekii* to 1.66 Mb in *R. bellii*. In the TG *Rickettsia*, G+C composition is about 29%. However, the SFG *Rickettsia* has a slightly higher G+C content value of 32-33% (16-19). The G+C content of the intergenic regions is estimated to be lower than the total genomic G+C content value in *Rickettsia*. As an example, in eight investigated intergenic regions, the GC content is about 21-23 % and 30-31% for the TG and the SFG respectively (paper IV).

Evolution of novel transport systems in *Rickettsia*

During the course of evolution towards an obligate intracellular life style, the bacteria appear to have abandoned much of the genetic material that was previously essential for their survival, such as a large number of biosynthetic genes. This has forced obligate intracellular bacteria to establish efficient transport systems in order to obtain essential metabolites and coenzymes from the host cell cytoplasm. Novel genes, such as those involved in protein transport, can be obtained by horizontal gene transfer or by gene duplication (20). Through horizontal gene transfer, genetic information is transferred from a donor organism to a distantly related recipient cell. This process is more common in free-living bacteria than in obligate intracellular bacteria (21). Another source of evolution of novel gene functions is gene duplication followed by sequence divergence. If we assume that the probability of these events are very low in obligate intracellular bacteria, we would expect that both of these evolutionary events have taken place early in the evolutionary history of *Rickettsia,* i.e. prior or at the time of its entrance to the intracellular habitat.

Comparative analyses of the distantly related obligate intracellular parasites may provide valuable information about the characteristics of these bacteria that have evolved as a consequence of their life style. Indeed, the comparative analyses of the two genomes of *R. prowazekii* and *Chlamydia trachomatis* revealed a few striking examples of reductive convergent evolution (22). In both genomes the genes coding for enzymes involved in the biosynthesis of purines or pyrmidines are absent (23,26). However, intracellular specialists such as *Rickettsia* and *Chlamydia* have found a way to cover this loss by gaining efficient membrane transport systems. For example, the genome sequence of *Chlamydia trachomatis* revealed the existence of 13 ABC transporters primarily associated with amino acid and oligopeptide transport. Furthermore, the identification of permeases in the *Chlamydia* genome makes this intracellular parasite able to transfer magnesium, phosphate, nitrate, and sulphate from the host cell (23).

In the next section, I will discuss the evolution of ATP/ADP translocases as an example of a novel and specific transport system in *Rickettsia.*

Duplication of ATP/ADP translocases in *Rickettsia*

The identification of genes coding for ATP/ADP translocases in the genomes of *Rickettsia*

and *Chlamydia* indicates the presence of a unique transport system in these bacteria (23-27). This transport system is also found in organelles and in the eukaryotic parasite, *Encephalitozoon cuniculi* but not in any other bacterial genomes known to date (28-41). In mitochondria, ATP is generated by oxidative phosphorylation and secreted into the cytoplasm by the aid of ATP/ADP translocases. Unlike the mitochondrial ATP/ADP translocases, these proteins are responsible for uptake of ATP from the host cell cytoplasm in the intracellular parasites *Rickettsia* and *Chlamydia*. ATP/ADP translocase proteins are 500 amino acid long monomers and have similar topology in *Rickettsia*, *Chlamydia*, and plastids (25,27). However, mitochondrial ATP/ADP translocase has six transmembrane domains and exhibits an internal sequence triplication of 100 amino acids containing two membrane-spanning α–helices (30-34).

There are five ATP/ADP translocases in the typhus group as well as in the two investigated members of the spotted fever group, *R. rickettsii* and *R. montana.* It has been shown experimentally that one of the translocases (*tlc*1) in *R. prowazekii* catalyze the transport of ADP and ATP (24). In addition, studies of transcription regulation in *R. prowazekii* reveals that the amount of the transcript is correlated with the concentration of ATP in the host cells cytoplasm. That is, the level of expression decreases in cells that are heavily infected with the *R. prowazekii.* At the same time, expression of genes encoding proteins involved in ATP-generating metabolic pathways are upregulated when the concentration of ATP is low in the host cell cytoplasm. This means that *Rickettsia* exploit the host cells ATP during an early phase of infection but are capable of producing their own ATP during the late phase of infection (42).

It has been previously shown that the five paralogous *tlc* genes in *R. prowazekii* are expressed in cell culture systems (42). In addition, sequence data based on these genes from the *R. typhi*, *R. montana* and *R. ricketsii* shows that they do not contain any stop codons or frame shifts, i.e. they are most likely also expressed in these species (paper III). So far it's not known whether all five *tlc* genes in *Rickettsia* encode proteins with identical functions or whether they code for proteins with somewhat different functions.

The discovery of five duplicated, conserved genes in *Rickettsia* seems not to be a characteristic feature of resident genomes. Since obligate intracellular parasites such as *Rickettsia* tend to lose genetic information during the course of evolution gene loss rather than gene duplication is expected. However, a relaxed degree of reductive evolution would be obtained if there were a strong selection for the function of the duplicated gene. Thus, the duplication of the *tlc* genes in *Rickettsia* is most likely explained by their important role in maintaining an efficient uptake and transport system of host cytoplasmic ATP.

The ancient origin of ATP/ADP transporters in *Rickettsia*

Sequence homology and phylogenetic analyses indicate an independent origin of the rickettsial ATP/ADP translocases and the mitochondrial ATP/ADP translocase (paper III). It has been suggested that mitochondrial ATP/ADP translocases have evolved by vertical transmission from an ancestral gene with a broader transport function (43).

In contrast, the sporadic distribution of ADP/ATP translocases in phylogenetically distant genomes may be explained by recent acquisition of these genes by horizontal gene transfer events. Indeed, Wolf, Aravind and Koonin suggest that two horizontal gene transfer events explain the acqusition of these genes in *Chlamydia* and *Rickettsia* (44). According to this suggestion, the *tlc* genes were first transferred to *Chlamydia* from plants and then by a second horizontal gene transfer event they were transferred from *Chlamydia* to *Rickettsia* (44).

However, an amino acid sequence comparison between the chlamydial and rickettsial ATP/ADP transporters did not reveal any strong sequence homology. There are only 19 amino acids that are uniquely shared between *Rickettsia* and *Chlamydia.* This is much less than the number of shared amino acids between *Chlamydia* and plastids translocases (148 amino acids). In addition, a phylogenetic reconstruction based on *tlc* amino acid sequences indicates that the rickettsial ATP/ADP translocases do not cluster with the plastid and the *Chlamydia* ATP/ADP translocases. Thus, they seem to have diverged prior to the divergence of ATP/ADP translocases in plastids and *Chlamydia*. Furthermore, superimposition of the two phylogenetic trees belonging to mitochondria and plastid/parasite translocases reveals that the origin of plastid/parasite translocases is as deep as for the mitochondrial translocases (paper III).

We speculate that, similar to the mitochondrial ATP/ADP translocases, duplication and divergence of membrane proteins with other functions may have resulted in the plastid/parasite type of ATP/ADP translocases. Indeed, we have found a possibly similar hypotetical protein in the ψ-proteobacterium *Xylella fastidiosa*. This protein is 441 amino acids long and is predicted to have 9 transmembrane helices. Amino acid comparisons of the known ADP/ATP transporters with the hypotetical protein from *Xylella fastidiosa* reveal identical amino acids at three conserved positions.

Taken together, our sequence and phylogenetic analyses are not consistent with a recent acquisition of the ATP/ADP translocases in *Rickettsia* by horizontal gene transfer. These transporter proteins seem to be of ancient origin, and may have originated by duplication and divergence of other membrane transport proteins.

ATP/ADP translocases provide insights into the evolution of mitochondria

Sequence comparisons of mitochondrial genomes with the relevant bacterial genome sequences together with phylogenetic reconstruction of related proteins have provided overwhelming support for an endosymbiotic, eubacterial origin of mitochondria (45-50). The question no longer is whether it occurred, but why, by what process (es), and what was the characteristics feature of the endosymbiont?

All standard models of the endosymbiotic theory rely on the original role of the mitochondria as the ATP-generating organelles of eukaryotes. According to these models, the most likely reason for the symbiosis leading to the mitochondrial lineages was the ability of the symbiont (free-living α–proteobacteria) to provide the anaerobic host with aerobically generated ATP. The discovery of a non-bacterial origin of mitochondrial ATP/ADP translocases has challenged this view (51-53). We now know that ATP/ADP translocaes are only found in the two obligate intracellular parasites *Chlamydia* and *Rickettsia,* since free-living bacteria are not capable of obtaining ATP from the environment (52-54). These data strongly suggest that, unlike the modern mitochondria, the free-living ancestor of mitochondria was unable to export ATP or import proteins. The hypothesis that the ATP generation was not the driving force for the origin of the mitochondria may sound a little astonishing to us since we are used to see the mitochondria as the powerhouse of the eukaryotic cells.

However, another functional role of the mitochondria is to reduce the intracellular oxygen to water through oxidative phosphorylation, a process accomplished of five membrane bound protein complexes including cytochrome oxidase and cytochrom *b*. These proteins form a monophyletic group with the corresponding proteins in *R. prowazekii* (55*)*. This suggests that the α–proteobacterial ancestor of mitochondria had already acquired an aerobic respiratory chain at the time of symbioses.

According to the so-called "Ox-Tox" hypothesis, the driving force for endosymbiosis was to reduce the concentration of ancient oxygen and thereby protect the anaerobic host from its toxic effects (53,56). Roughly two billion years ago, when the concentration of atmospheric oxygen began to rise dramatically, the need for an endosymbiotic relationship for the anaerobic host with an aerobic symbiont was essential (57). Even today, modern organisms have enzymes such as peroxidases, catalyses, and superoxide dismutases that protect the cell from the toxic effect of oxygen respiration.

In addition, based on the Ox-Tox model, the evolution of the mitochondrion

from the endosymbiont required a system for mitochondrial import and export that has been shown to be derived from nuclear, eukaryotic genes (58-60). The recruitment of nuclear derived ATP/ADP translocases by mitochondria is one example of such a system. These proteins are found in all mitochondria and evolved prior to divergence of the major branches of eukaryotes. As a result of the novel acquisition of ATP/ADP translocases in the mitochondrial proteome, the host cell had access to an efficient supply of ATP. For this reason, ATP/ADP translocases may serve as a hallmark for the transformation of the endosymbiont into an organelle (59).

Genome rearrangements in *Rickettsia*

Genome comparison between closely related species have revealed that gene rearrangements and genome downsizing appear to have resulted from extensive recombination events between dispersed repetitive sequences. Depending on the orientation of the repetitive sequences, the outcome of homologous intrachromosomal recombination can be either a deletion or an inversion of one or two repetitive sequences and the corresponding intervening sequence (60-62). Comparison of the two genome sequences of *Mycobacterium tuberculosis* (4.41 Mb) and *Mycobacterium leprae* (3.27 Mb) has shown that the smaller genome size of *M. leprae* is the result of large-scale rearrangements and deletions arising from homologous recombination events between related repetitive sequences. These repeptitive sequences include transfer RNA genes and copies of the three major repeats, RLEP, REPLEP and LEPREP (63-64).

Numerous research investigations of *Rickettsia*, including sequence analyses of the *R. prowazekii* genome, indicate that the genomes of these intracellular parasites are highly derived and rearranged (paper I&VIII, 65,66). The small genome size of *Rickettsia* is a result of reductive evolution. This reductive mode of evolution has had an effect not only on the genome size of these organisms but has also changed their genomic architecture. We have observed several examples of a disrupted gene orders in *Rickettsia* (66-68). It is noteworthy that the organization of these genes is otherwise highly conserved in a broad range of microbial genomes (69-76). The major mechanism responsible for genomic rearrangements in *Rickettsia* is homologous intra-chromosomal recombination. The outcome of intra-chromomsomal recombination between inversely oriented repetitive sequences is an inversion of the intervening sequences (60,61). Such an inversion has been observed in *Rickettsia* in the so-called super-ribosomal protein gene operon (77-79). This operon is highly conserved in a broad range of bacteria and archaea and consists of about 40 genes locating in seven operons with the conserved order *secE-nusG*, L11, L10, *str*, S10, *spc* and α (figure 3).

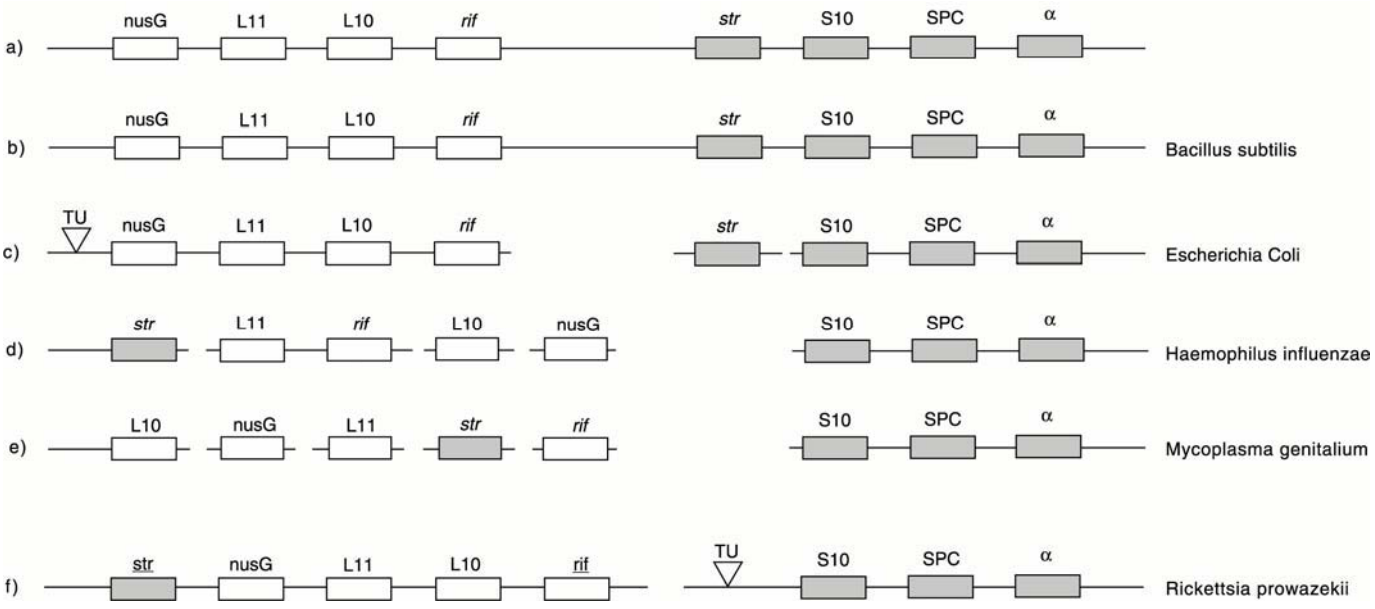

Fig. 3. Schematic representation of the organisation of super ribosomal protein gene cluster in a variety of bacteria. a) The expected organization of the ribosomal protein gene cluster in the common ancestor of Bacteria and Archaea. These operons are conservatively arranged in the genome of b) *Bacillus subtilis* and c) *Escherichia coli*, but have dispersed organisation in the genome of d) *Haemophilus influenzae*, e) *Mycoplasma genitalium* and f) *Rickettsia prowazekii.*

The conserved organization of 40 genes in these seven operons is thought to reflect the ancient origin of the super ribosomal protein gene operon (60). Despite the conserved orientation of the genes in the ribosomal protein gene operon in many bacteria such as *E. coli* and *Bacillus subtilis*, these genes are scattered around the genomes of *Haemophilus influenzae*, *Mycoplasma genitalium* and *R. prowazekii* (73,60). A more detail study of rickettsial genes encoding for elongation factor Tu (*tuf*) and G (*fus*) suggest that homologous intra-chromosomal recombination is the main mechanism responsible for the unique order of genes in these organisms (paper I). The *tuf* gene is found in two or multiple copies in a broad spectrum of microbial genomes (74-76, 78). In general, one copy is positioned in the so-called streptomycin operon (*str*) that comprises the two ribosomal protein genes, *rpsL* and *rpsG,* followed by the genes encoding for elongation factors G (*fus*) and Tu (*tufA*) respectively. The second copy of the *tuf* gene is located in *tufB* operon upstream of the two-membrane proteins *secE* and *nusG* (70-71).

The genome of *R. prowazekii* is atypical in that it contains only one *tuf* gene. Furthermore, the gene organization deviates from the conventional gene cluster. That is, the single *tuf* gene is not a member of the *str* operon and the flanking genes at the 3` end of the *fus* gene in the *str* operon are *secE, nusG* and *rpsJKL* that are normally located downstream of the *tuf* genes in other species. The present disposition of elongation factor genes and the

corresponding flanking regions in *R. prowazekii* can be explained by an inversion of the intervening sequences as a consequence of homologous intra-chromosomal recombination at the two ancestral *tuf* genes and the subsequent deletion of one *tuf* gene (paper I).

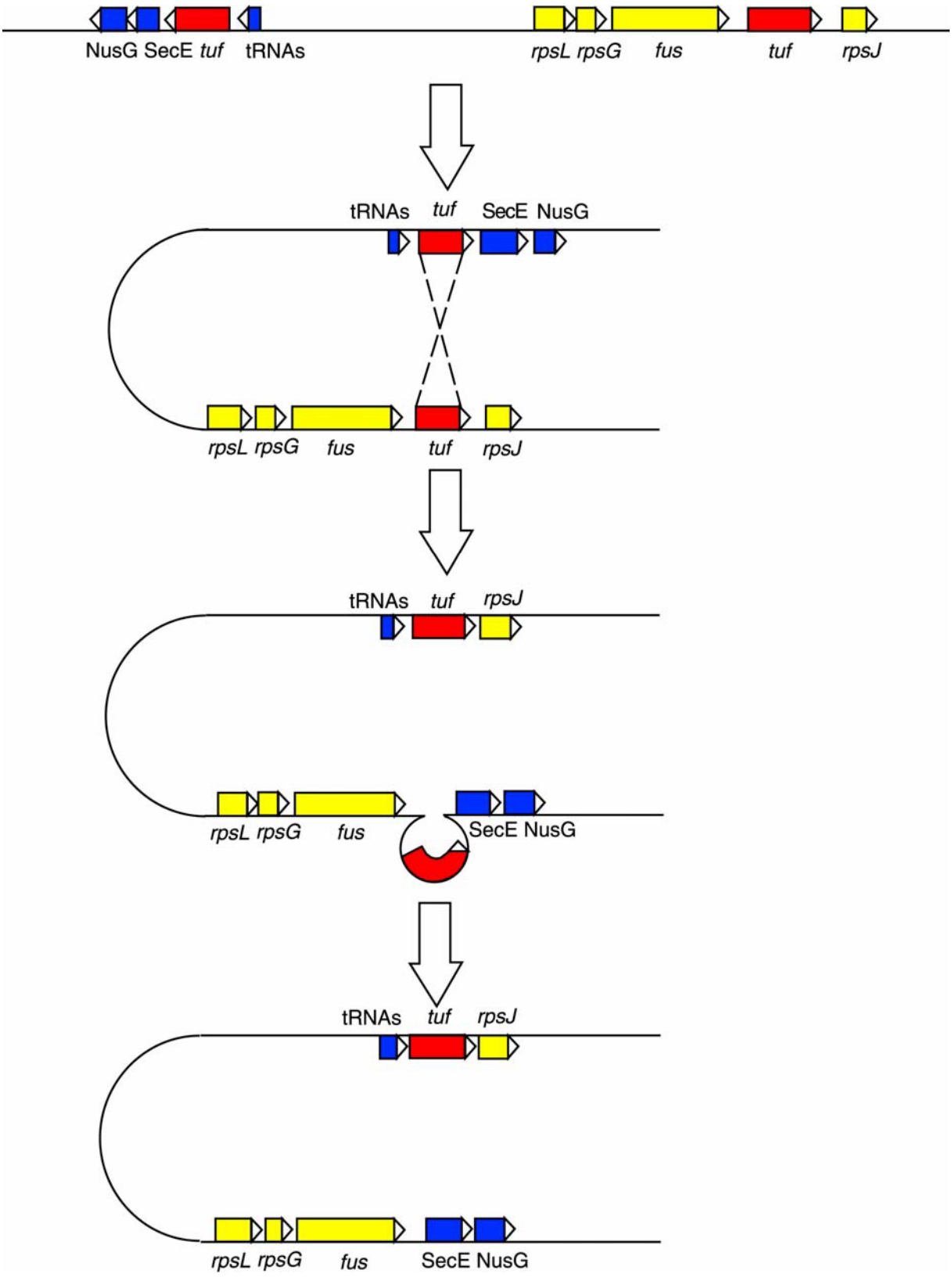

Fig. 4. Schematic representation of the postulated homologous intra-chromosomal recombination event between the two ancestral *tuf* genes that gave rise to the present gene organization and the subsequent deletion of one *tuf* gene. The figure is modified and taken from Ref.

Ancient origin of gene rearrangements in *Rickettsia*

Two major genomic rearrangements in *Rickettsia* occurred at the elongation factor genes and the rRNA genes. We suggest that the rearrangements at the *tuf* genes may have lead to a major inversion of the *Rickettsia* genome that occurred early in the evolution of *Rickettsia*. Indeed, our phylogenetic analyses based on the *tuf, fus* and the rRNA gene sequences indicate that the rearrangement of the duplicated genes and the subsequent deletions of the redundant *tuf* and rRNA genes occurred prior to speciation in *Rickettsia*. The finding of multiple *tuf* copies with the conventional organization in two closely related α–proteobacterial species, *Rhodobacter capsulatus* and *Agrobacter tumefaciens,* shows that this rearrangement event occurred subsequent to divergence of *Rickettsia* from *R. capsulatus* and *A. tumefaciens* (paper II).

Genomic streamlining in *Rickettsia*

Genome downsizing in bacteria is generally thought to be a consequence of a bias for deletions. The extent to which there is a bias for deletions in bacterial genomes has been examined for a broad range of bacteria (80). It has been proposed that this phenomenon plays a significant role in balancing genome size expansions in free-living bacteria and also in reducing the genome size of obligate intracellular bacteria. In other words, the compactness of bacterial genomes is a consequence of an ongoing deletional bias. This bias may result in the elimination of large segments of DNA by homologous recombination (rapid gene loss) or stepwise degradation and subsequent elimination of small segments of DNA (slow gene loss). The bias for deletions has been studied extensively in a variety of *Rickettsia* species and is the major point of the following sections in this thesis.

Rapid and extensive gene loss

It has been suggested that small obligate intracellular bacteria such as *Rickettsia* (1.1 Mb), *Mycoplasma* (0.58 Mb), *Coxiella* (1 Mb) and *Chlamydia* (1.0 Mb) have evolved from a free-living ancestor that had much larger genome size (1). During the course of evolution these obligate intracellular bacteria have lost a significant fraction of their genomes. As stated previously, large-scale deletions of genes and DNA sequences (rapid gene loss) can arise from intra-chromosomal recombination at directly oriented repetitive sequences.

The first victims of rapid gene extinction are those genes that are present in more than one copy in bacterial genomes, since the deletion of one or more copies will most likely not be lethal for the bacterial organism. Examples of such genes are those encoding rRNA genes and elongation factor proteins. In bacteria with an obligate intracellular lifestyle, such as *Rickettsia, Chlamydia* and *Mycoplasma, Buchnera*, these genes are found as single copy (52,53,54). Since duplicated genes such as *tuf*- and rRNA genes can serve as repetitive sequences, the deletion of one or multiple copies of these genes from the genome of intracellular bacteria may occur by intra-chromosomal recombination (61). Due to the consumption of repetitive sequences with each recombination event we expect that the rate of homologous recombination will decline during the course of evolution of obligate intracellular bacteria. Accordingly, we predict that the majority of homologous recombination-mediated deletion events in obligate intracellular bacteria occurred in the early

phase of the intracellular lifestyle.

Sequence comparison of the identified open reading frames (ORFs) located in the intergenic regions of different *Rickettsia* species has provided several examples of homologous recombination as the responsible mechanism for gene deterioration. For example, we found two pairs of short internal repeat elements in the putative ORF locating in the intergenic region of the *queA* gene, encoding S-adenosylmethionine tRNA and the gene encoding an abc transporter protein (*abcT3*). This ORF is absent from the TG *Rickettsia* and is inactivated and degraded within the SFG. We found that the size of this inactivated ORF varies between different members of the SFG. We suggest that size differences in the SFG are a result of independent homologous recombination between the internal repeat elements in this region (paper IV). The degree and positions of deletions caused by intra-chromosomal recombination in *Rickettsia* is different among the species, which suggests that the homologous recombination is an ongoing process that may result in an ongoing loss of genes under weak or no selection pressure.

Stepwise degradation and loss of genetic information in *Rickettsia*

The patterns and rates of mutations in resident genomes such as obligate intracellular parasites, endosymbionts and cellular organelles, is different from that of free-living bacteria. Small population size, lack of recombination, and bottlenecks during the transmission to the next host are characteristic features of resident genomes such as *Rickettsia (*81*)*. As a result, the fixation rate of the slightly deleterious mutations is higher in obligate intracellular parasites compared to that of free-living bacteria (82). The tendency to accumulate deleterious mutations in asexual populations living on the inside of a host cell is thought to be an effect of Mullers ratchet. Mullers ratchet may result in the gradual accumulation of harmful mutations, which in the absence of compensatory back mutations may lead to the extinction of the obligate intracellular parasites (83). It has been shown that Mullers ratchet is acting on the resident genomes such as those of *Buchnera* and *Rickettsia*, and it has been suggested that this phenomenon will generate smaller and smaller genomes (84-88).

We have already discussed that genomic streamlining may be caused by intra-chromosomal recombination at duplicated sequences resulting in the deletion of one or multiple copies of the duplicated sequence. Alternatively, the loss of genetic information may be accomplished by a gradual accumulation of small deleterious mutations. Due to the higher fixation rate of slightly harmful mutations in obligate intracellular bacteria, such as *Buchnera*

and *Rickettsia*, compared to free-living bacteria (89), we would expect an overall increase in the number of inactivated genes in these obligate intracellular bacteria. Detailed sequence comparison analyses of the pseudogenes and long intergenic regions belonging to the TG and the SFG provide numerous examples of gene shrinkage and loss in these intracellular parasites (paper IV&61). The genomic shrinkage can be triggered by accumulation of random mutations in DNA sequences that may not be any longer under purifying selection..

Indeed, it has been shown that the gene encoding for AdoMet synthetase is not active in the Madrid E strain of *R. prowazekii* since it has an internal termination codon (90). However, this gene is highly conserved among a wide range of bacteria as well as eukaryotes (65). The AdoMet synthetase is responsible for synthesis of AdoMet from methionine and ATP. The AdoMet is a source of methyl-groups that are essential for methylation of DNA, RNA, protein side-chain modifications, biosynthesis of amines, creatine, and carnitine (65). For these reasons, it is thought that the enzyme AdoMet synthetase is essential for all living organisms. Comparative analyses of several *Rickettsia* species have revealed that the AdoMet synthetase gene is intact in *R. typhi* and the Madrid B strain of *R. prowazekii* but contain stop codons and frame shift mutations (deletions/insertions) in *Rickettsia* species belonging to the SFG. The inactivation and degradation of essential genes such as AdoMet synthetase in the obligate intracellular parasites *Rickettsia* and *Chlamydia* may be compensated by establishment of a novel transport system by which the parasite can get access to essential compounds (90,91). Preliminary data suggests that such a transport system is present in *Rickettsia* (H. Winkler, personal communication).

Comparative analysis of long intergenic regions from several different species of *Rickettsia* provides additional support for an ongoing degradation and elimination of genes in *Rickettsia*. Examples of such genes are *gabD, ampG4* and *fic*. The *gabD* gene encodes succinate semialdehyde dehydrogenase oxidoreductase (SSDH) (paper IV). This gene catalyzes the formation of succinate from succinate semialdehyde by reducing NAD+/NAD(P)+ to NADH/NADPH. Succinate is one of the compounds participating in citric acid cycle. The enzyme SSDH is necessary for glutamate, tyrosin and butanoate metabolism (92). While this gene is intact in the SFG it is highly degraded in the TG *Rickettsia* (only traces of the gene at 3´ end are recognizable.) (paperIV).

The genes coding for AmpG are found in four copies in *R.conorii* and in three copies in *R. prowazekii* (93). These genes may make their carriers more resistant to some antibiotics such as β-lactam (93). The *ampG4* gene has been closely studied in five *Rickettsia* species, belonging to the TG and the SFG. The *ampG4* gene was found to be intact in three

members of the SFG, namely in *R. conorii, R. rickettsii* and *R. montana* while they are reduced in length in the TG. Furthermore, the gene contains many frameshifts and termination codons in members of the TG. The degradation of the *ampG* gene in the TG *Rickettsia* is consistent with the higher resistance to antibiotics observed in the SFG as compared to the TG. Finally, the gene encoding philamentation protein (fic) is intact in the investigated SFG *Rickettsia* but present as pseudogene in the TG *Rickettsia*. This gene contains deletions of 279 and 573 basepairs as well as 13 and 2 termination codons in *R. prowazekii* and *R. typhi*, respectively (paper IV). Apparently, these genes, *gabD, ampG4* and *fic* are still functional in the SFG, however, but accumulate mutations in a neutral manner in the TG *Rickettsia*.

Deletion profile in *Rickettsia*

Comparative analyses of long intergenic regions belonging to different *Rickettsia* species confirms that the unexpectedly high fraction of non-coding DNA (24%) in *R. prowazekii* represent remnants of inactivated and degraded genes (paper IV, 94). As was mentioned earlier, the degradation and subsequent deterioration of genes is a result of a bias for deletion mutations in *Rickettsia*, as observed by comparative analysis of the deletional profile of pseudogenes and highly fragmented genes. Reconstruction of the putative ancestral sequences for a large number of pseudogenes belonging to the TG and SFG *Rickettsia*, indicates that the number of deletions are considerably higher than the number of insertion mutations (94, paper IV). The size heterogeneity of pseudogenes between different species is a consequence of a mutation bias for deletions in these regions (paper IV).

The evolution of selfish DNA in *Rickettsia*

The term selfish DNA refers to repeat sequences that are capable of self-replication. Selfish genetic elements comprise a large diverge group, and due to their self-propagating feature they are found in most organisms (95). For example, a large fraction of eukaryotic genomes (up to 50%) consists of selfish DNA, such as retrotransible elements and transposons (96). The functional role of these self-replicating elements is still a matter of debate. There is now overwhelming evidence for a role in the regulation of gene expression. They are also responsible for gene or exon shuffling and duplications (97-99). Selfish DNA can also play a significant role in increasing the fitness of their host. For instance, the insertion of transposable element (P-element) in the third intron of the methuselah (*mth*) gene resulted in an extension (35%) of the life span of mutant flies (containing the inserted P-elements) compared to non-mutant flies (without no P-element insertion). The mutant line also showed increased resistance to starvation, high temperature and other stress factors. The accumulation of selfish DNA in the host genome may also have deleterious effects on the fitness of the host by inducing mutations in the genes (genetic diseases) (100). In this case, selection favours the survival of the propagating selfish DNA rather than the host's biology. How these selfish genetic elements originated and what was their role in the evolution of their host organisms has not been resolved.

Recently, a novel type of palindromic repetitive elements in the genome of *Rickettsia* has been identified (101). Due to the sporadic distribution of these repetitive sequences in *Rickettsia* genome, including insertion into the protein-coding genes, they have been classified as selfish DNA. However the selfish feature of these repeat elements has not been determined experimentally. In the following section, I am going to bring up questions concerning how early they have been acquired in *Rickettsia*, and what can be their possible destiny in the obligate parasite genome of *Rickettsia*.

Rickettsia palindromic elements (RPEs) - Long time residents

RPEs varies in length from 106-150 bp in *Rickettsia* and have palindromic structure (101). RPE-like repetitive elements with palindromic structure are also found in enterobacteria (102-104). However, unlike enterobacterial selfish DNA that is found in the non-coding regions, these elements are present in both genes and non-coding sequences. The abundance of palindromic selfish elements in protein coding genes has raised questions concerning their

origin and possible function in *Rickettsia*. There are 44 RPEs in *R. conorii*. Nineteen of these selfish elements are located in genes coding for a variety of proteins. Comparative analyses of the two *Rickettsia* genomes, *R. conorii* and *R. prowazekii,* has shown that RPEs are targeted to different genes in different species of *Rickettsia (*101*)*.

The different distribution of RPEs in various *Rickettsia* species can be explained in two possible ways. First, the variation in the distribution pattern may be explained by recent proliferation of RPEs in the *Rickettsia* genomes. If RPEs indeed represents new residents of the *Rickettsia* genomes, we would expect to find a considerable homology between their sequences within the same species. In other words, RPEs from the same species would cluster together with high bootstrap support and be separated by short branches in the phylogentic tree. However, phylogenetic reconstructions based on the nineteen RPEs in *R. conorii* indicate that these repeat elements are very diverged and do not cluster together (paper II). An alternative interpretation is that the RPEs were introduced into the *Rickettsia* genomes early in evolution and then deteriorated during the course of evolution. If so we would expect that RPE orthologs (RPE from different species) are more closely related than RPE paralogs (RPE within the same species). Indeed, sequence and phylogenetic analyses of RPEs located in the non-coding regions downstream of the genes encoding the elongation factors Tu and G (RPE-*tuf* and RPE-*fus* respectively) clearly indicates that these RPEs are long time residents of the *Rickettsia* genomes. In addition, the finding that the insertion sites of RPE-*tuf* and RPE-*fus* are identical in the different species suggest that their entrance into the *Rickettsia* genomes occurred some times prior to speciation of *Rickettsia* (paper II).

RPEs – Victims of reductive genome evolution

The independent loss of RPEs in some *Rickettsia* species is suggested to be responsible for species-specific differences in the host proteins targeted by RPEs (paper II). Indeed, it seems not unrealistic to believe that these selfish elements are temporary residents of the *Rickettsia* genome and that their elimination is caused by reductive genome evolution. The decay of RPEs in *Rickettsia* may be accomplished by homologous recombination between inversely oriented repeated sequences. One example of such a process has been identified in several *Rickettsia* species in which the RPE is located between two short repeated sequences. Alternatively, the RPEs may be degraded from the genome gradually by accumulating small deletions. This type of degradation is expected to happen more frequently than rapid loss, especially during the later stages of reductive evolution.

Indeed, the degradation of RPE-*fus* is an example of an RPE in different stages of degradation in different species of *Rickettsia.* This RPE is intact in some species of the SFG, it has accumulated some deletions mutations and termination codons in the TG and is hardly recognizable in *R. bellii*. Furthermore, we have found that deletions dominate over insertions in RPE-*fus*. This is consistent with the indel spectrum previously obseved in pseudogenes and other neutrally evolving genes in *Rickettsia*. In general there is a tendency for a higher deterioration rate in the TG compared to the SFG (paper II). This tendency has been seen in many sequences that are under weak selection pressure (paper IV,V). For this reason there is a considerably lower frequency of RPEs present in the *R. prowazekii* genome compared to that of *R. conorii*. Almost all of the few RPEs still present in the *R. prowazekii* genome are located inside ORFs. The RPEs that are incorporated inside the three genes, *kdtA, polA* and *RP545,* have been shown to have higher substitution rates than their flanking coding sequences (73). This provides additional support for the view that there is an ongoing degradation of RPEs in the *Rickettsia* genomes. Thus, it seems unlikely that the RPEs have proliferated recently in obligate intracellular parasites such as *R. conorii* that tend to eliminate the large fractions of their genomes during the course of evolution.

Are RPEs long-term hitchhikers?

The abundance of RPEs in different protein coding genes together with the conserved palindromic nature of these selfish elements is consistent with a possible functional role for the proteins in which they reside. It has been found that most of the RPEs presented in *R. conorii* are able to form stable hairpin-like mRNA secondary structures. However, the palindromic feature of RPEs belonging to *R. prowazekii* is less clear and only two out of ten are predicted to form hairpin structures (74). Indeed, the finding of unstable hairpin-like mRNA secondary structure for the majority of RPEs located in protein-coding genes suggest that this selfish DNA may have lost its function in most of the RPEs in *R. prowazekii.* This is also supported by a high evolutionary degradation rate in *R. prowazekii*. This means that the lack of purifying selection on RPEs has resulted in the extinction of most RPEs from the genome of *R. prowazekii*. However, they may still be under weak selection in the *R. conorii* genome, especially if inserted into protein coding genes. However, based on hydropathy analyses, they are predicted to be located at the surface of the various proteins (74). Due to their position on the surface of these proteins, the loss of these elements may not be critical for their hosts and may not affect the function of their carrier protein. Thus, these selfish

elements may possibly have had some functional role early in the history of *Rickettsia* but their role is currently thought to be insignificant and they seem to be in a process of rapid deterioration.

The influence of deletions on genome size in bacteria

In eukaryotes, genome size can vary from about 10^7 to 10^{11} basepairs (80). This broad variation in genome size is not related to the eukaryotic organisms complexity (the C-value paradox) (105). For instance, *Arabidopsis thaliana,* with the smallest genome in the plant kingdom known todate, has a 3000 fold smaller genome than the far simpler plant *Psilotum nudum* (sometimes called the "whisk fern"). It is also worth noting that the large fraction (80%) of the *P. nudums* genome consists of repetitive DNA. Other examples are the genomes of amphibians that contain 30 times as much DNA as human genome (106). The question is, if the enormous variation in genome size is not linear with organismal complexity, what is the explanation for this paradox? Dmitri A. Petrov and his coworkers have suggested that the answer to this long-standing mystery is that the rate at which junk DNA is eliminated from a genome is different among different organisms. The genome size of the Hawaiian crickets (*Laupala*) is about the 11-fold larger than that of *Drosophila*. It has been shown that the spontaneous deletions rate of the non-essential DNA is more than 40 times slower in *Laupala* than in *Drosophila*. The high deletion rate of unconstrained sequences such as dead-on-arrival (DOA) copies of non-LTR retrotransposable elements in *Drosophila* is suggested to be congruent with the general feature of mutation in this organism (105).

The C-value paradox may not solely be determined by differences in the insertion and deletion mutational spectra. For example, *C. elegans* has a very compact genome, smaller than that of *D. melanogaster*, although it has a lower rate of DNA loss (105). It is clear that the huge difference in genome size among organisms is not related to their complexity but the size of the non-coding region may be correlated to the organisms complexity. The most striking example is the human genome with a non-coding fraction that comprises ca 98% of the genome (106).

However, in bacteria, genome sizes do not vary as much as in eukaryotes. Known bacterial genome sizes vary over a 20-fold range from 580 kb in *Mycoplasma genitalium* to more than 10 Mb in several cynobacterial and myxobacterial species (108). In contrast to eukaryotic genomes, a large fraction of the bacterial genome consists of genic DNA and non-coding DNA comprises only 10% of the bacterial genomes. For this reason the size differences in bacterial genomes is often consistent with organismal complexity. The balance of two major forces, horizontal gene transfer and deletional bias can explain most of the genome size diversity among bacteria. The widespread impact of lateral gene transfer on

bacterial evolution should not be underestimated. Horizontal gene transfer occurs more frequently in free-living bacteria compared to bacteria in intracellular habitats. Lateral gene transfer results in extremely dynamic genomes and allow bacteria to adapt rapidly to changing environments by providing recipient organisms with new metabolic traits (for instance the acquisition of the lac-operon in *E.coli*), antibiotic resistance genes and genes that contribute to virulence such as pathogenicity islands (108). In addition to its significant role for creation of dynamic genomes, lateral gene transfer may contribute to expansion of genome size in bacteria. However, if lateral gene transfer were the only process affecting bacterial genomes, we would expect an ongoing expansion of genome size.

On the contrary, bacterial genomes are compact and the influence of horizontal gene transfer seems not to contribute to the expansion of genome size in bacteria. The compactness of bacterial genomes may be explained by reductive evolutionary forces that induce gene loss, which counteracts the influx of foreign DNA by horizontal gene transfer. The molecular details of reductive evolution are best understood by comparative analysis of known pseudogenes with their functional counterparts within the same taxa or closely related taxa. Since pseudogenes are neutral sequences, the profiles of deleterious mutations reflect the spontanous spectrum of mutations. As I discussed earlier there is a bias for deletions over insertions in obligate intracellular bacteria such as *Rickettsia* and *Buchnera*. Deletional bias also exists in a broad range of other bacteria such as *Neissera meningitidis, Helicobacter pylori, Sulfolobus solfataricus, Rhodobacter spharoides* (80). The effect of the deletional bias on genome size may be more visible in bacteria with obligate intracellular lifestyle where lateral gene transfer does not occur frequently and since the mutational degradation of genes as a consequence of this bias is normally counterbalanced by selection on gene function. Taken together, deletional bias contributes to maintaining constant genome size in free-living bacteria, but result in small genome size in obligate intracellular bacteria.

Deletion frequencies may depend on the density of repetitive elements

Studies on the processes of reductive genome evolution in obligate intracellular parasites have revealed the important role of repetitive sequences in affecting genome size and structure. The highly rearranged genome of *R. prowazekii* indicates that these organisms have undergone a large number of genomic rearrangements as well as extensive gene loss. Genome shrinkage in *Rickettsia* has been shown to be an ongoing process that results in loss of genetic material over time. As was mentioned earlier, the loss of large segments of genetic

information can occur rapidly by homologous recombination at directly oriented repetitive sequences. Since repetitive sequences are major substrates for homologous recombination, the variation in frequency of repetitive sequences in bacterial genomes and also in closely related species may contribute to different sizes and structures. Thus, we expect that genomes with a high number of repetitive sequences undergo homologous recombination more frequently, which in turn may result in a rapid deletion of DNA and a fast reduction of genome size over time.

A global survey of microbial genomes has shown that free-living bacteria with large genomes have a high content of repeated sequences and self-propagating DNA such as transposons and bacteriophages. In contrast, obligate intracellular bacteria tend to have small genomes with a low content of repeated sequences and no or few genetic parasites (paper VI). The frequency of the repeated sequences and genome size in microbial genomes may be determined by the lifestyle of the organism. With each homologous recombination (between directly oriented repeats) one or two repeat sequences will be lost. In free-living bacteria, the loss of repeated sequences by homologous recombination can be compensated by the import of novel repeat sequences from other bacteria by horizontal gene transfer or intra-specific recombination. Therefore, repeated sequences may have small effects on genome size in free-living bacteria. In contrast, Mullers ratchet will cause an irreversible loss of DNA including repeated sequences in obligate intracellular bacteria. Thus, the loss of repetitive sequences in isolated bacteria can not be compensated by a corresponding influx from other species and will therefore result in a small genome size.

The consumption of repeated sequences during the course of evolution may affect the rate of homologous recombination in resident genomes. We predict that the frequency of homologous recombination has declined during the evolution in obligate intracellular bacteria (paper VI). This in turn may affect the rate of genomic deletion. The latter will decrease over time under the assumption that the rate of fixation of slightly deleterious mutations is constant or less than zero. Genome size variation and differences in repeat content between free-living bacteria and obligate intracellular bacteria are the consequence of the lifestyle of these organisms. However, the difference in genome size and repeat content in obligate intracellular bacteria with identical lifestyles may be explained by a variety of factors, such as the number of hosts they infect, the extent of bottlenecks, the time since species divergence and the rate of evolution. Indeed, a comparative analysis of *R. prowazekii* and *R. conorii* as well as a separate study of eight non-coding regions in 5 different *Rickettsia* species (paperIV) suggests that reductive evolution has been influenced the TG *Rickettsia* to the larger extent than the SFG.

Concluding remarks

I have discussed the evolution of *Rickettsia* and the influence of obligate intracellular lifestyle on shaping their genomic architectures. Different evolutionary processes such as genome rearrangements, loss of genetic information and gene duplications have the been major focus of this thesis:

- **Genome rearrangements**

*R*ickettsia genomes are highly rearranged. Large genomic rearrangement is suggested to be the resulted of intra-chromosomal recombination at duplicated sequences. For example, we suggest that the unique organisation of the super ribosomal protein gene operon in *Rickettsia* is a consequence of intra-chromosomal recombination at the inversely oriented ancestral *tuf* genes followed by deletion of one *tuf* gene. Phylogenetic reconstructions based on *tuf* and *fus* genes from different species of *Rickettsia* and their closely related α-proteobacteria suggest that the rearrangement occurred at an early stage of the transition to an obligate intracellular life style, but prior to speciation within the genus *Rickettsia.*

- **Reductive genome evolution**

*R*eductive evolution is the major evolution process that has influenced the genomes of the *Rickettsia*. Reduction in genome size will primarily occur through by homologous recombination between directly oriented repetitive sequences. This will result in the deletion of the intervening DNA sequences. We suggest that a major loss of genetic information occurred via homologous recombination at repeated sequences (*tuf* genes, *rRNA* genes, and repetitive elements). The rate of homologous recombination has declined during the evolution due to consumption of repeated sequences during the recombination events. However, the elimination of genetic material will continue even after the consumption of repeated sequences, albeit with a slower rate by the accumulation of short deletion mutations. This in turn will give rise to the formation and temporary accumulation of inactive genes and junk DNA. The produced pseudogenes will serve as new victims of sequence degradation. Therefore, like any other neutrally evolving genes that are not under selective constraints, they are expected to slowly deteriorate from the genome. Gene loss in bacteria seems to be a consequence of a general bias for deletions. This process plays an important role in free-living

bacteria by balancing the influx of genetic information via horizontal gene transfer. In addition, the bias for deletions has resulted in small genomes in obligate intracellular bacteria such as *Rickettsia*. Comparative analyses of long intergenic regions that belong to the TG and SFG *Rickettsia* indicate that many genes that contain termination codons or frameshifts or that are completely deleted from one species may remain active in another species. Sequence and phylogenetic analyses suggest that deletion mutations occur more frequently than the insertion mutations in *Rickettsia*.

- **The invention of novel transport system**

The loss of essential biosynthetic genes during the transition to obligate intracellular life style has triggered the intracellular specialists to establish or develop novel transport system. One example of such a transport system is the ATP/ADP translocases that are only found in obligate intracellular parasites and organelles. We have found no support for the hypothesis that these transport systems have been transmitted among obligate intracellular parasites by recent horizontal gene transfer. On the contrary, these transport proteins seem to have an ancient origin and was acquired in *Rickettsia* long before speciation, possibly already during the establishment of the endosymbiont that later gave rise to the mitochondrion. We have suggested that the ATP/ADP transporters in *Rickettsia* have evolved by duplication and subsequent divergence from other transmembrane proteins long before the divergence of TG from SFG *Rickettsia.*

Future prospects

Comparative sequence analysis of a variety of *Rickettsia* species has provided a general overview of the different evolutionary processes that shape the genome of this obligate intracellular parasite. We have observed that reductive evolution is the major evolutionary force operating on these genomes, which has resulted in the inactivation and gradual elimination of genes and repetitive elements such as RPEs from *Rickettsia*. Comparison of the genomes of *R. prowazekii* (1.1 Mb) and *R. conorii* (1.3 Mb) indicate that the *R. conorii* genome contain a total number of 552 genes that are not present in *R. prowazekii*. A detailed comparative analysis indicates that as many as 200 of the unique genes in *R. conorii* correspond to inactivated gene fragments in *R. prowazekii* and that another 200 genes have

been completely eliminated from the *R. prowazekii* genome.

A comparison of active genes with their slightly degraded orthologs in *R. prowazekii* and *R. conorii* indicates that some of these degraded genes, also called split genes, are expressed (92). However, the substitution frequencies at non-synonymous and synonymous sites show no significant difference between the non-expressed ORFs and the expressed ORFs from the same set of split genes (paper V). This finding suggests that neither the expressed, nor the non-expressed split genes may be functional. In order to understand the mechanism of gene inactivation and degradation, we need to determine whether the expressed degraded genes are translated and if so whether the gene products are able to perform their original function and if so, how their expression is regulated. A genome-wide analysis of the *Rickettsia* genomes at the protein level will provide a clear picture of the functional roles of the encoded gene products. It will also highlight the role of genes involved in different metabolic pathways and tells us whether multiple functions may be encoded by some genes and define the role of the transport machinery in gaining essential compounds that can not be synthesized by the obligate intracellular parasite. In addition, functional genomics studies may shed light on the extent to which Muller ratchet have altered enzyme functions and kinetics in obligate intracellular parasites such as *Rickettsia,* and thereby their fait in the eukaryotic host cell.

References

1. Weiss, E., and J. W. Moulder. 1984**.** Order I. Rickettsiales **Gieszczkiewicz** 1939, 25 ^{AL} . In J. H. Holt (ed.), Bergeys manual of systematic bacteriology **vol. 1**. The Williams and Wilkins Co., Baltimore, pp. 687-704.

2.Gross, L. 1996. How Charles Nicolle of the Pasteur Institute discovered that epidemic typhus is transmitted by lice: reminiscences from my years at the Pasteur Institute in Paris. Proc. Natl Acad. Sci USA **93:**10539-10540

3. Raoult, D., Ndihokubwayo, J. B., Tissot-Dupont, H., Roux, V., Faugere, B., Abegbinni R. and Birtle, R. J. 1998. Outbreak of epidemic typhus associated with trench fever in Burundi. The Lancet. **352:**353-358.

4.Olsen, G. J., C. R. Woese, and R. Overbeek. 1994. The winds of (evolutionary) change: breathing new life into microbiology. J. Bacteriol. **176**:1-6.

5. Stothard, D. R., J. B. Clark, and P. A. Fuerst. 1994. The phylogenetic relationship of *Rickettsia bellii* to the spotted fever and typhus group of Rickettsia, and the antiquity of the genus. Int. J. Syst. Bacteriol**. 44:**798-804.

6. Stothard, D. R., and P. A. Fuerst. 1995. Evolutionary analysis of the spotted fever and typhus groups of *Rickettsia* using 16S rRNA gene sequences. Syst. Appl. Bacteriol. **18:**52-61.

7. Raoult D, Roux V. 1997. Rickettsiases as paradigms of new or emerging infectious diseases. Clin Microbiol Rev. **10:**694-719.

8. Azad, AF., Beard CB. 1998. Rickettsial pathogens and their arthropod vectors. Emerg Infect Dis **4:**179-86.

9. Thorner, A. R., Walker, D. H., Petri W. A. Jr. 1998 Rocky. Mountain Spotted Fever. Clin Infect Dis **6:**1353-9**.**

10. Wisseman, C. L. JR., E. A. Edlinger, A. D. Waddell, and M. R. Jones. 1976. Infection cycle of *Rickettsia rickettsii* in chick embryo and L-929 cells in culture. Infect Immun. **14:**1052-1064.

11. Tamura, A., N. Ohashi, H. Urakami, and S. Miyamura. 1995. Classification of *Rickettsia tsutsugamushi* in a new genus, Orientia gen. Nov., as *orientia tsutsugamushi* comb. Nov. int. J. Syst. Bacteriol**. 45:**589-591.

12. Ohkubo, S., Muto, A., Kawauchi, Y., Yamao, F., Osawa, S. 1987. The ribosomal protein gene cluster in *Mycoplasma capricolum.* Mol Gen Genet **210:**314-322.

13. Wright, F., Bibb, M. J. 1992. Codon usage in the G+C- rich *streptomyces* genome. Gene **113:**55-65.

14. Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, UK.

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

16. Eremeeva, M. E., Roux, V., 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:**105-12.

17. Roux, V., Raoult, D. 1993. Genotypic identification and phylogenetic analyses of the spotted fever group rickettsiae by pulsed-field gel electrophoresis. J Bacteriol. **175:**4895-904.

18. Tyeryar, Jr. F. J., Weiss, E, Millar, D. B., Bozeman, F. M., Ormsbee, R. A. 1973. DNA base composition of rickettsiae. Science **180:**415-7.

19. Schramek, S. 1974. Deoxyribonucleic acid base composition of Rickettsiae belonging to the Rocky Mountain spotted fever group isolated in Czechoslovakia. Acta Virol (Praha) **18:**173-4.

20. Li, W. H. and Graur, D. 1991. Fundamentals of Molecular Evolution. Sinauer associates, INC. publishers. U.S.a.
21. Andersson, S. G. E. and Kurland, C. G. 1998. Reductive evolution of resident genomes. Trends in Micro. Biol**. 6:**263-268.
22. Zomorodipour, A. & Andersson, S. G. E. 1999. Obligate intracellular parasites: *Rickettsia prowazekii* and *Chlamydia trachomatis*. FEBS Lett. **452:**11-15.
23. Stephens, R. S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R. L., Zhao, Q., Koonin, E. V. & Davis, R. W. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science **282:**7854-759.
24. Winkler, H. H. 1976. Rickettsial permeability: An ADP-ATP transport system. J. biol Chem **251:**389-396.
25. Williamson, L. R., Plano, G. V., Winkler, H. H., Krause, D. C. & Wood, D. O. 1989. Nucleotide sequence of the *Rickettsia prowazekii* ADP/ATP translocase encoding gene. Gene **80:**269-278.
26. Andersson, S. G. E., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T., Alsmark, U. C. M., 1998
Podowski, R. M., Näslund A. K., Winkler, H. H. &Kurland, C. G. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature **96:**133-140
27. Hatch, T. P., Al-Houssaine, E. & Silvermann, J. A. 1982. Adenine nucleotide and lysine transport in *Chlamydia psittaci*. J Bacteriol **150:**662-670.
28. Katinka, M. D., Duprat, S., Cornillot, E., Metenier, G., Thomara, F. et al., 2001. Genome sequence and gene compaction of the eukaryotic parasite *Encephalitozoon cuniculi*. Nature **414:**450-453.
29. Vignais, P. V. 1976. Molecular and physiological aspects of adenine nucleotide transport in mitochondria. Biochim Biophys Acta **456:**1-38.
30. Klingenberg, M. 1989. Molecular aspects of adenine nucleotide carrier from mitochondria. Arch Biochem Biophys **270:**1-14.
31. Heldt, H. W. 1969. Adenine nucleotide translocation in spinach chloroplasts. FEBS Lett **5:**11-14.
32. Pozueta-Romero, J., Frehner, M., Viale, A. M. & Akazawa, T. 1991. Direct transport of ADP glucose by an adenylate translocator is linked to starch biosynthesis in amyloplasts. Proc Natl Acad Sci USA **88:**5769-5773.
33. Neuhaus, H. E., Henrichs, G. & Scheibe, R. 1993. Characterization of glucose-6-phosphate incorporation into strach by isolated intact cauliflower-bud plastids. Plant Physiol (Rock) **101:**573-578.
34. Schunemann, D., Borchert, S., Flugge, U. I. & Heldt, H. W. 1993. ADP/ATP translocator from pea root plastids. Comparison with translocators from spinach chloroplasts and pea leaf mitochondria. Plant Physiol (Rock) **103:**131-137.
35. Kampfenkel, K., Möhlmann, T., Batz, O., Van Montagu, M., Inze, D. & Neuhause, H. E. 1995. Molecular characterisation of an *Arabidopsis thaliana* cDNA encoding a novel putative adenylate translocator of higher plants. FEBS Letters **374:**351-355.
36. Möhlmann, T., Tjaden, J., Schwöppe, C., Winkler, H. H., Kampfenkel, K. & Neuhaus, H. E. 1998. Occurrence of two plastidic ADP/ATP transporter in *Arabidopsis thaliana*. Eur J Biochem **252:**353-359.
37. Runswick, M. J., Powell, S. J., Nyren, P. & Walker, J.E. 1987. Sequence of the bovine mitochondrial carrier protein: structural relationship to ADP/ATP translocase and the brown fat mitochondria uncoupling protein. EMBO J **6:**1367-1373.
38. Saraste, M. & Walker, J. E. 1982. Internal sequence repeats and the path of polypeptide in mitochondrial ADP/ATP translocase. FEBS Lett **144:**250-254.

39. Klingenberg, M. 1993a. Dialectics in carrier research: the ADP/ATP carrier and the uncoupling protein. J Bioenerg Biomembr**. 25:**447-57.
40. Aquila H, Link TA, Klingenberg M. 1987. Solute carriers involved in energy transfer of mitochondria from a homologous protein family. FEBS Lett. **212:**1-9.
41. Hackenberg H, Klingenberg M. 1980. Molecular weight and hydrodynamic parameters of the adenosine 5′-diphosphate-adenosine 5′-triphosphate carrier in Triton X-100. Biochemistry. **19:**548-55.
42. Cai, J. & Winkler, H. H. 1996. Transcriptional regulation in the obligate intracytoplasmic bacterium *Rickettsia prowazekii*. J Bacteriol **178:** 5543-5545.
43. Kuan, J. & Saier, M. H. 1993. The mitochondrial carrier family of transport proteins: structural, functional, and evolutionary relationships. Crit Rev Biochem Mol Biol **28:**209-233.
44. **Wolf, Y. I., Aravind, L. & Koonin, E. V.** 1999. *Rickettsiae and Chlamydiae*: evidence of horizontal transfer and gene exchange. Trends Genet **15:**173.
45. Yang, D., Oyaizu, Y., Olsen G. J., Woese, C. R. 1985. Mitochondrial origins. Proc Natl Acad Sci USA **82:**4443-4447.
46. Gray, M. W., Cedergren, R., Abel, Y, Sankoff, D. 1989. On the evolutionary origin of the plant mitochondrion and its genome. Proc Natl Acad Sci USA**. 86:**2267-2271.
47. Gray, M. W. 1998. *Rickettsia*, typhus and the mitochondrial connection. Nature. **396:** 109-110.
48. Doolittle, W. F. 1998. A paradigm gets shifty. Nature. **392:**15-16.
49. **Gray, M. W.** 1992. The endosymbiont hypothesis revisited. Int. Rev. Cytol. **141:**233-357.
50. Gray, M. W., and B. F. Doolittle. 1982. Has the endosymbiont hypothesis been proven? Microbiol. Rev. **46:**1-42.
51. **Kurland, C. G., and Andersson, S. G. E.,** 2000, Origin and Evolution of the Mitochondrial proteome. Microbiol. Molecularbiol Rev. **64:**786-820.
52. Andersson, S. G. E. 1998. Bioenergetics of the obligate intracellular parasite *Rickettsia prowazekii*. Biochim. Biophys. Acta **1365:**105-111.
53. Andersson, S. G. E., and C. G. Kurland. 1999. Origin of mitochondria and hydrogenosomes. Curr. Opin. Microbiol. **2:**535-541.
54. Martin, W., and M. Mueller. 1998. The hydrogen hypothesis for the first eukaryote. Nature. **392:**37-41.
55. Vellai, T., K. Takacs, and G. Vida. 1998. A new aspect to the origin and evolution of eukaryotes. J. Mol. Evol**. 46:**499-507.
56. Sichertz-ponten, T., C. G. Kurland and S. G. E. Andersson. 1998. A phylogenetic analyses of the cytochrome b and cytochrome c oxidase I genes supports an origin of mitochondria from within the Rickettsiaceae. Biochim. Biophys. Acta **1365:**545-551.
57. Holland, H. D. 1994. Early proterozoic atmospheric change, p. 237-244. *In* S. Bengtson (ed.), Early life on earth. Columbia University Press, New York, N. Y.
58. Margulis, L. 1970. Origin of eukaryotic cells. Yale University Press, New Haven, Conn.
59. Karlberg, O., B. Canabäck, C. G. Kurland, and S. G. E. Andersson. 2000. The dual origin of the yeast mitochondrial proteome. Yeast Comp. Functional Genomics**. 17:**170-187.
60. Andersson, S. G. E. and Kurland, C. G. 1998. Reductive evolution of resident genomes. Trends in Microbiol**. 6:**263-268.
61. Krawiec, S. and Riley, M. 1990. Microbiol. Rev. **54:**502-598.
62. Segall, A., Mahan, M. and Roth, J. R. 1988. Science. **241:** 1314-1318.
63. Cole, S. T., Eiglmeier, K. *et al.* 2001. Massive gene decay in the *leprosy bacillus*. Nature. **409:**1007-1011.
64. Cole, S. T. *et al.* 1998**.** Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393:** 537-544.

65. Andersson, J. O. & Andersson, S. G. E. 1997. Genomic rearrangement during evolution of the obligate intracellular parasite *Rickettsia prowazekii* as inferred from an analyses of 52015 bp nucleotide sequence. Microbiology. **143:**2783-2795.
66. Andersson, S. G. E. *et al.* 1995. J. Bacteriol. **177:**4171-4175.
67. Keeling, P. J., Charlebois, R. L., and Doolittle, W. F. 1994. Curr. Opin. Genet. Dev. **4:**816-822.
68. Sicheritz-Ponten, T. and Andersson, S. G. E. 1997. Microb. Comp. Genomics **2:**123-139.
69. Munson, M .A. *et al.* 1993. Gene. **137:**171-178.
70. Friesen, An. G., and Friesen J. D. 1980. The nucleotide sequence of *tufB* and four nearby tRNA structural genes of *Escherichia coli.* Gene **12:**33-39.
71. Auer, J., **Lechner, K., and Böck, A.** 1989. Gene organization and structure of two transcriptional units from *Methanococcus* coding for ribosomal proteins and elongation factors. Can. J. Microbiol. **35:**200-204.
72. Buttarelli, F. R., Caloger, R. A., Tiboni, O., Gualersi, C. O., and Pon, C. L. 1989. Characterization of the *str* operon from *Spirulina platensis* and their evolutionary relationship to those of other prokaryotes. Mol. Gen. Genet. **217:**97-104.
73. Fraser, C. M., Gocayne, J. D., White, O., Adams, M., Clayton, R. *et al.* 1995. The minimal gene complement of *Mycoplasma genitalium*. Science **270:**397-403.
74. Furano, A. V. 1978. Direct demonstration of duplicate *tuf* genes in enteric bacteria. Proc. Natl. Acad. Sci. USA **75:**3104-3108.
75. Jaskunas, S. R., Lindhal, L., Nomura, M., and Burgess, R. R. 1975. Identification of two copies of the gene for elongation factor EF-Tu in *E. coli*. Nature. **257:**458-462.
76. Meng, B. Y., Shinozaki, K., and Masahiro, S. 1989. Genes for the ribosomal proteins S12 and S7 and elongation factors EF-G and EF-Tu of the *cyanobacterium Anacystis nidulans*: Structural homology between 16S rRNA and S7 mRNA. Mol. Gen. Genet**. 216:**25-30.
77. Pang, H., **Winkler, H. H**. 1993. Copy number of the 16S rRNA gene in *Rickettsia prowazekii*. J. Bacteriol. **175:**3893-3896.
78. Sela, S., Yogev, D., Razin, S., and Bercovier, H. 1989. Duplication of the *tuf* gene: a new insight into the phylogeny of eubacteria. J. Bacteriol. **171:**581-584.
79. Zurawski, G., and Zurawski, S. M. 1985. Structure of the *Escherichia coli* S10 ribosomal protein operon. Nucleic Acid Res. **13:**4521-4526.
80. Mira, A., Ochman, H., and Moran, N. A. 2001. Deletional bias and evolution of bacterial genomes. Trends in Genetic**. 17:**589-596
81. Brynnel, E. U., Kurland, C. G., Moran, N. A., Andersson., S. G. E. 1998. Evolutionary Rates for *tuf* genes in endosymbionts of Aphids. Mol. Biol. Evol. **15:**574-582.
82. Shigenobu, S., et al. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera sp*. Nature. **407:**81.
83. Muller, H. J. 1964. Mutat. Res. **1:** 2-9.
84. Felsenstein, J. 1974. Genetics **52:**737-756.
85. Chao, L. 1990. Fitness of RNA virus decreased by Muller ratchet. Nature. **348:**454-455.
86. Chao, L. *et al.* 1992. Muller ratchet and the advantage of sex in the RNA virus PHI-6 Evolution. **46:**289-299.
87. Chao, L. 1997. Evolution of sex and the molecular clock in RNA viruses.Gene. **205:** 301-308.
88. Andersson, D. I. & Hughes, D. 1995. Proc. Natl. Acad. Aci. U. S. A. **93:**906-907.
89. Moran, N. A. 1996. Accelerated evolution and Mullers ratchet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA**. 93:**2873-2878.

90. Andersson, J. O., & Andersson, S. G. E. 1999a. Genome degradation is an ongoing process in *Rickettsia.* Mol. Biol. Evol. **16:**1178-1191.
91. Andersson, J. O. & Andersson, S. G. E. 1999b. Insights into the evolutionary process of genome degradation. Curr. Opin. Genet. Dev**. 9:**664-671.
92. KEG Database, http://www.genome.ad.jp/kegg/kegg2.html.
93. Ogata, H., *et al*. 2001. Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. Science. **293:**2093-2098.
94. Andersson, J. O. & Andersson, S. G. E. 2001. Pseudogenes, Junk DNA, and the Dynamics of *Rickettsia* Genomes. Mol. Biol. Evol. **18:**829-839.
95. Makalowski, W. 1995. The impact of short interspersed elements (SINEs) on the Host Genome edited by Maraia, R. J., R. G. Landes Company, Austin.
96. Moran, J. V., Deberardinis, R. J., Kazazian, Ir., 1999. Science **283:**1530.
97. Kato, S*., et al.* 1998. Curr. Genet. **34:**303.
98. Britten, R. J., 1996. Proc. Natl. Acad. Sci. U. S. A. **93:**9374
99. Kidwell, M. G., Lisch, D. R., 2000, Perspective: Transposable elements, Parasitic DNA, and Genome Evolution. Evolution. **55:**1-24
100. Lin, Y. J., *et al.* 1998. Extended life-span and stress resistance in the *Drosophila* mutant *methuselah*. Science. **282:**943-946.
101. Ogata, H., *et al.* 2000. Selfish DNA in protein-coding genes of *Rickettsia*. Science. **290:**347-350.
102. Rudd, K. E. 2000. Novel intergenic repeats of *Eschericjia coli* K-12. Research in Microbiology. **150:**653-664.
103. Belkum, A. V., *et al.* 1998. Short-sequence DNA repeats in prokaryotic genomes. MMBR. **62:**275-293.
104. Herdman, M. 1985. The evolution of bacterial genomes. In: T. Cavalier-Smith, Editor, The evolution of genome size, John Wiley & Sons pp. 37-68.
105. Petrov, D. A. *et al.* 1996. High intrinsic mutation rate of DNA loss in *Drosophila*. Nature**. 384:**346-9. 106.
106. Mattick, J. S. 2001. Non-coding RNAs: the architects of eukaryotic complexity. EMBO reports. **2:**986-991.
107. Rocha, E. P. C. *et al.* 1999. Analyses of long repeats in bacterial genomes reveals alternative evolutionary mechanisms in *Bacillus subtilis* and other cometent prokaryotes. Mol. Biol. Evol. **16:**1219-1230.
108. Ochman, H., *et al.* 2000. Lateral gene transfer and the nature of bacterial innovation. Nature. **405:**299-304

Acknowledgements

I would like to thank everybody who have encouraged and helped me during this work at the former department of Molecular Biology and present department of Molecular Evolution. I am especially grateful to:

The former head of our department, Professor **Charles G. Kurland**, for giving me the opportunity to join this group. I greatly appreciate his outstanding scientific knowledge, encouragement and support. I also would like to express my gratitute to my supervisor and current head of the department, Professor **Siv Andersson**. You have been always understanding and helped me during the very tough and difficult period. I appreciate very much for all help that I got from you, especially during the last period of my research.
I am very grateful to all my colleges and friends. I have had great time with all of you and I am going to remember it as a very enjoyable time of my life.

I would thank more specifically:
Vildan, my friend, for the nice conversations and great time at the former department of Molecular Biology.
Ivica or Ivan or Tamas, my college, friend and former room maid, for always being helpful and kind.
Boris, you are wonderful.
Wagied, for all computer help and nice talking moments.
Carolin and Dave for critical reading of this thesis, and I am especially thankful to you Carolin for your efforts for arranging music for the disputation party.
Mats, you are and have been real gentleman to me. Thanks for all discussions and sharing knowledge.
Ann-Sofie and Kristina, thank you for all help and for nice and pleasant moments that you have shared with me during this time in this department.
Björn (known as gulligull), **Hans henrik, Ola, Ulla and Olof** for entertaining discussions during lunch, and Hans Henrik for your future help to fix the music, Ola for nice figures and Olof for great help with the manuscript.
Karin Olsson, for being helpful and kind whenever I need your help.
Robert, for being always nice and expanding my Swedish vocabulary by teaching me the ancient Swedish words.
A big thank to **Cessie, Micke, Lisa (lilla gumman), Cissi, Sean, Otto, Håkan, Karin, Rolf, Magnus, Linus ….** It has been wonderful to work with you.
Laleh, for your help, for the all PCR and sequencing reactions you did for me.
I also thank **Ann-Christin Syvenän** for her excellent supervising when I began my research at the former Molecular Biology department.
My precious family, **My parents,** for their great support during my whole life and specially in the last period of my scientific work.
My sisters, Shahnaz, Ladi, Maryam, My brother Sharid for their unconditional support and love.
Finally, I will thank you **Farshid** for your care, love, encouragements and understandings especially during the last weeks of my work.
My children **Arian and Lilian** for giving me happiness, and lots of fun which I needed to get through this.
Thank you every -body including those that I have forgotten to mention.